

CHAPTER ONE INTRODUCTION AND BACKGROUND

Introduction

In Michigan, the stakes attached to individual student and school district scores on the Michigan Educational Assessment Program (MEAP) tests are getting higher. Although no universally accepted definition of “high-stakes testing” exists, “high-stakes tests” may be defined as those whose results impact student placement, promotion, retention, graduation, diploma certification or endorsement, access to higher education, the distribution of dollars to students (typically by way of tuition payments) or to schools or school districts, or school accreditation, certification, ranking or the issuance of school or district level “report cards.” In Michigan, the highest stakes attached to the MEAP are in three categories: higher education tuition dollars for students, “Golden Apple” monetary awards for individual schools, and school accreditation.

The MEAP Merit Scholarship Program awards up to \$2500 tuition money to graduating high school students who pass four specified portions of the high school MEAP tests: mathematics, science, reading, and writing. The program also puts tuition funds in escrow (currently up to \$500) for middle grades students who score well on the various sections of the seventh and eighth grade MEAP tests.

The “Golden Apple” awards, given under the auspices of the Michigan Department of Education, provide a minimum of \$50,000 for schools which show dramatic improvement (a composite school score

increase of at least 60 points over three consecutive years) on the MEAP tests.

The Michigan Department of Education School Accreditation Program has historically relied on schools' MEAP scores for the determination of accreditation. Although recent proposals to base school accreditation solely on MEAP scores have been rejected by the State Superintendent of Public Instruction, it is likely that MEAP scores will remain a primary means of awarding or denying schools state accreditation.

As intended, these high stakes put pressure on students, teachers, and other educators to produce high test scores. This situation cannot be summarily labeled as negative. If the high stakes act as motivators for students and for educators, and if true higher student achievement results, then it may be that these high stakes are appropriate. Several relevant questions, however, present themselves:

1. Do higher standardized test scores and high scores on the MEAP¹ reflect genuine higher student achievement, particularly achievement in use of the higher cognitive skills of application, analysis, synthesis, and evaluation, or may they simply reflect narrow efforts on the parts of educators to raise scores?

2. Do standardized tests and MEAP tests generally assess students' abilities to be effective team members, real-life problem solvers, contributing members of society, responsible citizens, and innovators? If not, then should high stakes be attached to such tests?

3. If household income and the level of mother's formal education

¹ The MEAP tests are not standardized because student scores are not normed. However, they share characteristics with standardized tests, including their formality, objectivity, use of selected responses, administration to large numbers of test-takers at the same time, and reliability and validity. Therefore, some issues relating to standardized tests and to MEAP tests may be considered together.

are indeed the two best general predictors of students' standardized test scores, is it then appropriate to judge or rank schools or school districts based on those scores, or on the scores of MEAP tests? That is, in what ways is it legitimate to compare the standardized test scores or MEAP test scores of schools with differing demographics?

4. Might those potentially inappropriate comparisons tend to increase the opportunity gap between students of lower socioeconomic status and those of higher socioeconomic status? For instance, might the aforementioned Golden Apple Awards tend to be given to schools that are much less needy than other ineligible schools?

5. By their very nature, nearly all standardized tests, and certainly all of the MEAP tests, are also reading tests. Might it be that students who are sufficiently proficient in the areas of mathematics, science or social studies, but who have reading deficiencies, are unable to demonstrate their proficiencies by way of traditional testing? If this is so, again, are high stakes appropriate?

Those five questions take on additional urgency in light of Grant Wiggins' "Five Classroom Assessment Premises":

1. Assessment should improve performance, not just audit it.
2. Teachers should assess that which is valued, so that they value the assessment.
3. Assessment is instruction.
4. Assessment should inspire performers and encourage quality work, not intimidate performers or encourage perfunctory compliance.
5. Local assessment is better.

Many cautionary warnings have been put forth regarding the use of standardized test results for high-stakes purposes. An analysis of National Assessment of Educational Progress (NAEP) data by the National Center

for Fair & Open Testing (FairTest) found

...that students in high-stakes testing states were more likely to score below the NAEP level of “basic” in both math and reading. Students from states without required graduation exams more frequently scored at or above “proficient.” According to the NAEP data, states with high school graduation tests were less likely to show significant improvement from prior years (p.1).

In an article in the *American School Board Journal*, Kaufhold (1995) sounded a strong warning that school board members and administrators should not rely heavily on the results of standardized tests. He cited numerous reasons: curricula vary, money makes a difference to academic success, children are not standardized, good facilities lead to higher test scores, different types of textbooks can be influential, school climates vary, some schools coach students in test taking and some don't. Kaufhold's observation that “money makes a difference to academic success” sets the tone for the remainder of this introduction.²

The 2001 flap over MEAP “testing irregularities” added an element of urgency to the test debates in Michigan. In June of that year, the Michigan Department of Education and the MEAP office released a list of 71 schools (mostly from the Detroit Public Schools) that were under suspicion of testing irregularities during the January 2001 administration of the tests to fifth graders. The most common irregularity (as identified by the Michigan Department of Education) was the appearance of common topics, themes and even phrases in students' writing samples.

² This is not to imply that the author accepts that high MEAP scores are good indicators of student academic achievement. That is a separate question.

By July, most schools had responded to the allegations, with the most frequent explanation being that students had memorized definitions and were using them in their writing. (At least one school was simply on the list because of a clerical error). Even though few of those “testing irregularities” proved to be substantial, the affair begged the question: Is it possible that the increased pressure on schools to simply produce high MEAP scores has contributed to an atmosphere that may promote cheating on the tests? If so, the reliability of MEAP scores is called into question, and use of the scores for high-stakes purposes becomes problematic.

Michigan teachers already are suspicious of the MEAP and of MEAP scores. A study of teachers’ perceptions of the MEAP (Ray-Taylor, 1993) found teachers frustrated with what they perceived as pressure to improve MEAP test scores with little direction as to how to do that. By way of several focus group and individual interviews, Ray-Taylor found that teachers repeatedly described there being “too much work with too little time to do it” (p. 11). The same study found that teachers felt that there were insufficient guidelines concerning what constituted appropriate test preparation. These concerns, Ray-Taylor noted, led Michigan teachers to feel cautious regarding allowing testing to lead any curriculum reform efforts or to use MEAP test results for high-stakes purposes.

While not conclusively demonstrating that standardized testing is illegitimate, other studies provide a general caution concerning the use of such tests (and, by implication, the MEAP) and attaching high stakes to the results of such tests. Williams (1993) collected data from 23 Michigan high school students who had received “test sophistication training” (specifically, the “Ace That Test Preparation for Success

Guide,” a test preparation course). Williams found that they scored significantly higher on the MEAP tests than 23 randomly matched students who did not receive such training. The results of good assessment instruments are minimally impacted by any student test preparation (beyond the review of content); therefore, the implication is clear: If students can score higher on a test after generic, noncontent-related preparation, is not the validity of the test called into question?

High stakes are getting higher. A national academic testing program for students in grades three through eight, proposed by the Bush administration and approved by Congress, puts some school accreditation and funding at risk. The President has frequently spotlighted the high-stakes statewide standardized testing program in Texas, (the Texas Assessment of Academic Skill, or “TAAS”), and has stated his desire to see a similar test given nationwide. This initiative may be seen as an intrusion into what has traditionally been a function of the states, and the situation highlights the need to debate the wisdom of attaching high stakes to such tests.

Evidence exists that socioeconomic factors impact student achievement in general and standardized test scores in particular³. Roscigno (1996), Grinion (1999), Smith (1999), Markel-Fox (1993), Adams (1994) and Ainsworth-Darnell (1999) all provide recent evidence supporting the long-established notions that individual student characteristics (such as family income, number of parents living in the home, etc.) are highly correlated with poor student performance on standardized tests, and that school and district demographic factors (such as low median family income, per-capita income, median area

³ I do not necessarily accept that standardized test scores are good indicators of student academic achievement, any more than I accept that MEAP scores are good indicators.

housing prices, etc.) are also correlated with poor school and district test scores.

Most research investigating factors that impact MEAP scores in particular has focused on nonsocioeconomic factors, such as pupil/staff ratio, class size, district per-pupil spending⁴, principal leadership characteristics, years of service of staff, teacher style, etc. Compared to data on the SES-standardized test scores relationship, there is less available evidence that socioeconomic factors impact MEAP scores in particular.

There exist no full correlational studies of a wide range of SES factors for all of Michigan's 500-plus non-charter⁵ school districts (median district family income, percent of lone parent households, free and reduced lunch percentages, etc., as the independent variables) vs. MEAP scores (as the dependent variable).

This study was undertaken, therefore, to determine the strengths of those correlations. The first research question was: "How do Michigan public school district SES factors correlate with aggregate district MEAP scores?"

The generation of the attached set of correlational coefficients was a prelude to answering the second research question, "How can individual correlation coefficients (district free/reduced lunch percent vs. district MEAP scores; percent of lone parent vs. district MEAP scores, etc.) be combined to form an equation strongly predictive of district MEAP scores?"

⁴ Although district per-pupil spending is an economic factor, it is considered an "internal" factor for purposes of the proposed study, and not a student SES factor.

⁵ There is more on the exclusion of Public School Academies (Charter Schools) in the Methodology section.

Statement of the Problem

Despite considerable evidence that high-stakes testing distorts teaching and does not give very stable information about school performance, test results have become the dominant way states, politicians, and newspapers describe the performance of schools...The consequence of statistical accountability systems will be the narrowing of purpose for schools, impatience with reform, and the continuing erosion of political support for publicly funded schools (Dorn, 1998).

Whether or not one agrees with Dorn, the continuing development of public education policy, particularly at the state and national levels, is at an important juncture, with an unmistakable increased focus on school accountability. President Bush's education policy emphasizes higher standards, increased accountability and consequences for failure, and consistently equates educational progress, success and achievement with student test scores. Some form of national standardized testing for elementary, middle grades and high school students is in the offing, and policymakers appear to be operating under two assumptions:

1. Standardized tests and tests such as the MEAP tests provide a valid and reliable indicator of individual student academic achievement.
2. Aggregate student scores on such tests provide valid and reliable indicators of the quality of schools and school districts; that is, schools whose students receive high test scores may legitimately be labeled "successful," and schools whose students receive low test scores may legitimately be labeled "failing."

If it is true that individual student SES factors provide reliable general predictors of academic performance (as indicated by standardized or MEAP test scores), and that the general SES of students in schools or school districts provides a reliable general predictor of school or district test scores, the legitimacy of assumption number 2 above is called into question, in that a factor beyond the schools' control (SES)⁶ is associated with (correlates with; is predictive of) student test scores, a presumed outcome of schools.

The problem, therefore, is that education policymakers may be operating under a significant false assumption, namely, that external (to schools) student/school/district SES factors are not associated with aggregate district student test scores, and that, therefore, use of such scores for high-stakes purposes (particularly the labeling of schools as “successful” or “failing”) is fair and reasonable. This study determined, with noted limitations, how several SES factors are associated with aggregate district MEAP scores, as defined below.

Purpose of the Study/Justification and Significance

This study had a narrow and specific focus: It established the strength of relationship between several district SES factors and aggregate district student MEAP scores. A finding of even a single significant correlation between any individual district SES factor and aggregate district MEAP scores provides information to education policy makers that calls into question the use of MEAP test scores for any high-stakes purposes, including labeling schools and districts as successful or

⁶ SES factors are generally associated with individuals, rather than schools or school districts. For the purposes of this study, however, district averages and district household averages will be referred to as school and district SES.

failing. It also provides a reliable starting point for further research into these relationships.

The second purpose of the study was to formulate a predictive equation, using multiple regression, based on a combination of SES factors that were found to relate significantly to district MEAP scores. The ability to develop such an equation spotlights the illegitimacy of use of an essentially uncontrollable factor (SES) for high-stakes purposes. The predictor equation provides some additional degree of guidance for policymakers who influence education resource allocation. A possible implication for them might be that public policies that hold the most potential to decrease poverty, and child poverty in particular, will also positively impact student achievement, at least as indicated by student test scores.⁷

The study did not address all of the issues noted above. It only examined the relationship between district SES factors and aggregate district high school MEAP test scores. The district level of analysis was chosen for two reasons:

1. Most MEAP score reporting is done at the school district level, with published comparisons done at that level as well.⁸
2. As a practical matter, SES indices are most widely available at the district level.

A logical and much larger follow-up study would examine the relationship between individual school SES factors and school MEAP scores.

There was personal significance to the study. My experience as a K-12 public school teacher and administrator in Michigan, with extensive

⁷ Again, this depends upon whether or not one accepts the notion that test scores are legitimate indicators of student academic achievement.

⁸ Note: Neither the MEAP office nor the Michigan Merit Scholarship Award office publish school or district MEAP test scores in rank order.

involvement with the MEAP, engendered a passionate interest in the relationships that this study examined. Furthermore, as a beginning researcher, this study gave me needed experience in basic correlational, predictive, and multivariate research.

Delimitations

This study used a specific definition of “aggregate district MEAP scores.” (See the Methodology section.)

The Michigan Department of Education, for reporting purposes, defines each of the approximately 150 state Public School Academies (Charter Schools) as individual school districts unto themselves. Because the nature of individual Charter Schools is different from that of traditional school districts, Charter School “districts” were not included in this study, and no inferences regarding their MEAP test scores is valid. Furthermore, the handful of Michigan non-K-12 public school districts were excluded from this study, again, based on their distinctive nature. Of the 784 entities identified by the Michigan Department of Education as school districts, only the 522 “unified” districts (K-12) were studied.

As noted by Bobbett, French, and Achilles (1993), “The researcher’s selection of a particular statistical treatment has a large impact on the study's conclusions” (p. 13). The only unchallengeable outcomes of this study are the specific correlational coefficients generated by the Pearson Product Moment test, based on the input data. All other conclusions, therefore, are inferences, subject to examination and debate.

In that the SES factors that were used were district averages or percentages, results of the study provide no predictions regarding

school-level SES/MEAP test score correlations. More importantly, the results say nothing about student potential or about how individual students may score on MEAP tests,.

The very largest Michigan school district (Detroit) was excluded from participation in this study, in that it covers an extremely large geographic area, at least one order of magnitude larger than the next largest district (Grand Rapids).

Bobbett et al. (1993), in a study which explored the relationship among several Tennessee school district variables and the relationship between each variable and average student test scores, concluded that the variables did not have the same impact at the elementary, middle school, high school and district levels. The conclusions of this current study therefore apply only to aggregate district high school MEAP scores and not to MEAP scores of lower grades or to individual high school buildings (unless a district has only one high school).

Limitations

The study examined the relationship between district SES factors and aggregate district scores for one MEAP reporting year only (2000). Further research is needed to establish that the discovered correlations and the multivariate predictive equation are reliable over time.

The SES data and MEAP scores used are from several different years (1997 to 2000). All data, however, are the latest available.

Assumptions

This study did not address the issue of whether MEAP scores are legitimate indicators of desired student achievement in the first place. Addressing this issue requires, at a minimum, defining “desired student achievement,” a difficult task itself (Whose definition of “desirable?” “At what levels?”, etc.)

As with all correlational studies, even the discovery of high correlational coefficients would not confirm causality between factors; unstudied and unidentified intervening factors may explain high correlational coefficients. This study, therefore, did not establish or attempt to establish causality between SES factors and MEAP scores.

Methodology

This study provided a correlational/predictive and multivariate analysis of several district-level SES factors and aggregate high school student MEAP achievement. Michigan Department of Education data bases, United States Census data bases, and Standard and Poor’s data bases provided the data sources. Support for this design is noted at the end of the Methodology section.

A review of the literature provided guidance in the identification of SES factors that might correlate significantly with MEAP test scores. It has been noted that the MEAP tests are not standardized, norm-referenced tests. However, the literature review found suggestions of correlations between a wide variety of SES factors and standardized test

scores, so those factors were used in this study, which focused on MEAP test scores only.

The definition of each factor was established; these are noted and discussed in Appendices A and B. Identified SES factors for all 522 Michigan unified (K-12) school districts were collected on a computer spreadsheet (using Microsoft Excel). As noted in the Delimitations section above, there are many ways to define a district's "MEAP scores." For purposes of this study, the definition of "district MEAP scores" was "percent of district high school tests that were given a 'passing' score on four of the five MEAP tests, the Social Studies test being excluded." In that the preceding acts as the criteria for determining which students earn the Michigan Merit Scholarship, "district MEAP scores" may be alternately (and significantly) defined as "the percent of district students taking the high school MEAP tests in the 1999-2000 school year who qualified for the Michigan Merit Scholarship Award." Students who qualified for this award were those who scored in the top two state-reported achievement categories ("Exceeded Michigan Standards" and "Met Michigan Standards") for the science, mathematics, reading, and writing tests. (Neither the Department of Education or the Merit Award Office uses the terms "passing" or "failing" in any of their reports.)

This study focused on public policy implications of SES-student achievement correlations. The Michigan Merit Scholarship Award Program is public policy: that is why the definition of "MEAP scores" used in this study was tied to the Scholarship Program.

Drawing from the spreadsheet, individual correlation coefficients for each SES factor/aggregate district MEAP test scores relationship were generated using the SPSS computer program. The Pearson Product-Moment correlation was used to produce the coefficients. For each

correlation, a scatterplot was generated as an aid in determining if there existed sufficient linearity to consider the coefficient as a quality indicator of the studied relationship. Krathwohl (1997) offered that “Anyone using correlations is well advised to create a scatterplot of the observations” (p. 380).

For purposes of this study, a Pearson Product-Moment correlation (R) of positive or negative 0.5 was considered strong. SES factors that exhibited strong correlations with MEAP test scores were then added into various multiple regression combinations. Combinations that offered the power to predict MEAP scores more powerfully than the individual factors were considered. (The study did not examine the relationship among the variables.)

The results of the individual correlational tests and the use of multivariate analysis are presented, analyzed, and discussed in Chapters Four and Five.

Support for the study’s methodology came from several sources. Gay and Airasian (2000) offered that “Correlational studies typically investigate a number of variables believed to be related to a major, complex variable, such as achievement” (p. 321). Further,

In larger correlational studies, a number of variables believed to be related to a complex concept such as achievement may be examined. Variables found not to be highly correlated to achievement may be eliminated from further consideration, while variables that were highly correlated might prompt further examination (p.12-13).

“Elimination from further consideration” of low-correlation variables and “further examination” of high-correlation variables are precisely what were done in this study.

Fraenkel (2000) identified three major types of problems that are the focus of correlational studies, one of which is “What are the relationships among a large number of variables and what predictions can be made that are based on them?” (p. 368). Again, from Gay (2000), “The purpose of a correlational study is either to establish a relationship (or lack of it) or to use relationships to make predictions” (p. 12).

Sowell (2001) added “Correlational designs are used to test hypotheses about the magnitude and direction of possible relationships between (among) variables. Such studies provide explanations about associations among data taken on two or more variables” (p. 100).

Correlational studies such as this one can become predictive studies. Fraenkel (2000) stated that “predictive research is a subset of correlational research.” (p. 361). Gay (2000) offered that “More than one variable can be used to make predictions. If several predictor variables each correlate well with a criterion, then a prediction based on a combination of those variables will be more accurate than a prediction based on any of them” (p. 332).

Sowell (2001) concurred: “Prediction studies are used to see if known relationships among variables...” (which the first research question will establish) “...can be used for prediction purposes. Research begins with variables known to be strongly correlated. One or more predictor variables in this set of variables is (are) used to make predictions about other variables in the same set, called criterion variables” (p, 101). This reflects the approach used in this study.

Numerous other sources indicate that multiple regression equations use variables that are known to individually predict (correlate with) the criterion to make a more accurate prediction. McMillan (2001) stated that “...by having several predictor variables, you would be able to

make a more accurate prediction...All of the predictor variables can be combined to form what is called a **multiple regression predictor equation**" (p. 293).

Definition of Terms⁹

correlational research. Research in which data are collected to determine the extent to which a relationship exists between variables.

factor. An attribute which can be expressed by way of an assigned value.

high stakes. In relation to academic testing, such as standardized tests and the MEAP test: student academic placement, promotion, retention, graduation, diploma certification or endorsement, school or district accreditation or designation (such as "failing"), or the distribution of dollars to districts, schools or students (usually by way of tuition payments.)

MEAP. The Michigan Educational Assessment Program, which provides, administers and scores tests given in mathematics, science, language arts and social studies annually to Michigan students at specific grades, fourth through high school.

multiple regression. A statistical technique for using a predictive equation that includes two or more variables, combined to predict a criterion, or independent, variable.

⁹ Commentary on each of the variables used in this study, including definitions and identification of data sources, is found in Appendices A and B.

Pearson Product-Moment Test. A statistical test that determines the direction, shape and strength of the relationship between two factors or variables.

quantitative research. A study approach that attempts to answer a question or confirm a hypothesis. Data collection precedes an attempt to find statistical correlation or causality between or among studied factors.

SES. Socioeconomic status. Generally refers to the level of an individual's income, housing, education, accessibility to community services, etc. May also refer to an entire community as well.

scatterplot. Coordinate planes on which are plotted--as ordered pairs--two factors at once. Highly linear (or strong) relationships between variables result in scatterplots resembling flattened ovals. The flatter the oval, the more linear (or strong) the relationship.

standardized testing. Generally formal, objective, selected response testing given to large numbers of test-takers at approximately the same time. Individual scores are typically given as comparisons with others' scores. Reliability and validity may be characteristics of standardized tests.

Organization

The remainder of this document is organized in the following manner:

Chapter Two provides a review of the literature relevant to the topics of high-stakes testing, state and federal education policy (especially as they relate to high-stakes testing), and correlations between various SES factors and student achievement on standardized tests or MEAP tests. The review also provides a sociological, educational, and economic context for the study.

Chapter Three details the research methodology, primarily the steps in generating both the first level correlational coefficients (district free/reduced lunch percent vs. HS MEAP passing rate; percent of lone parent households vs HS MEAP passing rate, etc.) and the predictive (multivariate) equation.

Chapter Four consists of a presentation of the generated data in chart and narrative form, along with an analysis.

Chapter Five presents a summary, conclusions, possible inferences, and policy implications based on the results of this study, and recommendations for further research.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Introduction

This chapter provides a review of literature (research and articles) related to the areas of high-stakes K-12 public school student testing, public policy regarding such testing, and the relationship between socioeconomic factors and achievement test scores. Literature that pertains to Michigan Educational Assessment Programs tests and to other states' tests receives special attention.

Although descriptions of federal and state-level education policy in this chapter come from various sources, the author acknowledges that *Education Week* has been consistent in its reporting of the topic, and those reports are drawn on extensively. The 1977 report of The National Academy of Education to the Assistant Secretary of Education, *Improving Educational Achievement*, is treated as a seminal document regarding the evolution of federal education policy, especially for the GOP, and is referred to frequently in this chapter.

High-Stakes Testing

It has been noted elsewhere in this document that there exists no formal or universally accepted definition of "high-stakes testing" as the term relates to standardized tests, but that the term generally refers to tests whose results impact student placement, promotion, retention, graduation, diploma certification or endorsement, the distribution of

dollars to students (usually by way of tuition payments) or to schools or school districts, or school accreditation, certification, or ranking. In a study of Utah teachers, Beck (1995) found that simply publicly reporting and comparing school and district state-mandated test scores made teachers feel as though the tests were high-stakes tests, even if no other stakes were attached.

There exists a demand for higher public school student test scores and accompanying high stakes. Carter and Cunningham (1997) stated that, although true excellence takes time, the American public is demanding that standardized test scores be up now. Popham (1998) explained that state and federal governments have enacted accountability laws tied to student standardized test scores as a method of making schools show whether or not their students are making academic progress.

Several studies can be interpreted as supporting the practice of attaching high stakes to standardized test scores, and several authors have written in favor of the practice in general. In his dissertation, Herrick (1999) studied how various state-level tests impacted teacher classroom practice, and he concluded that some elements of state-level testing indeed improved teacher practice. (It should be noted, however, that only performance assessments were found to have such a positive effect.) Vogler (2000) examined the impact of the Massachusetts Comprehensive Assessment System (MCAS) tests, which he described as high-stakes, on teachers' classroom practice, and noted an increase in the use of rubrics, critical-thinking questioning, problem-solving activities, writing assignments, and inquiry-based investigation. The MCAS contains relatively few traditional multiple choice and binary (true-

false) questions, and as with Herrick, Vogler studied the impact of performance assessment.

Supporting and, perhaps to some extent, explaining those two studies, Demars (1998) found that test item type (multiple choice, true-false, short answer, constructed response items, etc.) on standardized tests mattered. Specifically, Demars found that students were more likely to respond positively to constructed response items. He also found that when standardized tests had consequences (high stakes), students tended to score higher.

In a widely-quoted article appearing in the *American School Board Journal* (1997), Smith and Jorgenson defended attaching high stakes to student standardized test scores by noting that standardized tests are recognized in the “real world.” They criticized teachers and administrators who view standardized tests as wastes of time and who do little or nothing to help prepare students for such tests, claiming that such an attitude will “leave gaps in students preparations for tests” (p. 46).

Further, they stated that the widespread teacher and administrator ambivalence toward standardized testing has “a devastating effect” on students’ test performance. (It appears to be an assumption underlying such statements that standardized test scores tell us what we want to know about student learning.)

While the public in general may be supportive of high-stakes standardized testing, the practice, and even the topic, has divided many communities. According to Hardy (2000), the issue has the potential to polarize both communities and boards of education, and he noted that there are numerous protests, including lawsuits, against such tests, and that this is a fairly new phenomenon.

Shepard (2000) cited Skinner (1954) to provide a possible theoretical explanation for the general public support of high-stakes standardized testing: “The whole process of becoming competent in any field must be divided up into a very large number of small steps, and reinforcement must be contingent upon the accomplishment of each step” (p. 5). Shepard noted that such notions of efficiency echo the scientific management theories of Thorndike, Taylor and Hawthorne, which, she stated, remain the paradigm of public education in spite of generations of reform efforts. Shepard offered that the measurement approach to classroom assessment (as opposed to more holistic approach) “is exemplified by standardized tests and teacher-made emulations of those tests” (p. 4) and warned that such an approach presents a barrier to the implementation of more constructivist (and non-Skinnerian) approaches to instruction.

While the general public often offers support for high-stakes testing-as-accountability, the researcher and writers cited earlier represent a minority in education; few researchers or practitioners have offered strong support of the practice. Wasserman (2001) offered perhaps the most devastating anecdotal view of the impact of high-stakes standardized testing on students. Observing several first grade classrooms in the midst of preparation for testing, he noted that “The preparations alone were exhausting and nerve-wracking. Normally spirited, they [students] were now extraordinarily quiet; the stress in the air palpable” (p. 30). As testing moved to actual administration, he noted “Students crying, frustrated to the point of being red-faced, or treating the test as a coloring exercise. It didn’t take an expert to know that something was terribly rotten about the use of such a test and about the educational significance of the results” (p. 30). Wasserman (2001) noted:

When it comes to assessing quality, accuracy in measurement is even more elusive. Take, for example, “expert” opinions on which nominee would win the Academy Award for best actor. Or which actors should have been nominated and were ignored. Take, for example, the widely disparate judgments of informed opinion about a particular work of art. A piece of music. A novel. A candidate’s ability to govern. A politician’s performance in office. An impeachable offense. A prison sentence for a non-offender. Take, for example, the widely differing views about what constitutes adequate health care...(p. 31).

Even Oscar Buros, editor of the widely-used and widely-respected Buros’ Mental Measurement Yearbooks, was quoted by Wasserman as stating “Many of you know that I consider that most standardized tests are poorly constructed, of questionable or unknown validity, pretentious in their claims, and likely to be misused more often than not” (p. 32).

Popham (1999) was more specific regarding that misuse:

These days, if a school’s standardized test scores are high, people think the school’s staff is effective. If a school’s standardized test scores are low, they see the school’s staff as ineffective. In either case, because educational quality is being measured by the wrong yardstick, those evaluations are apt to be in error...Most educators do not really understand why a standardized test provides a misleading estimate of a school staff’s effectiveness. They should (p. 8).

Wasserman concluded his criticism of high-stakes standardized testing by quoting Edward Fiske, former editor of the *New York Times*,

who attacked the use of standardized tests because (according to the author),

1. They assume a single, correct answer to problems.
2. They measure how good students are at recognizing information, not generating it, ignoring synthesis and problem solving.
3. Most are timed, so they reward speed, not profound thinking.
4. They focus on basic skills.
5. They emphasize isolated learning, not integration.

Although Wassermann's criticisms were unusually pointed, others, by way of research or opinion, are on the same side of the high-stakes standardized testing fence. Rothman (1996) argued that the emphasis on standardized testing is harmful to students, academically at-risk and minority students in particular, and that many useful alternative assessment and accountability tools exist. Vornberg and Hart (2000), who have extensively examined the Texas Assessment of Academic Skills (TAAS), noted that high stakes accountability for student performance has caused a narrowing of the curriculum and forced teachers to deliver instruction strictly from a behaviorist viewpoint (echoing Shepard's citations of Skinner). Vornberg and Hart have also found that Texas principals and teachers feel that too much time is spent on preparation for the TAAS, and that they see in the TAAS a push for Skinnerian-behaviorist theory. The same Hart (1999) noted that the TAAS has not impacted curriculum itself, but has impacted only planning and delivery of instruction. Teachers, Hart found, feel they devote too much time to teaching specific objectives measured by TAAS and that teacher training in Texas has focused on raising scores to the neglect of current research. (More on the TAAS in the section entitled State-Level Tests.)

K-12 practitioners (as opposed to policymakers) have at best mixed perceptions regarding standardized testing and state-level tests, with more negative than positive attitudes reported. Danielson (1999) found that North Carolina public school principals believed the following:

1. While the public school rating system in North Carolina (based primarily on state test scores) had a “profound impact” on participants, morale issues in schools with low ratings are a pressing issue.
2. The first release of school ratings in 1997 was responsible for strained relationships within districts and schools.
3. The nature of North Carolina district curricula is narrowing to become more aligned with the state test.
4. Time spent on state test preparation has increased “dramatically” in all schools.

The North Carolina principals who participated in Danielson’s study also reported that there has been a return to tracking of students, specifically to meet the demand for higher test scores.

Practitioners’ generally negative perceptions of high-stakes testing underscore Jadallah and Partin’s (1995) observation that the best educational lesson from the twentieth century is that school changes are successful only if educators accept and incorporate the changes.

The authors of at least two reports in *Educational Measurement: Issues and Practices* discussed the curriculum-instruction-assessment alignment issues posed by high-stakes testing. Herman and Golan (1993) reported on their attempts to measure the effects of standardized testing (and “standardized-like” testing, such as most state-level tests) on teacher practice. They found that teachers in high-stakes situations tended to spend substantial time on test objectives during lesson planning, often excluding objectives that they felt were more important

for students, resulting (according to the teachers) in a strong discrepancy between what should be taught and what standardized tests emphasize. Teachers in those situations also took additional time to review test-taking strategies with students. Herman and Golan concluded that such high-stakes testing may trivialize the instructional process, distort the curriculum, and usurp valuable instructional time. Their findings echoed Smith (1991), who examined “the unintended consequences” of external (typically standardized or state-level) testing in elementary schools. He found that pressure to increase scores caused some teachers to neglect objectives that the tests did not assess.

Cooley (1993) commented that “...tests such as...(the) NAEP are not keyed to a specific curriculum. They are not examinations on what has just been taught” (p.10).

Tuch (1996) examined the impact of high-stakes *performance* testing on teachers and on student learning, and his conclusions mirrored those of Herman and Golan and Smith: the use of high-stakes performance tests (in place of standardized tests) still has negative effects on teachers. He found that such tests cause teachers to feel overwhelmed, that they perceive an atmosphere of threat, and that they sense pressure to get scores high as opposed to keeping quality high.

Of course, given perfect curriculum-instruction-assessment alignment, issues of “letting the test tail wag the instructional dog” might not be of concern. Bushweller (1997) noted that teaching to the test is fine if there is true alignment between instruction and assessment, but that generally what is billed as “curriculum alignment” is simply teaching to the test. As long ago as 1988, Madaus warned that, contrary to the traditional (and “correct”) “flow” from curriculum to instruction to assessment, testing is rapidly replacing the role of curriculum as the

mechanism for defining what school is about. Popham (2000) called this “the score-boosting game,” which he described as dangerous in that standardized tests fail to give teachers sufficient descriptive detail to plan and deliver instruction (as would be expected of a well articulated curriculum). Spady (2001) stated that the calls for standards and standardized testing represent “a legacy of the narrowest and most mechanistic assembly line set of ‘educational’ prescriptions imaginable.” The push to simply raise minority students’ test scores was seen by McNeil (2000) as leading to a focus on the test to the exclusion of many other forms of education.

Thomas (2001) offered a pithy view of the entire issue: Rather than teaching toward the test, and using results from standardized tests as an opportunity for reflective curriculum restructuring, some schools are actually using the test itself as the curriculum...The (Bush) proposal rests on two faulty assumptions: standardized tests are the best way to measure students’ academic success, and schools are failing simply because they are not trying hard enough and the threat of sanction will transform schools...Studies show that heavy emphasis on standardized tests reduces both the quality and quantity of what is taught in school. Often, the curriculum is truncated by pressures to elevate test scores....Advocates of high stakes standardized tests maintain that these assessments are objective indicators of instructional and school quality. However, testing experts disagree, pointing out that standardized, nationally normed tests give children, parents, school systems, legislatures and others misleading reports on student

achievement and how schools are faring. The inappropriateness of using such test results to gauge instructional quality is due, in large part, to the lack of alignment between the test and the curriculum emphasis in a given locality (pp. 1-3).

Saks (1999), recognizing that nearly every state has some sort of public school accountability system (although not necessarily state-sponsored tests), offered that “High stakes tests cause schools to focus instruction on test objectives only” (p. 6).

The *Phi Delta Kappan* has been the venue of choice for several writers who take issue with high-stakes testing. Bernauer and Cress (1997) provided a virtual summary of Danielson’s study in the *Kappan* by offering that “high-stakes accountability is driving instruction in the wrong direction” (p. 72). In another, more recent *Kappan* article, Rotberg (2001) put it plainly:

Accountability has become the centerpiece of political rhetoric on education reform. The underlying assumption is straightforward: hold teachers accountable for students’ scores on standardized tests, and academic standards will rise. Sounds good. But it doesn’t work. High stakes testing weakens academic standards (p. 170).

Jones and Whitford (1997) charged in the *Kappan* that accountability forces teachers to focus on whatever is thought to increase scores, not on individual student needs. Shepard (1994) stated that high-stakes testing in early grades tends to distort the curriculum and promotes developmentally inappropriate instructional practices, including long periods of seat work and overuse of worksheets. Similarly, Edelman (1994) has charged that pressures to improve test scores

causes some teachers to adopt instructional strategies that mirror test formats and focus their time on shallow, disconnected learning.

A heartfelt, summary protest of state-level testing and accountability systems was offered by Bigelow (1999), cited by Anderson (2001):

I want the State to abandon its effort to turn me into a delivery system of approved social information. I want it to support me and other teachers as we collaborate to create curriculum that deals forthrightly with social problems, that fights racism and social injustice. I want it to acknowledge the legitimacy of a multicultural curriculum of critical questions, complexity, multiple perspectives, and social imagination. I want it to admit that wisdom is more than information, that the world can't be chopped up into multiple choice questions and that you can't bubble in the truth with a number two pencil (p. 323).

Fullan (1999) went so far as to offer this hypothesis:

The greater the emphasis on academic achievement through high stakes accountability, the greater the gap becomes between advantaged and disadvantaged students. The main reason for this is that poor performing students do not need more pressure, they need greater attachment to the school and motivation to want to learn. Pressure by itself in this situation actually demotivates poor performing students. We need to work on motivation and attachment along with academic achievement (p. 18).

When public school board of education members have bothered to weigh in on the topic at all, it is generally with some ambivalence,

although there is recognition that relying heavily on standardized and state-level test results may not be wise. Board members are often aware that curricula vary, that money can make a difference, that good facilities can lead to higher test scores, that “children are not standardized,” that textbooks can be influential, that school climates vary, and that some schools’ teachers coach students in test-taking (Kaufhold, 1995). Hurwitz and Hurwitz (2000) offered school board members this advice: there are common characteristics shared by states and districts with “successful” high-stakes testing programs, and these include bipartisan political support and the support of a broad coalition of special interest groups over time.

Nonetheless, the National Association of State Boards of Education (the NASBE) has gone on record as stating the state assessments of student achievement should have consequences, both for the students who take them and for the schools who administer them. Achilles and Mitchel (2001-2002) characterized such calls as “ill-advised ‘get-tough-on-kids’ political mantra(s)” (p. 10). (The same NASBE report did suggest that states should use several types of tests to measure student learning before judging a student or a school.) (Lawton, 1997). Hess and Brigham (2000) warned public school board members that they should insure that any high-stakes testing must promote the benefits claimed, must not do harm, and must limit costs.

In summary, high-stakes testing, for all its ubiquitousness, suffers from a general lack of support from educators. As discussed in the Policy section of this literature review, the chief supporters of high-stakes tests tend to be policy makers.

Policy

As might be expected, federal policy regarding education changes with each new Democratic or Republican administration. Left- and right-wing views on high-stakes testing, however, approach unanimity at the national level. Numerous authors have noted that the recently enacted education bill, which will shortly require annual tests in reading and in mathematics for all third-through-eighth grade public school students in America, enjoyed the backing of a broad coalition of politicians, including strange political bedfellows George W. Bush and Edward Kennedy. Prince (2001) flatly stated that there is bipartisan support for annual testing.

The watchword for current federal education policy is “accountability.” *Education Week* has reported extensively on President Bush’s education policy proposals and has stated that the President “plans to hold schools accountable for federal funding and to judge schools’ performance based on annual testing.” (Johnston, 1999). The Center on Education Policy (Jennings, 2000), in a pre-election analysis of the major parties’ education policies, stated that both Democrats and the GOP are committed to “standards-based reforms,” “aligned assessments.” and accountability systems based on those assessments (and that Republicans at the national level have abandoned vouchers as a route for school reform, at least for now). That is, federal policy will support use of standardized tests and state-level test scores to define school success and failure.

Again, support for this approach is bipartisan: a 1999 “Pre-election Preview” of presidential candidates’ preferences on education, appearing in *The Education Digest*, stated that both George W. Bush and (eventual

losing Democratic presidential candidate) Al Gore both emphasized “the need for more accountability”. The Bush administration's own published policy paper outlining the President’s education proposals specifically equates “setting high standards,” “progress,” and “achievement” to student test scores (President Bush’s education policy, 2001).

Democratic support pre-dates Al Gore: One of former President Bill Clinton’s education advisors wrote late in Clinton’s second term that “The first principal of Clinton’s ‘Call to Action for America’s Schools’ is a national crusade for education standards and an effort to develop national tests of student achievement in reading and math...” (MacEwan, 1997, p. 2).

Lewis (2000) acknowledged that greater national involvement in education is what the public wants, but offered a negative summation of federal policy: “Increased federal funding and further demands for accountability from the federal government, unaccompanied by substantive debates about the benefits and drawbacks of a realignment of governance will once more abandon the nation’s public schools to a silver bullet” (p. 184).

State governors generally support the high-stakes testing and accountability themes of the federal government. Hoff (2000) reported that many governors are pushing for the National Assessment of Educational Progress (NAEP) to be administered more frequently, and some are prodding federal agencies to help make that happen. In a study of the history of the NAEP, Epstein (1998) stated that the NAEP was originally designed in the 1960s to be a neutral, non-political census-like measure of educational attainment, and that many educators first supported it because of the promise of political safeguards. Now, Epstein claimed, the NAEP has been thoroughly politicized, and that it is the

precursor to a national test and a national curriculum (as opposed to the potpourri of testing currently allowed by the new education bill.)¹⁰ Indeed, a committee convened by Congress during the Clinton administration warned that the various standardized and state-level tests now in place across the country would not work as a substitute for a national test (American School Board, 1998).¹¹

The “given” status of standardized and state-level tests as elements of accountability systems was stated by Hill (1999):

A number of states’ primary indicators of student academic achievement for the public schools accountability system are standardized achievement tests. Though the factors that affect student achievement are complex, it is generally held that such tests potentially inform all educational stakeholders of how effective public schools are at meeting the academic needs of all children served (p. 30).

Dorn (1998) offered that it is a given that high-stakes testing and accountability systems are what the public wants and politicians at all levels intend to deliver. “Politically popular school accountability systems in many states already revolve around statistical results of testing in high-stakes environments. The future of high stakes testing thus does not depend on what happens on Capitol Hill” (p. 1), and “Alternatives to standardized testing which do not ...also connect schooling with the public at large will not be politically viable” (p. 1). Carter and Cunningham

¹⁰ Abedi and Lord, 2001, reported that students in low-level and average mathematics classes, ELL, and low SES students did significantly better on the mathematics portion of the NAEP when test items were “linguistically modified,” providing a warning that the scores of non-linguistically-modified NAEP tests may produce lower scores for those students.

¹¹ The recently passed education bill does not mandate a national test.

(1997) put it bluntly, stating that, although true excellence takes time, the public is demanding that test scores be up now. A report on high-stakes testing in *Education Week* (Viadero, 1998) noted that

State and national policy makers [sic]--including President Clinton--have pointed to such tests as a way to improve student learning. In addition...Quality Counts, a report card on all 50 states' efforts to improve education published this month by Education Week, also awards points to states that have such testing programs in place (p. 1).

In Michigan, state-level testing continues to be strongly supported by the current governor and by the Michigan Department of Education and the Treasury Department (each of which is responsible for various elements of the MEAP program in the state.) A trend in the late 1990s was for many Michigan high school students to skip taking the MEAP tests, with many students (and some of their parents) stating that they questioned the value of the tests. As recently as 1997, about two-thirds of the students eligible to take the high school MEAP tests in affluent Bloomfield Hills, Michigan, simply did not take the tests (Johnston, 1997a). Michigan's governor, John Engler, reacted with anger at such boycotts, stating in 1997 that he was mad over apparent efforts to undermine the MEAP program. The governor restated his belief that the tests helped to make districts more accountable (Johnston, 1997b). In reaction to the situation, the Michigan School Board issued a warning to schools and districts that only students with disabilities or limited English ability should be excused from the MEAP tests (Johnston, 1997c).

There is no shortage of criticism of current and recent federal education policy, or at least warnings regarding apparent federal

premises. Eccles' 1999 study of federal policy concluded that contemporary federal education policy rests on three strongly-held assumptions:

1. Educational success can best be measured by traditional, objective test scores.¹²
2. The most appropriate tests are those concentrating on mathematics and science.
3. The major value of education in America is to enhance economic competitiveness.

Eccles claimed that while some aspects of federal policy were helpful (in impacting graduation rates, attendance, overall school funding and more equitable finance), federal policy over the last several decades tended to have negative social effects. The author concluded that federal policy has not helped to promote traditional democratic goals of building social capital, improving access to the political system, or promoting the general welfare. The author specifically states that the federal emphasis upon test scores and accompanying accountability schemes do not promote the ideals of Jefferson, Mann, or Dewey.

Finn (1998) went further:

The assumptions and strategies underlying most contemporary federal interventions in K-12 education are not just archaic and ineffectual but also damaging. The damage is caused by three large errors in federal policy for K-12 education: a grave misreading of current fundamental problems, the selection of the wrong client, and the endless cycle of a failed reform paradigm (p. 52).

¹² A strong caution regarding the reliability of traditional standardized tests was put forth by Markel-Fox (1993). That study found that modifying standard reading tests by way of embedding the assessments within games resulted in significantly higher scores.

Metcalf (2002) put this into a political context by stating that “The Bush revolution in education is the culmination of a decade of educational reform spearheaded by conservatives and business leaders” (p. 1).

In 1977 a committee of the National Academy of Education (NAEd), chaired by Stephen Bailey, was charged by then-Assistant Secretary of Education Mary Berry to study “within-school behaviors” that might be contributing to declines in SAT scores in American public schools. The committee was also charged with investigating the supposedly “deplorable and declining quality of written composition” over the twentieth century (p. 1).

Many of the findings which the NAEd committee offered in 1977 mirror the policies of the current federal administration, and assumptions of that committee seem to be the same assumptions held by President Bush. That report may be seen to contain statements which, at first glance, are in contradiction of elements of current federal policy. For instance, the report stated there are “no easy substitutes for well-trained and committed teachers, principals who are helpful to and supportive of staff, a disciplined atmosphere in schools, improved pupil nutrition and health, and ultimately the raising of standards of living for the poor” (p. 67). It also stated that:

...the goal of increasing test scores per se is a trivial one, and could possibly become a pernicious one. That is, if the goal leads to coaching for the test or to efforts to “beat the test” it can pervert education. The goals of eradicating illiteracy and of increasing educational achievement generally--so long as these do not lead to an overriding preoccupation with test scores--are surely appropriate goals for the educational system (p. 8).

Furthermore, the NAEd report stated that “If social policy increases the trauma associated with educational sorting devices by increasing the stigmas of scholastic failure, more is at stake for the society at large than the elevation of average standards” (p. 8).

Yet, the report offered that “...*a series of standardized tests at the lower grade levels used for diagnosing individual student weaknesses, pinpointing remediation needs, and building public pressures if school-wide performances in basic skills continue over time to be consistently low, could be positive influences on student learning*” (p. 9).

Moreover, the report claimed that “Unfortunately, there is not today a situation in which the public seems willing to trust the judgments of teachers and principals in certifying their children's' competence....So, tests are needed” (p. 10). Other sources certainly concur with the report's assertion that “There is a growing emphasis on holding the schools responsible for what and how well students learn” (p. 11). Yet, the report concludes with the somewhat startling statement that “...there is a general agreement that they (standardized tests) are testing the right kinds of things” (p. 10).

A report issued by the Center on Education Policy and the American Youth Policy Forum (2000) indirectly challenged many of the assumptions of the NAEd Report, while offering that the current popular view of U.S. public schools is that they are failing: “Most people believe American public schools are failing. Such views are not surprising, since critics of public education emphasize only what's wrong with the public schools, and negative stories about education appear frequently in the media” (p. 5). The report went on to list “what's right” with American education, asserting that the list of findings contained only those “supported by objective national data banks, such as the National Center

for Education Statistics” (p. 5). The entries on the list that are pertinent to public education, high-stakes testing, and accountability policy include the following.

1. American high school students are taking more challenging courses.
2. High school students are taking tougher mathematics and science courses.
3. Student mathematics and science achievement is improving.
4. SAT and ACT scores increased during the 1990s.
5. American teachers have stronger backgrounds in academic disciplines than critics have asserted.

It is difficult to state with certainty why policy-makers may ignore research findings, but Price offered insight (1990):

Education, unlike other professions, is forever saddled with what has been called by various writers the “everyone went to fourth grade” syndrome, that is, common sense tells us what needs to be done...research findings became but one of the many inputs that compete for influence, and other inputs have legitimacy as well...What seems clear is that current methods for the dissemination of important research findings are inadequate and often do not make a compelling case to the practitioner...(p. 3).

Price illustrated his point via the issue of technology and the schools:

School administrators throughout the United States are currently spending huge sums of money on technology (primarily computer hardware and software) in the belief that the use of technology will improve student

achievement levels and better prepare students to compete in a world economy. These policy decisions, which involve huge cost, are driven by intuition (in addition to strong political forces and corporate interests) as research findings have found few, if any, statistically significant links between school achievement levels and the use of technology (p. 2).

A specific warning concerning standardized testing was offered by Illig (1998), who noted that early formal testing may be developmentally inappropriate. Rotberg (2000) offered a warning to President Bush before passage of the new education bill: "Poverty and associated societal factors are the most significant contributors to low student achievement" (p. 72). (More on this in this chapter's Socioeconomic Factors section.)

Dorn (1998) warned that the consequences of statistical accountability systems will be the narrowing of purpose for schools, impatience with reform, and the continuing erosion of support for public schools, stating that high-stakes testing "creates perverse incentives against good teaching" (p. 1).

Tennessee has been a leader in the reporting and use of statistical school measures. The title of Bobbett, French, and Achilles' 1995 study of Tennessee's "District Report Cards"--*District-level Report Cards are More Smoke than Fire*--summarized the authors' conclusions. Quoting from the abstract:

A preliminary analysis of the variables on Tennessee's "District Report Card" (1988-89) suggests that they do not represent well what influences student outcomes in that state. As "Report Cards" seem to be in vogue as a way to "measure" school districts--in some states they are the

basis for “warning” and take-over plans by state agencies--
results of this study might impel more careful analyses of
the proposed policy (p. 1)

The authors challenged policy-makers “who are contemplating, creating or modifying report cards on schools” similar to those in Tennessee to use caution, and offered that the data contained in such report cards may be no more than “interesting” (p. 15).

It may be reasonable to assume that President Bush brings to Washington the views on education, high-stakes testing, and accountability that he held while serving as governor of Texas. The cornerstone of his state policy was the TAAS: the Texas Assessment of Academic Skills, state-level mandatory tests.

Haney (2000), (quoted by Achilles, et al., 2001), stated that although the TAAS has been frequently credited with dramatically raising the level of student achievement in the state, the so-called Texas miracle may be more myth than real. They noted that the percentages of students passing all three parts of the TAAS in recent years has shown a sharp and consistent decline, especially among minority students, and that much of the supposed increase in other student scores is due to the high dropout rate among minority students. Achilles, Finn, Gerber and PannoZZo (2001) warned policy-makers that they are ill-advised to use TAAS results as the building block of federal policy (instead focusing on consistent small class size as a better way to close the minority/majority academic achievement gap.) Anderson (2001) charged that the TAAS program “calls for closing the education gap while failing to problematize the growing social and economic gaps. The language of ‘no excuses’ and ‘zero tolerance’ of the current accountability movement seems to silence any discussion of... ‘savage inequalities’” (p. 325). Anderson (2001) continued

the indictment of the political right on the issue of school reform:

...education reformers have throughout recent history colluded in “blaming schools for not solving problems beyond their reach”...it is hard to understand why they (on the right) so thoroughly ignore the growing lack of social equity in the United States that forms the broader context for the Texas reforms...The problem I have with strategically supporting the Texas reforms is that we are ultimately supporting a conservative corporate agenda--one that heavily lobbies and funds both major political parties and one that does not have the welfare of poor and minority children at its core. If the welfare of poor and minority children has suddenly appeared on the radar screens of corporate American, it is because they are viewed as potential human capital (pp. 326-7).

A similar, skeptical view of the TAAS and the agenda of its supporters came from Shirley (1997), who observed that current Texas school reform seems to be a classic case of social engineering directed from sources outside the children’s communities.

The organization FairTest (which generally opposes standardized testing) has offered numerous criticisms of current high-stakes testing programs at all levels. This includes the recent claim that students in states that require them to pass a test in order to graduate from high school tend to fare worse on the NAEP than other students, according to FairTest’s own analysis (Viadero, 1995). Commenting on that report in *Education Week*, then-FairTest acting Executive Director Monty Neill stated that “What we’re saying and what this shows is, ‘Where is the evidence that it (high-stakes testing) does in fact make things better?’”

(p. 1). Echoing that sentiment is Clark Kerr, President Emeritus of the University of California, as quoted by Berliner and Biddle (1995): “Seldom in the course of policymaking in the U.S. have so many firm convictions held by so many been based on so little convincing proof” (p. 2).

Kohn (2000) has charged that high-stakes testing is deeply unfair, especially in that it fails to examine the more important skills and behaviors.

Along with Kohn and Spady, Berliner has been one of the more vocal opponents of high-stakes student testing. He has stated outright (1995) that it is a myth that student achievement in America has declined, that such myths lead to poor ideas for educational reform, and that “Evidence suggests that many of these (high-stakes testing) programs are not merely foolish, but that, if enacted, most would seriously damage America’s students, educators, or school programs” (p. 7). Likewise Tate (1997), in a review of quantitative studies on trends in mathematics achievement, especially those that focused on SES, gender, and language proficiency issues, found that basic mathematics skills had increased for all groups in the United States over the previous 15 years, and that the white/nonwhite gap was at least slowly closing.

Spady (2001) offered a scathing and pointed, if rhetorical, summation of the entire standards and standardized testing themes of current state and federal education policy:

They may get their version of improvement in the form of *higher test scores*, but they’ll never get *change*. Only the *rhetoric* of this movement is about change....when you consider that some states are expecting *every student* to meet *all* of them (standards) in *every subject*, they cease to be ideals at all. *These standards ignore the greater*

realities of life and living as creative, contributing human beings. Standards are about mastering basic skills and specific academic content, but there's very little of that content which directly enhances the complex competencies adults need in their career, civic, and family life roles. Traditional curriculum content relates to only a couple of the multiple intelligences....and hardly any of the domains of living, future conditions, or life performance abilities...*These standards cost a fortune...These standards are hypocritical*, simply because almost no well-educated, successful adult knows or can do them-including teachers (p. 80, 86).

Spady's current attacks are no match for Karier's premise, offered more than a generation ago, that standardized and I.Q. tests are designed to maintain social order and control lower classes (1972). This view was echoed by Carter et al. (1997), who found that teachers tend to see state-level tests as threats to educators--as evaluative, quality-controlling devices, rather than as aids to curriculum and instruction.

Missing from the debate over the use of high-stakes tests and accountability systems is a substantive public discussion of the appropriateness of multiple choice questions on such tests. (Multiple choice questions are the item type of choice, in spite of occasional attempts at more authentic assessment.) While educational psychology experts and educators generally see multiple choice items as inadequate for assessing student knowledge and skills, it may be that the dramatically lower cost of using multiple choice items (over, say, open-ended or constructed-response questions) has stifled that debate (Lawton, 1996).

In summary, federal and state-level policy-makers generally favor the use of high-stakes student testing and accompanying “accountability” plans.

State-Level Testing

Many states now administer their own achievement tests to students. It is instructive to examine what some education experts have had to say about those testing programs, and to examine the limited number of research studies that have focused on state tests. In that President Bush was a chief supporter of the Texas version of such tests, and, as previously noted, brings to Washington a standards-based, pro-standardized test, accountability stance, the Texas Assessment of Academic Skills (TAAS) receives here special consideration, as does the Michigan Education Assessment Program (MEAP).

It has already been mentioned in this chapter that the general public and federal and state-level policymakers tend to support high-stakes testing and accountability systems. This is in spite of a lack of convincing evidence that such programs increase student learning. Pedroza (1997) studied the impact of high-stakes testing on disenfranchised students, and found “generally negative consequences,” and little change in patterns of achievement due to the testing. Valenzuela (2000) and McNeil (2000) offered a longer list of concerns regarding the TAAS in particular. They charged (among other things) that the TAAS:

1. Reduces the quality and quantity of the curriculum
2. Re-directs district dollars toward test-preparation instead of

curricular instruction

3. Provokes instruction aimed at low levels of information
4. Violates what we know about how children learn
5. Imposes exit measures inappropriate for Limited English Proficiency students
6. Widens the achievement gap between majority and minority children.

“Critics of the critics” have offered their own warnings to non-supporters of the TAAS. Trueba (2001) summarized those warnings:

At the risk of making politically incorrect statements, I think that some of the strongest positions against the TAAS are overly protective and can be interpreted as patronizing....The last thing they (poor students) need is overprotection...we all recognize that deficit thinking affects our judgment and actions in education. Many of us, educators, tend to think that students who fail in school have internal deficiencies (cognitive, emotional, motivational, social and cultural), and, as a consequence, they suffer clear dysfunctions that exonerate us from the obligation to teach them successfully....academicians have developed a complex process to justify deficit thinking: an inventory of deficits, a sophisticated explanation of such deficits, a predictive system for the occurrence and perpetuation of deficits, and a range of possible interventions...it (TAAS) has clearly improved dramatically the performance of minority and low-income children in Texas and has resulted in significant instructional changes in many school districts (p. 340).

While few authors defend the TAAS, Hinojosa (1997) found that certain in-school factors were significant in raising TAAS scores: strong building-level instructional leadership, and the collaboration of staff, administration, and parents. It may be that the best interpretation of Hinojosa's findings is that leadership and collaboration are good things, not necessarily that he has provided a list of "TAAS-score-raisers."

Zlatos (1995) claimed that some school systems across the country --including Texas--are now excluding low-achieving students to artificially inflate test scores, and that cross-district, cross-state, and longitudinal comparisons are therefore invalid. Young (1994) found that any motivational effect of high stakes testing is limited to suburban students, thus calling into question the wisdom of using such tests as means to improve low-achieving students' scores.

There are significant manifestations of high-stakes testing and accountability programs in the various states. The State School Board of Kentucky has decided to use norm-referenced test results in the formula for determining school rewards and sanctions (although that Board has also stated that such scores would be only a part of a school or district's total score.) (Hoff, 1998).

Similarly, North Carolina has instituted the "ABCs of Public Education," a high-stakes accountability program designed also to "focus curricula on the basics." Program literature contains the dubious claim that the ABCs will "increase local control." Yarbrough (1999) studied teacher perceptions of the program and found them to be generally negative, with perceptions varying somewhat by community, SES factors, and percent of minorities.

Virginia's statewide tests spotlight the difficulty of standards-setting: 97% of public schools "failed" the first round of tests in 1998

(Saks, 1999). Issues of content, instrument quality, and relevance have provoked an unusual public outcry against those tests. Issues of reliability have arisen in Rhode Island as well, based on widespread allegations of “testing irregularities” on the state’s tests. A special commission appointed by the Rhode Island Education Commissioner found that more than 50 teachers in 21 different school districts used actual copies of the state test as practice exams with students.¹³

At least one study has found that teachers have made positive changes in their teaching practices in response to high-stakes state testing. Hanley (2000) found that teachers in Massachusetts tended to use more open-ended student questioning and writing prompts in direct response to the institution of state-level testing. Hanley also found that teachers tended to focus more instruction on the specific standards of the Massachusetts Curricular Framework, upon which the state test is based (although that could be either a positive or a negative development, depending upon the quality and relevance of the Framework.)¹⁴

Bucking the trend toward rigid use of paper-and-pencil, standardized test-like instruments, Vermont has for several years been a leader in use of alternative assessments for mandated state testing. Vermont is one of only a few states using or considering using student portfolios as one indicator of academic achievement (Manzo, 1996).

It has been noted that practitioners (teachers and administrators) tend to have negative views of standardized tests, similar state-level tests, and accompanying accountability systems. Michigan educators are no

¹³ This raises the entire question of whether cheating on high-stakes testing is inevitable. This dissertation does not address the issue, but the author recognizes it as important.

¹⁴ Although not a student test, Hanley, Fowler, Wheelock, Bebell and Marlec (1999) reported that an ad hoc committee of nationally known education experts examined and rated the Massachusetts Teacher Tests as highly unreliable, full of questionable content. The panel recommended to the Massachusetts Board of Education that the test (used to qualify teachers for their positions) be suspended immediately.

exception. Ray-Taylor (1991) studied Michigan teachers' descriptions of their use and interpretation of MEAP data and found that they were frustrated with pressure to improve scores. Teachers reported that they felt there were few guidelines offered as to how improve test scores, that they had too little time to teach the curriculum and to prepare students for the MEAP tests, and that testing was inappropriately leading curriculum reform in Michigan.

A more recent incident in Michigan might indicate that such frustrations still exist. In the spring of 2001, the Michigan Department of Education (MDE) released a list of 71 public schools whose winter 2001 MEAP scores (mostly fifth grade social studies) were seen to have, according to the MDE, "irregularities." One such school was Hornung Elementary School in Brighton, Michigan. The responses of the district superintendent, David Pruneau, and Brighton's State Representative, Lynn Scranton (although a politician and not an educator), provide the flavor of local views regarding the allegations (which turned out to be the product of clerical error). In a letter to district parents, Pruneau offered:

Hornung Elementary School was named in a press release by the State MEAP office and the Department of Education as one of those schools with testing irregularities. We received no further information from the department prior to the press release...Needless to say, I am indignant that Hornung Elementary would be named in the press by the department without clear confirmation and proof. The Brighton staff works extremely hard and is proud of the reputation we have for integrity and excellence in education. Unfortunately, the MEAP office has damaged

that reputation due to what appears to be incompetence on their part (Pruneau, 2001, p. 1).

A press release from the office of State Representative Judith Scranton was stronger yet in its outrage:

I am appalled and horrified at the damage the Michigan Merit Office has done to the Brighton Area Schools...The accusation of cheating is bad enough. Not having the facts to support the accusation is horrendous...Because of the high-handed, arrogant manner in which this was handled, Hornung Elementary students, staff, and parents have had their reputation irreversibly damaged. It is unconscionable that a high-powered state agency has so little concern, and could so quickly judge before checking their facts...I am disgusted that the department has so little respect for the work any school district does...I need to know why. The public needs to know why. The Brighton area schools need to know why. Every school district in Michigan deserves to know why...An apology will never be enough (Scranton, 2001, p. 1).

In that the researcher of this study attempted to identify SES factors that may be helpful in predicting overall district high school scores on the MEAP tests, it was instructive to review the literature for earlier findings in this area. There are few studies that deal directly with the issue of SES impact on MEAP scores. The most significant study to date in this area is the Standard and Poor's Statewide Evaluation Services' "Statewide Insights" document, which offered demographic, financial, and academic information on all of Michigan's school districts. In addition to an extensive database, the report included some

correlational data that paired district academic achievement scores with several SES factors. The two most important findings offered in the first Standard and Poor's report for Michigan (relevant to this study) are these:

1. There is a strong relationship between wealth and student achievement on the MEAP tests.
2. Household income is often found to be a strong correlate of student MEAP results.

This study confirmed those Standard and Poor's findings (and used some of the data given in the S&P report to examine other correlations; more on this in the Methodology section.)

One older study (Makki, 1992) paralleled this author's current work, except that Makki attempted to determine predictor variables that would help to identify individual students likely to "fail" the MEAP in one large urban Michigan district. Makki's conclusion was that ethnicity was the factor with the greatest predictive value for low MEAP scores.

Several other sources have offered findings related to factors that may impact MEAP scores (individual student scores, school aggregate, or distinct aggregate scores) and are noteworthy. Snyder (1995) examined per-pupil funding rates compared with student achievement on MEAP tests and found only weak correlation coefficients between those variables. Tucker (1996) examined *groups* of schools with similar per pupil spending and found that among groups there were no significant differences in achievement. Tucker reported that, overall, as per-pupil funding increased, there was no corresponding increase in student MEAP achievement. There were all levels of achievement within groups, and districts with fewer dollars per student did as well on the MEAP tests as districts with more dollars.

Burde (1989) studied the impact of class size on student MEAP test achievement and found no significant relationship. (That study only examined 4th grade MEAP reading and mathematics test scores, and compared class sizes of 16-20 students and 30-34 students.) The MEAP test has undergone evolution since the time of Burde's study, possibly limiting its relevance.

Building principal "leadership factors" were identified by Anspaugh (1995) as contributing significantly to higher MEAP scores at the middle school level. One of the more interesting, and perhaps provocative, studies (Williams, 1993) found that Michigan high school students who underwent extensive "test sophistication training" (test preparation) scored significantly higher than did students who did not receive such training, calling into question the reliability of MEAP test scores (based on the assumption that good assessment instruments are minimally impacted by noncurricular student test preparation).

Finally, Farfsing (1988) looked for a significant correlation between school staffs' perceptions of themselves and their MEAP scores, but he failed to find one.

Beyond the new Standard and Poor's document, there are few studies that attempt to statistically correlate socioeconomic factors with individual, aggregate school-level or aggregate district-level MEAP scores.

Introductory Notes Concerning SES Definitions

Evidence suggests that there exist significant correlations between and among various socioeconomic factors (individual student, school, and district) and student achievement. Just as the term "high-stakes" has no universally accepted definition, so too with "socioeconomic factors." In

general, those factors may be defined as individual student or community characteristics, particularly those which are describable in social or economic terms. A listing of such factors might include the following (with some factors referring to individual students, some to school or community averages, and some to both):

Household income

Per capita income

Level of parents' education

Number of parents and/or adults present at home

Home value

Location (urban, rural, suburban, small town)

Student transiency rate

Percent of students receiving free or reduced cost lunches at
school

Minority student percent

Multicultural student percentage

Percent college-bound

Dropout rate

Per-pupil spending

Class size

Teacher salary

The list leaves room for debate. For instance, how important is "home value," when both mean and median community home values vary drastically according to local cost of living? That is, might a median home value of \$100,000 mean two very different things in two different communities? How is class size defined? If the ratio of newer teachers-to-veteran teachers in a district can vary, how helpful is "teacher salary"?

Even if experts were to agree on the appropriate elements of a list of socioeconomic factors relevant to academic student achievement, there would remain much disagreement regarding definitions. For example, should “family income” be defined simply as “total gross household income” or should that figure be divided by the number of household residents in order to produce a more useful bit of data? Should “per-pupil spending” be based on total average local revenue, or should federal revenue be included? Instructional spending alone? Does considering “state homestead equalized value” provide more useful data than market prices?

It is not the researcher’s or the commentator’s task to produce a definitive list of socioeconomic factors; rather, he or she must take pains to provide the definitions for terms which appear in his or her work, and, where appropriate, to note why such definitions are used. It may be that useful existing data (from United State Census Bureau databases, for instance) require the researcher to accept definitions as provided by the source. Depending upon how a researcher defines a particular socioeconomic factor (or student achievement), the study may take on one or more delimitations, and it becomes the researcher's responsibility to note these.

In some of the works cited in this section, authors defined their terms. In others, terms are used with the apparent assumption that their meanings are obvious to the reader. This may limit the usefulness of some studies and articles.

*Socioeconomic Factors and Student Achievement*¹⁵

Various studies have examined the relationships between SES factors and student achievement. Some have considered individual student factors such as level of parent education, number of adults living in the student's home, family income, etc. Others focus on school attendance area, district or community factors such as *average* level of education, *average* number of adults living in students' home, *average* family income, etc. Any studies or articles that examined one or more SES factors at any level and compared them with student achievement levels, whether for individuals or for groups, were considered for this literature review.

Studies that examined the relationships between district per-pupil spending rates and student achievement have provided mixed findings. Tucker (1996) found that average teacher salaries offered some predictive value regarding student achievement, but that in general, an increase in district per-pupil funding did not correspond with increased MEAP scores. (Significantly, Tucker examined several categories of funding, including instructional spending per-pupil.) Earlier, Snyder (1995) found only weak correlations between average district teacher salaries and per-pupil funding and student MEAP scores. In Minnesota, however, Smith (1999) found a significant correlation between higher teacher salaries and students' state-level mathematics and reading test scores. While comparisons of different states' tests' results should be done with caution, Siminski (2000) reported that, in Massachusetts,

¹⁵ For purposes of this study, "student academic achievement" is defined as "achievement on standardized tests and standardized-like tests, such as state-level tests." Use of this definition does not imply that the author agrees with the proposition that such tests fully and accurately measure desired student academic outcomes.

increased per-pupil funding did correlate with significantly higher mathematics and reading state test scores.¹⁶

Those studies seem to confirm VanValkenburg (1990), who found that district per-pupil expenditures were not significantly related to student MEAP scores. However, that study found that, in heavily urban/minority population Wayne County, there were strong correlations between average district-level of adjusted gross income, average home value, and student performance on the MEAP tests at all levels.

Payne and Biddle (1999) claimed that “evidence supporting the effects of differences in funding levels on student achievement has proven elusive.” They continued:

Many studies have reported positive correlations between school funding and student achievement...But since level of advantage in students' homes is also related to student achievement, and (in the United States) home advantage is also tied to level of school funding, one must always control for the effect of home advantage when estimating the net impact of school funding. Prior studies involving such controls have generated confusing findings. Some studies have indeed found significant net effects for funding, but many others have reported only non-significant results (p. 4).

Using federal definitions of poverty, the U.S. Department of Education's National Center for Education statistics (1996) examined the nation-wide, school-level relationship of student poverty to standardized

¹⁶ Adams' 1994 study of the relationship between several district SES factors and SAT scores found that low per capita income was associated with a decline in mean SAT scores. Increased performance on SAT was associated with increased instructional per pupil spending, but not general fund expenditures per pupil. Also, a high “at-risk index “ presence was associated with lower SAT scores, with poverty as a significant variable.

test scores, finding that high levels of poverty (as well as urban location) were strong predictors of low student standardized test performance. Using data from the 1990 U.S. Census, the CORE of Educational Data, and the National Education Longitudinal Survey (NELS), Roscigno (1996) confirmed that poverty depresses general educational achievement, with family income level and family structure (i.e., number of adults) identified as the most significant variables. Smith (1999), noted above, also found that poverty, ethnicity, and the presence of student developmental disabilities were three strong predictors of student achievement on Minnesota's state-level tests. In a review of several studies, Viadero reported that, generally, elementary students in areas of high poverty lag behind their wealthier counterparts on measures of academic achievement.

Kim (1992) claimed that previous studies had noted many correlations between indicators of poverty and low academic achievement, and attempted to reproduce such results. Kim found that children's time spent in poverty is the chief predictor of academic achievement, explaining that much of the effect can be attributed to the children's birth weight, mother's cognitive skill, and the home environment.

In New York, Hendrie (1999) found that the state's poorest districts, particularly in urban and rural areas, tended to fare most poorly on the state literacy tests. Grinion (1999) examined the relationship between family SES (free/reduced lunch) and achievement on 11th grade standardized tests in Jefferson County, Kentucky, and found that poverty was significantly correlated with lower achievement (and that poverty had a greater impact on boys than on girls).

The Scholastic Aptitude Test (now called simply the SAT) is a standardized test long used by high schools and higher education admissions offices to predict student success in college. Adams (1994) looked for relationships between mean school district SAT scores and six variables: operating expenditures per pupil, instructional expenditures per pupil, per capita income, district enrollment size, population density, and “at-risk index characteristics.” Adams’ most significant finding was that low per capita income was associated with lower SAT scores. In his conclusion, Adams contributed a powerful statement regarding public policy:

The data and the study strongly suggest that, if school authorities, legislators, private business and parents continue to use the SAT scores as a prime barometer and target for educational success, we should immediately begin to compensate dramatically for the at-risk and per capita income deficits in individual students and impacted schools, and maximize financial resources into proven classroom instructional strategies. (p. 40).

Student academic achievement in Texas, and on the TAAS in particular, have been the subjects of several studies. Again, the results of these studies are of interest in that President Bush brings his philosophical support of TAAS to Washington.

Polinard, Wrinkle, and Meier (1995) studied the relationship of poverty rates to TAAS achievement levels for Hispanic and African American students and concluded that higher community poverty rates meant lower student passing rates on that test. (The authors also found that minority students tended to be more successful on the TAAS when there was a higher percentage of minority teachers and leaders in the

school district). Hill (1999) confirmed that high district poverty rates were associated with lower student TAAS scores.

“Parent and teacher expectations,” while perhaps not SES factors per se, were used as the independent variables in a 1997 study (Gill). Strong parental and teacher expectations were found to be significantly associated with higher mathematics and reading scores in pre-school-to-third grade low-income and minority youngsters. The author suggested as the key implication of the study that programs and policies that promote high educational expectations among parents and teachers may contribute to such children’s academic success.

Similarly, Ainsworth-Darnell (1999) found that neighborhood characteristics were predictive of educational outcomes. (The predictive power of neighborhood characteristics depended upon the type of educational outcome being considered.) The presence of positive role models--at home or in the community--was found to be associated with higher achievement test scores (but unrelated, surprisingly, to student dropout rates). Over a decade ago, Keusenkothen (1990) found that certain family characteristics were positively correlated with higher eighth grade achievement on the California Test of Basic Skills (CTBS) in reading, language and mathematics. Specifically, Keusenkothen found that family cohesion, expressiveness, independence, intellectual-cultural orientation, and active-recreational orientation (as defined by the Family Environment Scale) were related to student achievement. Paralleling Gill, Keusenkothen stated that the implication of the study was that schools should promote programs which help to develop parenting skills.¹⁷

¹⁷ Such implications imply that causality between the independent and dependent variables has been established; and that is not so. It may be that an unidentified variable is responsible for both “quality family characteristics” and high student achievement, that is, the relationship between the variables in the cited studies may be *only* correlational. Nonetheless, strong correlation hints at causality.

Goldstein (1991) identified factors that are helpful in predicting on-time school completion (and, incidentally, early child-bearing). All of the following factors were found to be related to on-time graduation, in order of significance: poverty, race, frequency of residential moves, non-marital changes in family structure, head-of-family level of education, parents' occupational status, and time as a welfare recipient. While not technically "student academic achievement," it may be reasonable to assume that "on-time school completion" is related to student achievement.

At least one study examined several in-school factors to determine their relationships with student achievement on the MEAP tests (grade four and five; mathematics, reading, social studies, and science). Secumski-Kiligian (1993) found no significant correlations between teacher job satisfaction, principal's managerial style, and student MEAP achievement. However, this study examined only the two elementary schools in one large suburban district, limiting the generalizability of the results. Nonetheless, this study hints at the stronger influence of out-of-school SES factors on student academic achievement. Ferguson, in the well-respected *Harvard Journal on Legislation* (1991) stated flatly that about half of student achievement is accounted for by SES factors, and about half by teacher and school effectiveness.

Three Significant Reports

In this portion of the Socioeconomic Factors and Student Academic Achievement section, three reports are spotlighted.

The Center on Education Policy report, "It Takes More Than Testing: Closing the Achievement Gap" (2001), offered several

summaries of research relevant to this dissertation, including the following:

1. Public school quality has not declined over the past 25 years.
2. All subgroup standardized test scores are up over the last 25 years.
3. A wide gap between African American/Hispanic and white test scores remains.
4. That achievement gap is present before students begin school.
5. The achievement gap is not due to innate abilities, nor is it simply the result of biased test questions.
6. There is no simple explanation of the gap, but a variety of school, community and home factors seem to underlie the gap (p. 21).

The Center identified the following as factors that may contribute to the gap:

Curriculum and course taking patterns

Teacher qualifications

Teacher expectations

Resources disparities

Concentrations of low income and minority students in certain schools

School climate

Student performance anxiety

Peer pressure

Access to high quality pre-school

Effects of poverty

Legacy of discrimination

Home and community learning opportunities

Parenting practices and education (p. 21).

Based on those findings, the Center supported general raising of standards and accountability, but stated that “policy makers [*sic*] have been timid about providing the supports needed to close the gap,” and that

...policy makers [*sic*] are being irresponsible if they lead the public into thinking that testing and accountability alone will close the gap. Policy makers [*sic*] must be cautious in attaching consequential penalties to test results, and should closely monitor and quickly address any negative effects of high-stakes testing on minority students (p. iii).

The report offered possible strategies for closing the gap, including these:

Offering challenging curriculum and instruction

Keeping well qualified teachers

Setting high standards and accountability for subgroups

Assuring equitable resource distribution

Encouraging diverse school enrollments

Developing the knowledge and capacity for school improvement

Reducing class size

Extending learning opportunities

Expanding pre-school

Attending to research

Encouraging supportive, motivating home culture

Encouraging parent education and involvement

Supporting community academic activities

Improving social conditions (p. 21).

Cooley (1993), in a report issued under the auspices of Pennsylvania Educational Policy Studies and the the Educational Research Service, using data from Pennsylvania schools, calculated that more than 60% of “the variation in average district school performance (test scores) among these school districts can be explained by...three simple census factors” (p. 5). At-risk for low achievement are students who are

1. living in poverty
2. with a single parent
3. who is not a high school graduate.

Cooley commented on the three factors, stating that while all three factors make unique contributions to the prediction of student performance, “...if any of these three census predictors were dropped there would be a significant loss in the predictive power of the resulting multiple regression equation” (p. 5) (a finding of importance for this dissertation).

The author provided policy implications:

...if you rank order states on the basis of the difficulty of the educational task, you get about the same rank orders as are produced using the NAEP average proficiency for these states. Therefore, one can clearly not make inferences about the relative quality of the math programs in these 42 states when over 75 percent of the variation in the math means among these states can be explained by the nature of the populations being served by the schools in those states...A very large problem...is that there are educational systems that are doing a good job with difficult-to-educate students, but their successes go unrecognized and unrewarded when their unadjusted results are unfavorably

compared with systems that had the easy job....Not doing so results in invalid inferences...such comparisons (among district test scores) reveal little or nothing about differences in the quality or effectiveness of the educational programs that are represented by those results.

Cooley introduced a concept that he believed puts into perspective the entire accountability debate, stating that “What those differences in test results primarily reveal are differences in the difficulty of the educational task....Districts with lots of poverty children have a more difficult educational task...”

Payne and Biddle (1999) compared existing child poverty data (including school funding levels) from 205 classes in 67 school districts in 32 states with results of the SIMS (Second International Mathematics Study) for eighth graders. The authors summarized:

...our results suggest that level of school funding and child poverty have substantial and statistically significant net effects on average student achievement among the school districts of American and that these effects stand up even when juxtaposed with those of two crucial, district-level control variables: level of curricular instruction, and race (p. 11).

The study suggested, therefore, that both poor school funding and child poverty affect student academic achievement, and that such effects are largely independent of one another.

In their report, the authors also discussed the issues of poverty and test scores by way of comparing America’s school district’s SIMS score “standings” vs other countries. They found that “advantaged” U.S. school

districts scored second in the world (below Japan), that “disadvantaged” districts scored near the bottom of the ranking, above only Nigeria and Swaziland, and that the overall American ranking was significantly below the median.

The authors suggested an explanation for the relatively poor overall America scores that is relevant to this dissertation. They claimed that poorly funded schools and communities with high levels of poverty are rare in other developed countries, and that education in America is uniquely handicapped because of tolerance for poorly funded schools and child poverty. Specifically, the child poverty rate in America stands at between 20 and 25%, far above that of most other advanced nations. Further: child poverty tends to be high in families headed by a single mother, and America has a large single-mother family population. Payne and Biddle went on to claim that, therefore, none of the current reform efforts (vouchers, “higher standards,” accountability schemes, charters, etc.) are likely to improve achievement.

The authors stated that the U.S. differs from most other industrialized nations in that most school dollars come through local taxes and bonds, and that, naturally, funding levels vary tremendously. In other countries, they reported, there is generally even funding. How large are the gaps in America? From \$4000 to \$13,000, depending on where one lives, according to the authors.

Payne and Biddle quoted Rainwater and Smeeding (1995, p. 22):

All nations begin with the issue of preventing severe poverty and disadvantage for at-risk children. The major difference we note is that other nations are able to produce [much] lower levels of child poverty than in the United States...Our high-income children do very well in real terms compared to

similar children in other nations. No one wants to take away these advantages for which the parents of these children work long and hard. What is needed is a reasonable response to the real needs of low-income American children. And as other nations have shown, there is an answer that we can find if we have the national will to face up to the sobering facts...(p. 7).

Poor children are “uniquely handicapped” for education, stated the authors. They added that poor homes provide little access to books, writing materials, computers, or other supports; poor students are more often distracted by diseases; they tend to live in neighborhoods affected by crime, decay, drugs, and drug dealing; and their homes tend to be dysfunctional, with parents often incarcerated or disturbed. The authors concluded that poor children are likely to have a much harder time in school, and that aggregate achievements of American schools are likely to be lower than those from other countries where child poverty is low.

CHAPTER THREE

METHODOLOGY

Introduction

As stated in Chapter One, the problem under consideration in this study is that education policy-makers may be operating under a significantly false assumption. That assumption is that student, school, and district SES factors are not associated with composite district student test scores, and that, therefore, use of such scores for high-stakes purposes is appropriate, and that state and federal education policy may be reasonably based on that premise. The review of literature, summarized in Chapter Two, seemed to support this claim. The author of this study, then, attempted to determine, with limitations, if school district SES factors are indeed significantly correlated with composite district MEAP scores.

Design

The questions posed in Chapter One suggested a quantitative, correlational, multivariate/predictive study design. That is, relevant and available district SES factors (in numeric form) needed to be identified and subjected to statistical correlational tests. The goal was to determine if it was possible to develop an equation, wherein the variables consisted of various district SES factors, which would provide a reliable predictor of district MEAP scores (hence refuting the assumption that SES factors are not associated with district student test scores).

Specifically, the study's design called for a review of literature that would suggest a list of SES factors to be paired with district MEAP scores, with sources of that data identified. The creation of a spreadsheet which included various district SES factors and MEAP scores would then allow for statistical correlational testing. This was to include single SES factor/MEAP score correlational tests, using the Pearson Product-Moment statistic.

That test produces a number, a "correlational coefficient," which expresses the direction, shape, and strength of a relationship. If a coefficient is sufficiently distant from zero, it allows for reasonable prediction. (In this study, that means SES factors being predictive of MEAP scores, as defined.)

If significant correlations were discovered, the factors with which they were associated were to be combined by way of multiple regression. The goal of such an approach is to determine if some combination of factors (in this case, district SES factors) is *more* predictive of the independent variable (district MEAP score) than any single factor. Multiple regression also produces a new coefficient for each factor, which, along with a generated constant, can be used to produce a multivariate predictive equation.

Multiple regression testing also produces two important statistics: "R," which, as with simple bivariate correlations, simply identifies the strength of the combined factors in relation to the dependent variable, and "R²," which gives the percent of the variance of scores due to the influence of the combined factors.

Sample size for this study was near 100% for all identified factors. Variability was thereby decreased and validity increased. N (the entire population under consideration) was 522. Free and reduced lunch data

were available for all 522 districts. For all other factors, n (the actual number of data points available) was equal to 519¹⁸.

The data gathered would always produce the same correlational coefficients using the Pearson Product-Moment test, so study reliability is high.

Methods

Step One: Collection of Data

The review of literature suggested that the SES factors noted below should be the individual independent variables initially considered; these data were collected.¹⁹

1. Percent of District Students Eligible for Free or Reduced Price Lunches (1999)
2. State Equalized Homestead Valuation (SEV) Per State Aid Member (2000)
3. Percent of District Children Poor, Ages 5-17 (1997)
4. Percent of District Lone Parent Households (2000)
5. Mean District Household Income (2000)
6. Median District Household Income (2000)
7. Percent of District Households With Annual Income Under \$30,000 (2000)

(Dates noted represent the most recent available data.)

¹⁸ Data for Percent of District Lone Parent Households, Median and Mean Annual Household Income, Percent of District Households With Annual Incomes Under \$30,000 were missing from archival databases for the school districts of Hillman, Jonesville, and Lake City, for unknown reasons.

¹⁹ The list also reflects the availability of archival district-level SES data. Although other, possibly useful, SES data exist (median district housing price, for instance), they were not available, short of conducting a district-by-district search. Such a task was beyond the scope of this study.

All of the data were available via the World Wide Web. The sources were databases on the Michigan Department of Education website (for K-12 Enrollment and Percent of Students Qualifying for Free and Reduced Price Lunches, State Equalized Homestead Valuation Per State Aid Member, and Pupil to Teacher Ratio); the United States Census Bureau website (for Total District Population and Percent of Children Poor Ages 5-17), and the Standard and Poor's website (for Percent of Lone Parent Households with Children, Mean and Median Household Income, and Annual Income Under \$30,000.) The Standard and Poor's site also provided District Composite MEAP Passing Scores, with the social studies test scores excluded.

Step Two: Cleanup of Data

The various databases used did not always present data that were “clean” for purposes of this study. For example, as previously noted, the Michigan Department of Education considers Public School Academies (Charter Schools) school districts for reporting purposes, but such “districts” were excluded from this study. Additionally, several “nonunified” schools districts (typically K-5 and K-8) had to be identified and removed. This necessitated customizing a new, unique spreadsheet, so that only the traditional, “unified” public school districts remained. Thus, a customized spreadsheet was produced from all noted databases, with one column for each SES factor.²⁰

²⁰ All computer work was done on an iMac, using Microsoft Word 98 for word processing, Microsoft Excel 98 for spreadsheets, and SPSS 10.0 for statistical tests.

Step Three: Correlational Analysis

Next, individual correlation coefficients for each SES factor/district MEAP test scores relationship were generated using the Pearson Product-Moment within the SPSS computer statistical program. These coefficients indicated the level or strength of the relationship between each SES factor and the corresponding district MEAP score. SES factors found to have a correlational coefficient²¹ higher than 0.5 were arbitrarily considered to have significant value in the prediction of district high school MEAP scores, as defined previously (see Table 1 in Chapter Four).

Step Four: Examination of Scatterplots

The SPSS computer program allows researchers to produce scatterplots of paired factors (see “Definitions” in Chapter One). Scatterplots were produced for each of the SES factors under consideration, paired with district MEAP scores. The scatterplot of each relationship with a correlation coefficient of 0.5 or higher was examined to determine if any irregularities might disqualify the factor from further consideration as a quality indicator of the studied relationship.

Step Five: Multiple Regression

The SPSS program allows for the examination of several independent factors in relation to an identified dependent factor at once. As previously noted, this is called *multiple regression*, and this simultaneous examination of several variables can potentially generate an equation with stronger independent variable predictive power than any of the individual factors. Multiple regression allows researchers to produce

²¹ “R,” a number from -1 to +1, indicating the strength of each relationship, with R values more distant from zero indicating stronger relationships.

such multivariable equations wherein, when the several variables are replaced with specific values (in this study, numeric district SES factors), the independent variable is predicted. The higher (stronger) the multivariate coefficient (R), the more reliable the factors in predicting the value of the independent variable (in this study, district MEAP scores.) This procedure was used to produce such a predictive equation.

SES factors that were found to have significant correlational coefficients in relations to MEAP scores were combined in various configurations until the combination that produced the highest power of prediction was discovered.

Step Six: Drawing Conclusions (Chapter Five)

The study's design called for the statistical results to be analyzed, conclusions drawn, and possible inferences and policy implications made.

Methodology Notes

This study discovered the correlational coefficients between several district-level SES factors and composite district student MEAP scores. The study did not establish causality. However, as with all correlational studies, the discovery of strong correlational coefficients between independent and dependent variables suggests possible causality or the existence of intervening variables.

As noted in the introduction, the district level of analysis was chosen because most MEAP score reporting is done at the school district level, with published comparisons done at that level as well, and, as a practical matter, SES data are most widely available for the district level.

Furthermore, high school MEAP scores can be seen to some degree as summative, in that they are generated as students are ending their K-12 school careers. It was recognized that student transiency, which this study did not consider, represented a possible confounding factor.

For purposes of this study, “composite district MEAP scores” was defined as “the percent of district high school students passing the MEAP tests,” further defined as “the percent of district high school students taking the MEAP tests in the 2000-2001 school year who qualified for the Michigan Merit Scholarship Award.” Generally, students who qualified for this award were those who scored in the top two state-reported achievement categories (“Exceeded Michigan Standards” or “Met Michigan Standards”) for four high school MEAP tests (Science, Mathematics and Reading and Writing.) These students are popularly referred to as having “passed the MEAP,” although neither the Department of Education nor the MEAP Merit Award Office use the terms “passing” or “failing.”

The specific definition of “MEAP scores” chosen for this study was additionally significant in that the awarding of Michigan Merit Scholarships on the basis of MEAP scores is public policy. It is important, therefore, to emphasize that the correlations discovered in this study provide predictions of which Michigan K-12 public school districts tend to receive more of the scholarship dollars.

As noted above, the Michigan Department of Education, for reporting purposes, defines each of the approximately 150 state Public School Academies (Charter Schools) as individual school districts unto themselves. Because the nature of individual Charter Schools is different from that of traditional school districts, Charter School “districts” were

excluded from this study, and no inferences regarding their SES factors and MEAP test scores are valid.

Data on the Detroit Public Schools were excluded from this study. All other Michigan public school districts have no more than five high schools (and only four have four or five), while the Detroit district has 37. That number of high schools puts Detroit in a unique enrollment class. The 37 schools represent a far larger geographic area, with far greater SES diversity, than any other Michigan district, hence the exclusion.

In that the SES factors used in this study were district averages and percentages, and MEAP scores were only from high school MEAP tests, results of this study provide no predictive power for student or school-level SES-MEAP test score correlations. (See Bobbett, French and Achilles, 1993). Their exploration of the relationships among several Tennessee school district variables, and the relationship between each variable and average student test scores, concluded that the variables did not have the same impact at the elementary, middle school, high school and district levels).

The chosen methodological approach allowed this researcher to provide definitive answers to the questions posed in Chapter One.

CHAPTER FOUR

PRESENTATION AND ANALYSIS OF DATA

Introduction

In this chapter the results of the statistical analyses outlined in Chapter Three are presented, interpreted, and analyzed.

Results of the Pearson Product-Moment tests (PPM) for all identified SES factors are given in tabular and graphic (scatterplot) forms. Interpretations and analyses of the tables and scatterplots are provided.

Multiple regression results are given in table form; again, with interpretation and analysis provided.

As noted in Chapter Three, before applying various statistical tests to the data, they were organized into a single spreadsheet. Spreadsheet rows were labeled with the 522 unified (K-12) public school districts in Michigan. The columns of the spreadsheet were given these titles:

1. Michigan School Code Number
2. District Name
3. Total District Population
4. Total District Student Enrollment
5. Percent of District Students Eligible for Free or Reduced Priced Lunches
6. State Equalized Homestead Valuation (SEV) Per State Aid Member
7. District Pupil to Teacher Ratio
8. Percent of District Children Poor, Ages 5-17
9. Percent of District Lone Parent Households

10. Median Annual District Household Income

11. Mean Annual District Household Income

12. Percent of District Households With Annual Incomes Under
\$30,000

13. District High School MEAP Composite Score Without Social
Studies

Exact definitions of spreadsheet column labels are found in
Appendix B. The spreadsheet was too large for inclusion in this
document; Table 1 is provided as a reference sample.

Table 1
Sample of Data Spreadsheet

	name	totalpop	pupenrol	f_rlunch
1	ADAMS TOWNSHIP SCHOOL DISTRICT	2470	522	44.3
2	ADDISON COMMUNITY SCHOOLS	7479	1345	18.6
3	ADRIAN CITY SCHOOL DISTRICT	31293	4394	33.2
4	AIRPORT COMMUNITY SCHOOL DISTR	17175	2997	22.4
5	AKRON FAIRGROVE SCHOOLS	3523	506	41.6
6	ALBA PUBLIC SCHOOLS	1082	171	50.0
7	ALBION PUBLIC SCHOOLS	14268	2081	65.3
8	ALCONA COMMUNITY SCHOOLS	7164	1050	39.9
9	ALGONAC COMMUNITY SCHOOL DISTR	16678	2588	20.0
10	ALLEGAN PUBLIC SCHOOLS	16386	2996	26.1
11	ALLEN PARK PUBLIC SCHOOLS	22275	3245	9.6
12	ALLENDALE PUBLIC SCHOOL DISTRI	9626	1659	19.0
13	ALMA PUBLIC SCHOOLS	15196	2679	34.8
14	ALMONT COMMUNITY SCHOOLS	7043	1571	11.0
15	ALPENA PUBLIC SCHOOLS	31673	5552	33.6
16	ANCHOR BAY SCHOOL DISTRICT	26192	5047	16.4
17	ANN ARBOR PUBLIC SCHOOLS	145106	16276	14.6
18	ARENAC EASTERN SCHOOL DISTRICT	2557	477	47.8
19	ARMADA AREA SCHOOLS	7926	1781	6.4

Results of Bivariate Correlational Analysis

For this study, MEAP scores (as defined) were individually paired with the following factors in a PPM test:

1. Percent of District Students Eligible for Free or Reduced Priced Lunches
2. State Equalized Homestead Valuation (SEV) Per State Aid Member
3. District Pupil to Teacher Ratio
4. Percent of District Children Poor, Aged 5-17
5. Percent of District Lone Parent Households
6. Median Annual District Household Income
7. Mean Annual District Household Income
8. Percent of District Households With Annual Incomes Under \$30,000

Each correlation table identifies which two factors were paired in that PPM test. For example, it is noted in the left hand column of Table 2 that the dependent variable District High School MEAP Composite Score Without Social Studies was paired with the independent variable Percent of District Students Eligible for Free or Reduced Priced Lunches.

The table indicates several other things:

1. The Pearson Product-Moment was the statistical test used, and it produced a correlational coefficient (MEAP Score vs. Percent of District Students Eligible for Free or Reduced Price Lunches) of -0.701, a strong correlation. (The coefficients of 1.000 simply indicate that each factor correlated with itself perfectly.)
2. The “significance” of the statistical pairing was given by the SPSS program as 0.000, indicating that probability that the finding was due to

chance is near zero.²²

3. N indicates the number of data points for each factor used in the computation. Although there are 522 unified school districts in Michigan, data on the two factors (MEAP Scores and Percent of District Students Eligible for Free or Reduced Price Lunches) came from slightly fewer districts (519), due to missing data on the database used.

Table 2
MEAP Scores vs. Free/Reduced Lunch: Correlations

		MEAP Composite w/o Social Studies	% of Free and Reduced Lunch Eligibility
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	-.701*
	Sig. (2-tailed)	.	.000
	N	519	519
% of Free and Reduced Lunch Eligibility	Pearson Correlation	-.701*	1.000
	Sig. (2-tailed)	.000	.
	N	519	522

** . Correlation is significant at the 0.01 level (2-tailed).

²² This statistic is derived from the value of the correlational coefficient coupled with the sample size, with a strong coefficient and a larger sample size giving a significance nearer zero.

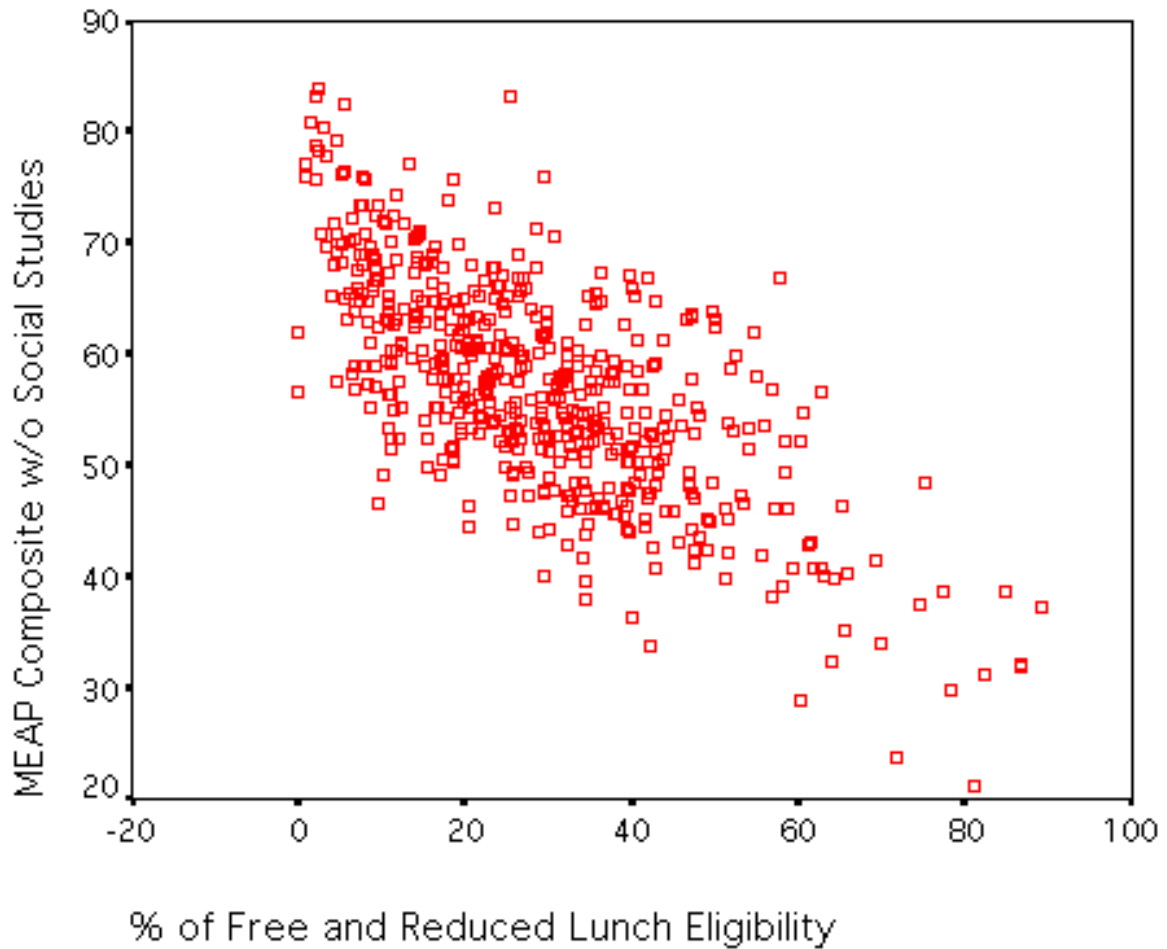


Figure 1. MEAP Scores vs. Free/Reduced Lunch: Scatterplot.

Each correlation table is accompanied by the scatterplot produced by that PPM test. The points on the scatterplot (Figure 1) represent the MEAP Scores (as defined) and Percent of District Students Eligible for Free or Reduced Price Lunches (as defined) for all Michigan unified school districts. There may appear to be fewer than 522 points on the scatterplot; this is due to point overlap.

The general oval shape formed by the points and the negative direction of the oval provide a graphic representation of the information

on the table, particularly the correlational coefficient of -0.701.

Summarizing the information in Table 2: There is a strong relationship (-0.701) between a district’s MEAP Score and its Percent of Students Eligible for Free or Reduced Priced Lunches. The sample size was near 100% (519 out of 522 for both factors), and the probability that the findings were due to chance is near zero.

On the following pages, the PPM test results and scatterplots for each pairing (MEAP score and SES factor) are presented and summarized.

Table 3 gives the results of pairing district MEAP Scores with State Equalized Homestead Valuation (SEV) Per State Aid Member (2000). The results indicate a moderately strong relationship (-0.512) between those factors. The sample size was near 100% (519 out of 522 for both factors), and the probability that the findings were due to chance is near zero. The scatterplot (Figure 2) indicates that several districts are “outliers”; that is, their ordered pairs (SEV Per Pupil, MEAP Score), are far outside the general oval formed by the other districts’ points.

Table 3
MEAP Scores vs. Total SEV Per Pupil: Correlations

		MEAP Composite w/o Social Studies	Total SEV per pupil
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	.512**
	Sig. (2-tailed)	.	.000
	N	519	519
Total SEV per pupil	Pearson Correlation	.512**	1.000
	Sig. (2-tailed)	.000	.
	N	519	522

** . Correlation is significant at the 0.01 level (2-tailed).

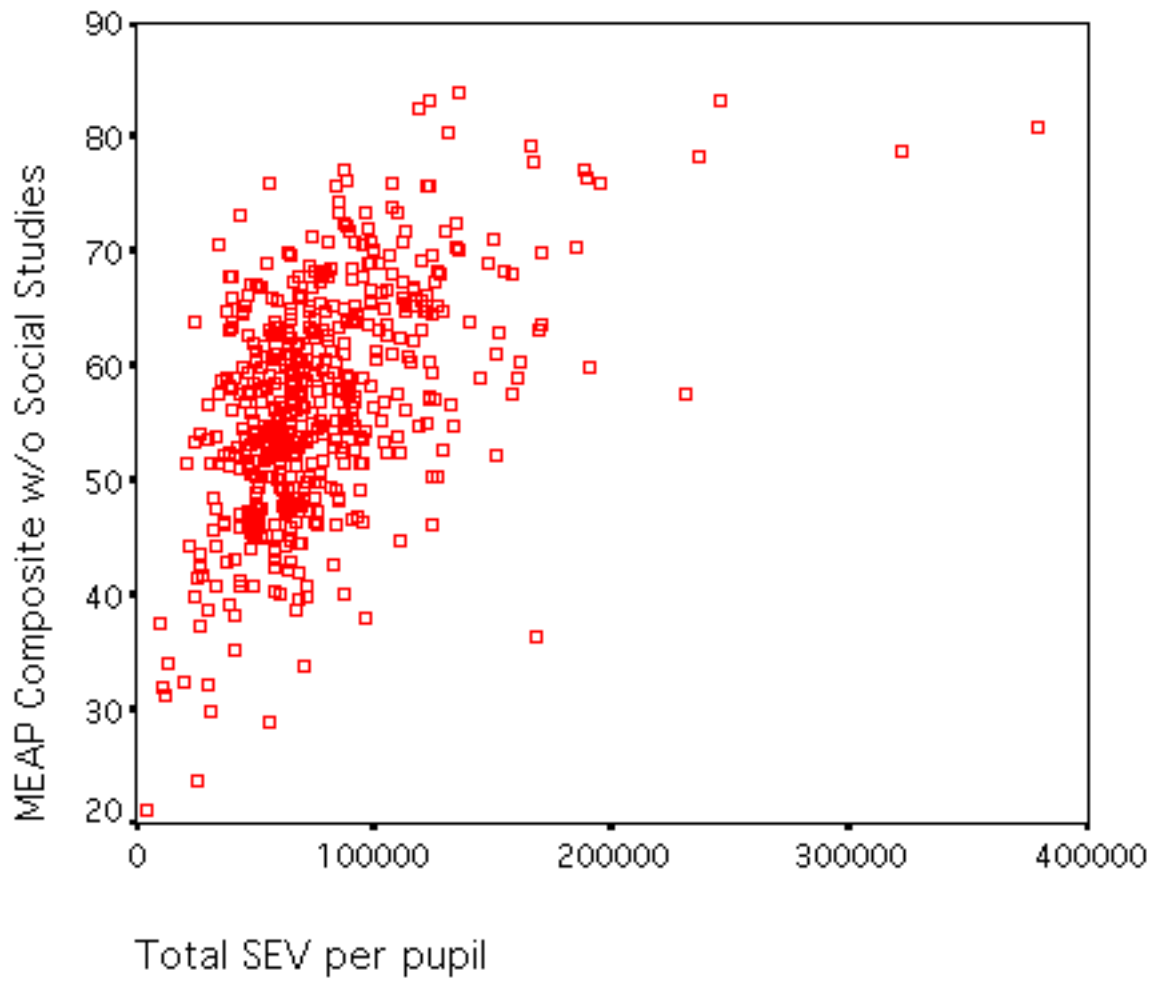


Figure 2. MEAP Scores vs. Total SEV Per Pupil: Scatterplot.

Table 4 gives the results of pairing district MEAP Scores with District Pupil to Teacher Ratio. (As noted in Appendix B, this statistic does not include support staff, but does include Special Education teachers.) The results indicate essentially no relationship (-0.088) between those factors. The sample size was near 100% (519 out of 522 for both factors), and the probability that the findings were due to chance is .044 (low). The somewhat circular shape formed by the points reflects the near-zero correlation of the factors. The appearance of vertical lines in the scatterplot (Figure 3) is due to the fact that Pupil to Teacher Ratio data were whole numbers, hence vertical lines at 20, 21, 22, etc.

Table 4
MEAP Scores vs. Pupil to Teacher Ratio: Correlations

		MEAP Composite w/o Social Studies	Pupil to Teacher Ratio
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	.088*
	Sig. (2-tailed)	.	.044
	N	519	519
Pupil to Teacher Ratio	Pearson Correlation	.088*	1.000
	Sig. (2-tailed)	.044	.
	N	519	522

*. Correlation is significant at the 0.05 level (2-tailed).

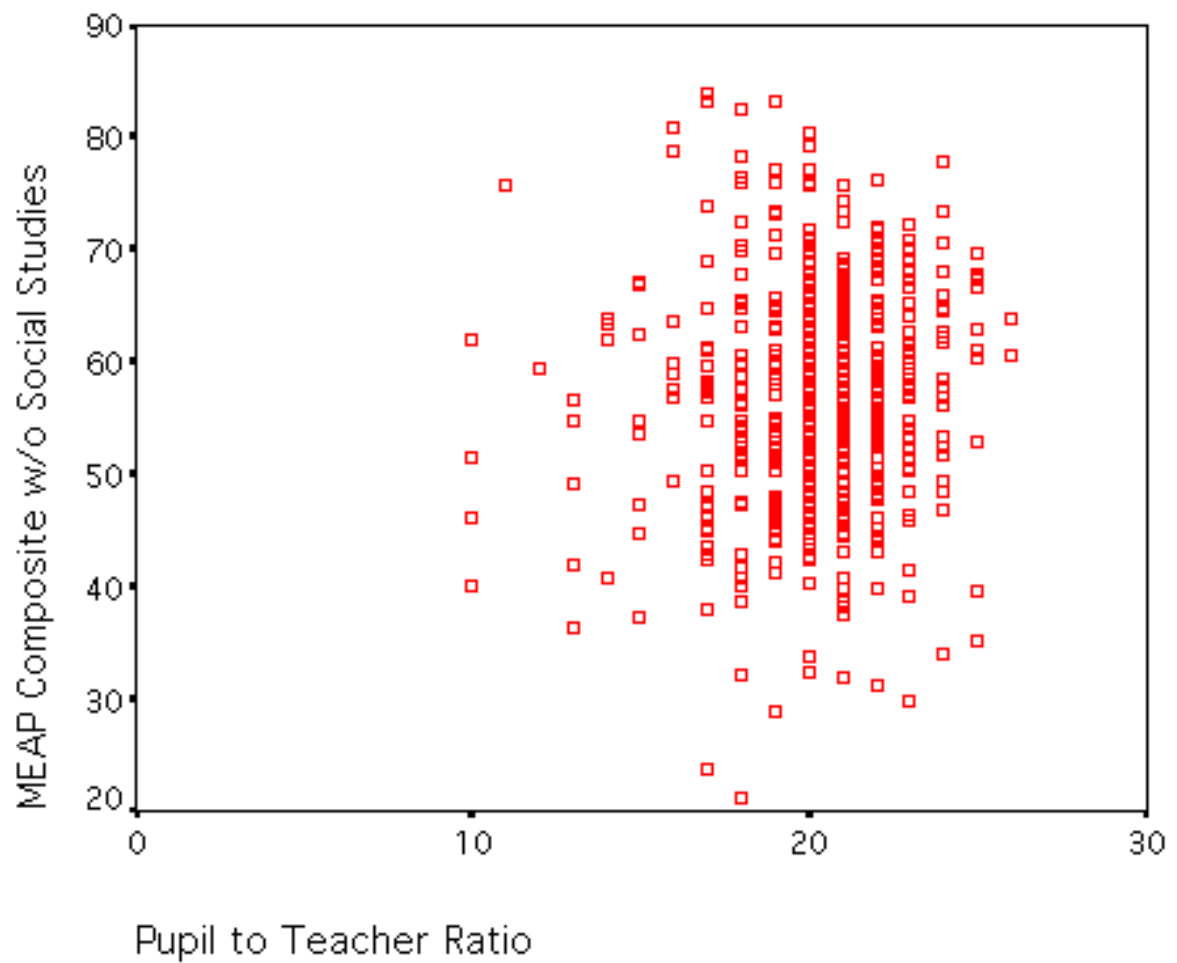


Figure 3. MEAP Scores vs. Pupil to Teacher Ratio: Scatterplot.

Table 5 gives the results of pairing district MEAP Scores with the Percent of District Children Poor, Ages 5-17. The results indicate a fairly strong relationship (-0.655) between those factors. The sample size was near 100% (519 out of 522 for both factors), and the probability that the findings were due to chance is near zero. The scatterplot (Figure 4) reflects the information from Table 5, with some outliers evident.

Table 5
MEAP Scores vs. Children Poor: Correlations

		MEAP Composite w/o Social Studies	% of poor children, ages 5-17
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	-.655*
	Sig. (2-tailed)	.	.000
	N	519	519
% of poor children, ages 5-17	Pearson Correlation	-.655*	1.000
	Sig. (2-tailed)	.000	.
	N	519	522

** . Correlation is significant at the 0.01 level (2-tailed).

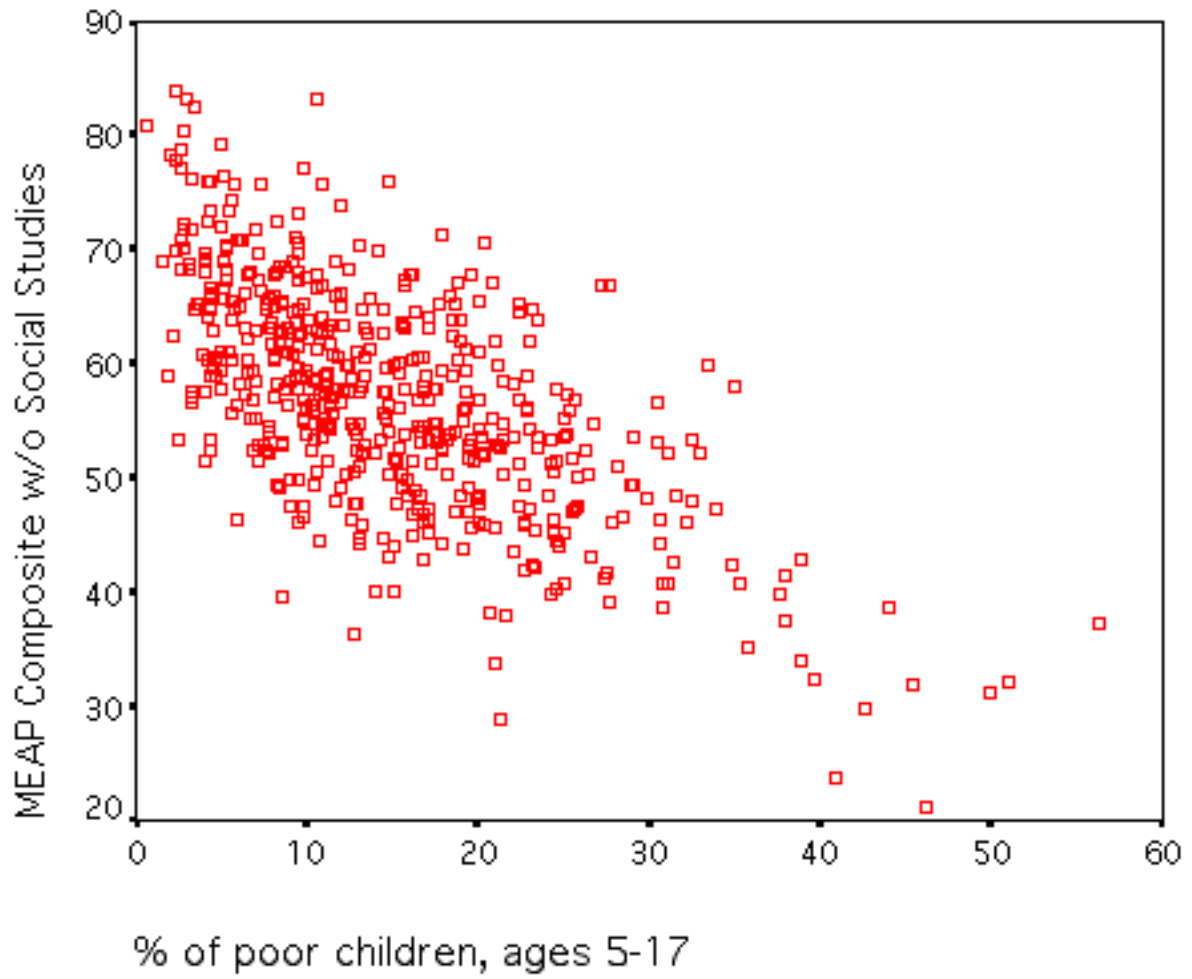


Figure 4. MEAP Scores vs. Children Poor: Scatterplot.

Table 6 gives the results of pairing district MEAP Scores with Percent of District Lone Parent Households. The results indicate a moderately strong relationship (-0.557) between those factors. The sample size was 100%, and the probability that the findings were due to chance is near zero. The scatterplot (Figure 5) reflects the information from the table, with some outliers evident.

Table 6
MEAP Scores vs. Lone Parent Households : Correlations

		MEAP Composite w/o Social Studies	% lone-parent households
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	-.557*
	Sig. (2-tailed)	.	.000
	N	519	519
% lone-parent households	Pearson Correlation	-.557*	1.000
	Sig. (2-tailed)	.000	.
	N	519	519

** . Correlation is significant at the 0.01 level (2-tailed).

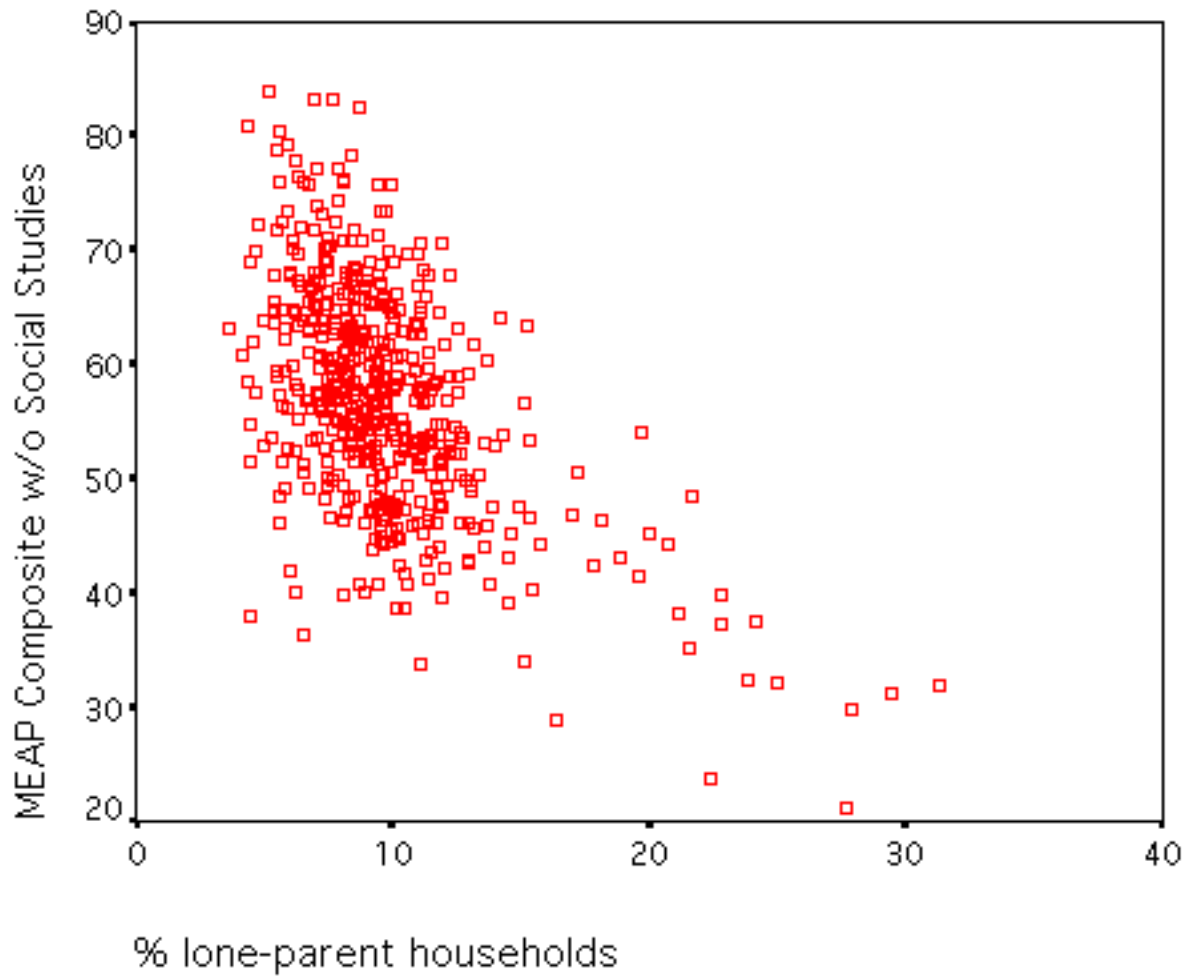


Figure 5. MEAP Scores vs. Lone Parent Households: Scatterplot.

Table 7 gives the results of pairing district MEAP Scores with Median Annual District Household Income. The results indicate a moderately strong relationship (0.530) between those factors. The sample size was 100%, and the probability that the findings were due to chance is near zero. The scatterplot (Figure 6) reflects the information from the table, with some outliers evident.

Table 7
MEAP Scores vs. Median Income: Correlations

		MEAP Composite w/o Social Studies	Median Income
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	.530*
	Sig. (2-tailed)	.	.000
	N	519	519
Median Income	Pearson Correlation	.530*	1.000
	Sig. (2-tailed)	.000	.
	N	519	519

** . Correlation is significant at the 0.01 level (2-tailed).

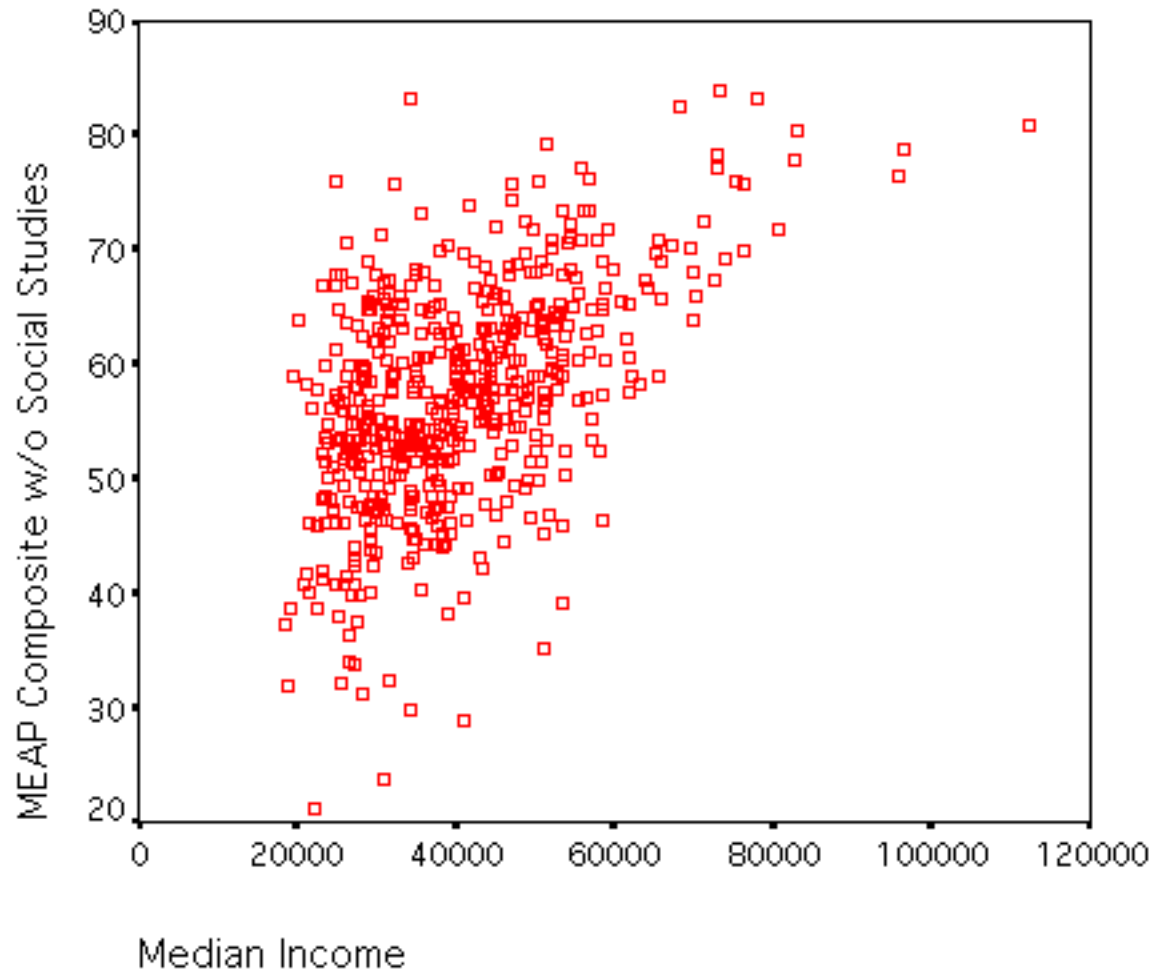


Figure 6. MEAP Scores vs. Median Income: Scatterplot.

Table 8 gives the results of pairing district MEAP Scores with Mean Annual District Household Income. The results indicate a moderately strong relationship (0.582) between those factors. The sample size was 100%, and the probability that the findings were due to chance is near zero. The scatterplot (Figure 7) reflects the information from the table, with few outliers evident. The most distant outlier, with near-\$200,000 mean income, represents the Bloomfield Hills School District.

Table 8
MEAP Scores vs. Mean Income: Correlations

		MEAP Composite w/o Social Studies	Mean Income
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	.582*
	Sig. (2-tailed)	.	.000
	N	519	519
Mean Income	Pearson Correlation	.582*	1.000
	Sig. (2-tailed)	.000	.
	N	519	519

** . Correlation is significant at the 0.01 level (2-tailed).

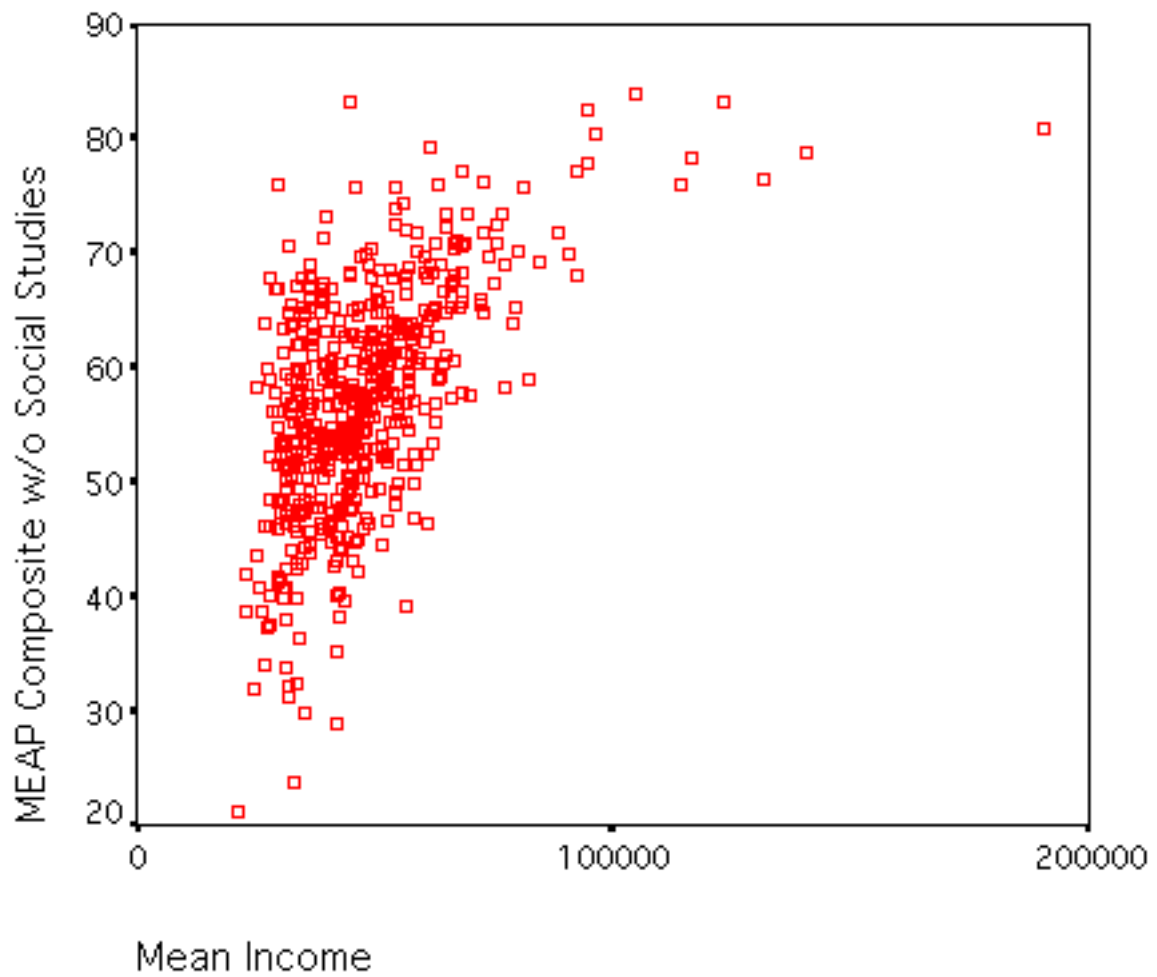


Figure 7. MEAP Scores vs. Mean Income: Scatterplot.

Table 9 gives the results of pairing district MEAP Scores with Percent of District Households With Annual Income Under \$30,000. The results indicate a moderately strong relationship (-0.494) between those factors. The sample size was 100%, and the probability that the findings were due to chance is near zero. As would be expected with a somewhat weaker correlational coefficient, the scatterplot (Figure 8) presents a less-defined oval.

Table 9
MEAP Scores vs. Income Under \$30,000: Correlations

		MEAP Composite w/o Social Studies	% of households with income < \$30K
MEAP Composite w/o Social Studies	Pearson Correlation	1.000	-.494**
	Sig. (2-tailed)	.	.000
	N	519	519
% of households with income < \$30K	Pearson Correlation	-.494**	1.000
	Sig. (2-tailed)	.000	.
	N	519	519

** . Correlation is significant at the 0.01 level (2-tailed).

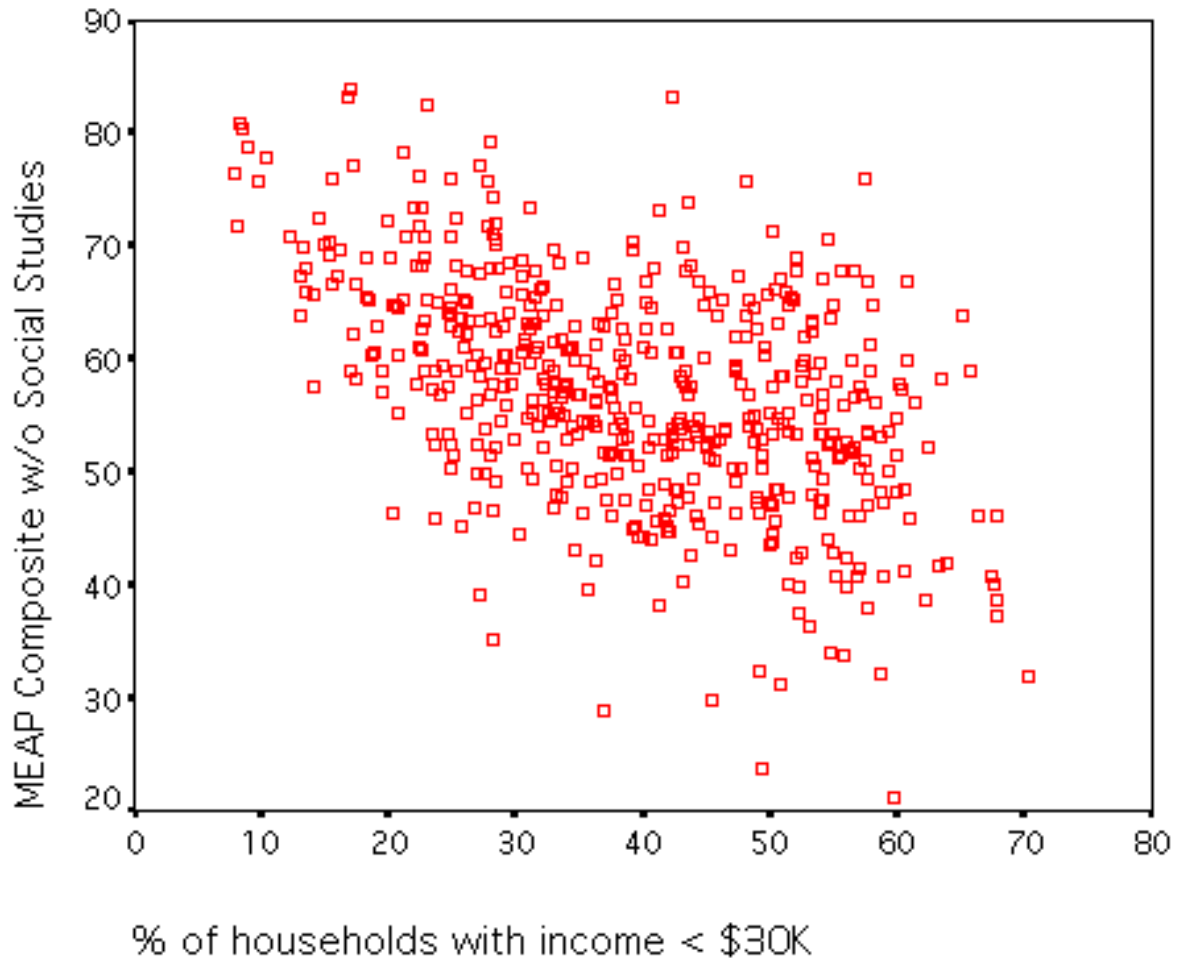


Figure 8. MEAP Scores vs. Income Under \$30,000: Scatterplot.

Summary of Bivariate Correlational Findings

Identified socioeconomic factors correlate with Michigan unified school district aggregate MEAP scores with strengths ranging from moderately strong to strong, with the exception of Pupil to Teacher Ratio, which had a correlational coefficient of 0.088. The weakest of the other identified factors, Percent of Households With Annual Income Under \$30,000, had a coefficient of -0.494. Percent of District Students Eligible For Free and Reduced Price Lunches produced the strongest coefficient, -0.701.

Here is the complete list of SES factors identified, along with the correlational coefficients produced when the factors were paired with Composite District MEAP Scores as defined (in order from strongest to weakest correlation).²³

Percent of District Students Eligible for Free/Reduced Price Lunches.....	-0.701
Percent of District Children Poor, Aged 5-17.....	-0.655
Mean Annual District Household Income.....	0.582
Percent of District Lone Parent Households.....	-0.557
Median Annual District Household Income.....	0.530
State Equalized Homestead Valuation (SEV) Per State Aid Member.....	0.512
Percent of District Households with Annual Incomes Under \$30,000.....	-0.494
District Pupil to Teacher Ratio.....	0.088

²³ Positive and negative signs are ignored when identifying strength of coefficients.

Results of Multiple Regression Test

As noted in Chapter 3, multiple regression is a statistical technique that allows for the simultaneous examination of several independent variables in relation to a dependent variable. In this study, individual district SES factors (as defined) were independent variables; district MEAP Scores (as defined) were the dependent variables.

The goal of multiple regression testing in this case was to produce an equation composed of several variables (representing several of any school district's identified SES factors), the accompanying coefficients, and a numerical constant, the sum of which would provide a reasonably accurate prediction of that district's MEAP Score.

Using the SPSS statistical computer program, various SES factors were combined, with the goal of finding the combination of factors which had the highest predictive power.

The combination of 1) Percent of Students Eligible for Free or Reduced Price Lunches, 2) Percent of Lone Parent Households, and 3) Mean Annual District Household Income produced a predictive equation with the most power (0.749). This was higher than any of the individual bi-variate correlational coefficients.

Table 10 summarizes that model. It indicates that the power of the equation to predict district MEAP scores is 0.749, considered a strong correlation in the social sciences.

Table 10 gives R Square as 0.561. This number is computed by squaring R (0.749×0.749), and it gives the percent of MEAP score variance (district-to-district) attributable to the three identified SES factors. That is, according to the multiple regression, somewhat over half

Table 10
Results of Multiple Regression

R	R Square	Adjusted R Square	Std. Error of the Estimate
.749 ^a	.561	.558	6.598

a. Predictors: (Constant), Mean Income, % lone-parent households, % of Free and Reduced Lunch Eligibility

of district MEAP Score values is accounted for by, or is attributable to, the three SES factors. The remaining 0.439 (1 - 0.561) is attributable to other unidentified factors. The Adjusted R Square statistic (0.558) is slightly lower than the given R Square; this is a reflection of the small number of missing data from the spreadsheet.

The model produced numerical coefficients for each SES factor:

For Percent of Students Eligible for Free or Reduced Price

Lunches; -0.226.

For Percent of District Lone Parent Households; -0.767.

For Mean Annual District Household Income; 0.00014.

Those three values (along with a generated numerical constant of 64.533) allow for the production of the predictor equation:

$$-0.226a + -0.767b + 0.00014c + 64.533 =$$

(Predicted) District MEAP Score, where

a = Percent of Students Eligible for Free or Reduced Price Lunches

b = Percent of District Lone Parent Households

c = Mean Annual District Household Income

Illustrations of the Equation's Predictive Power

Following are three examples that illustrate the predictive power of the equation.

Adams Township School District

For the Adams Township School District in Michigan's Upper Peninsula, the values for the three identified SES factors are as follows:

- a (Free/Reduced Lunch, %) = 44.3
- b (Lone Parent Households, %) = 9
- c (Mean Household Income, \$) = 30,223

Substituting those values for a, b, and c in the equation gives $-0.226(44.3) + -0.767(9) + 0.00014(30,223) + 64.533$, or a Predicted District MEAP Score of 51.833. The actual MEAP Score for the Adams Township School District is 51.7.

Adrian Public Schools

For the Adrian Public Schools, the SES values are as follows:

- a (Free/Reduced Lunch, %) = 33.2
- b (Lone Parent Households, %) = 12
- c (Mean Household Income, \$) = 51,686

Using those values in the equation, the predicted MEAP Score for the Adrian Public Schools is 53.033; the actual score is 53.

West Bloomfield School District

For the West Bloomfield School District the SES values (radically different than those in the two previous examples), are as follows:

$$a \text{ (Free/Reduced Lunch, \%)} = 5.5$$

$$b \text{ (Lone Parent Households, \%)} = 6$$

$$c \text{ (Mean Household Income, \$)} = 132,080$$

Using those values in the equation, the predicted MEAP Score for the West Bloomfield Schools is 77.2; the actual score is 76.6.

Most Michigan unified district SES factors provide a similar level of MEAP Score prediction. For some districts, the SES factors provide a less accurate prediction. Overall, the strength of the predictive formula is 0.749 (with 1.0 indicating perfect prediction.)

Summary of Multiple Regression Findings

The values of three district SES factors--Percent of District Students Eligible for Free or Reduced Price Lunches, Percent of District Lone Parent Households, and Mean Annual District Household Income, when substituted for a, b and c in the expression $-0.226a + -0.767 + 0.00014c + 64.533$ --produce a fairly reliable predictor of that district's Composite High School MEAP Score, as defined.

CHAPTER FIVE
SUMMARY, CONCLUSIONS, INFERENCES, IMPLICATIONS, AND
RECOMMENDATIONS FOR FURTHER STUDY

Summary

This researcher discovered the level of correlation between several school district socioeconomic factors and aggregate district student achievement on the High School MEAP Tests. For purposes of this study, “District High School MEAP Score” was defined as the the percentage of district high school students who took the bank of MEAP tests in spring of 2000. The strongest relationship discovered was that between Percent of District Students Eligible for Free and Reduced Priced Lunches and District High School MEAP Score (-0.701).

Additional research used multiple regression to show that combinations of SES factors could produce a higher power of prediction (“R”) than any of the SES factors considered independently. The combination producing the highest power of prediction (-0.749) consisted of the following three factors:

1. Percent of District Students Eligible for Free and Reduced Priced Lunches
2. Percent of District Lone Parent Households
3. Mean District Annual Household Income

Multiple regression assigned numerical coefficients to each factor. Using those coefficients and the numerical constant generated by the test, an equation predictive of a district's High School MEAP Score was developed:

$$-0.226a + -0.767b + 0.00014c + 64.533 =$$

(Predicted) District High School MEAP Score

where

a = Percent of Students Eligible for Free or Reduced Price Lunches

b = Percent of District Lone Parent Households

c = Mean Annual District Household Income

Squaring R gave 0.561. This gave the percent of MEAP score variance (district-to-district) attributable to the three identified SES factors. As noted in Chapter Four, this means that about half of district MEAP Score values is accounted for by the three SES factors, with the remaining 0.439 attributable to other factors.

Conclusions

This research was undertaken to answer several questions. The first, “How do Michigan public school district SES factors correlate with aggregate district MEAP scores?” was answered by the generation of correlational coefficients for each factor studied. Those relationships were generally found to be strong, with measures of average family income and average number of parents per household being the strongest.

The second, “How can individual correlation coefficients be combined to form an equation strongly predictive of district MEAP scores?”, was answered with the generation of the equation given above.

The results of this study are in agreement with Cooley (1993), who found that three family socioeconomic factors (student living in poverty with a single parent who is not a high school graduate) account for 60% of the variation in average district test scores among Pennsylvania school

districts. Similarly, a study of Texas school district “Report Cards” (which report, among other things, school scores on the TAAS) found that “percent of economically disadvantaged students” had a fairly strong (-0.588) relationship with the Report Card indicators (Bobbett et al. 1996). The results of the present study also reflect the findings of an Educational Research Service study that showed that poverty alone accounted for 56% of the variance among state average test scores in the NAEP-92 Trial State Assessment in mathematics. That same study showed that a stunning 89% of those variations were due to poverty and just three other, outside-of-school demographic factors (number of parents living at home, parents’ education, and community type.) (Educational Research Service, 1994).

Inferences and Implications

In Chapter One, this question was posed somewhat rhetorically: Is it appropriate to judge or rank schools or school districts based on MEAP tests? If it is clear (as this study has shown) that those scores are closely associated with a district’s socioeconomic status, such a practice is at least questionable, and at most demands a stronger explanation than that provided thus far by policymakers.

As noted in Chapter One, a limitation of this study was that the researcher examined MEAP data from one year only. However, an informal longitudinal examination of MEAP achievement and district demographic data, where available, suggests that such data do not change quickly; that is, district-level descriptive data remain fairly stable year-to-year. It is likely, then, that district high school MEAP scores (and,

therefore, the percentage of students eligible for MEAP Merit Scholarships) can continue to be predicted with a good degree of accuracy before students take their tests. Specifically, more dollars will be (and have been) sent to districts of higher SES and fewer dollars to those with lower SES. In that the MEAP Scholarship money essentially represents college tuition, the Scholarship Program is serving (perhaps inadvertently) to widen the higher education accessibility gap.

The current study, then, provides an unsettling answer to the other question raised in Chapter One: Might the use of student test scores for high stakes tend to increase the opportunity gap between students of lower SES and those of higher SES? The answer is yes, and a blunt way of stating this is that disadvantaged students are having to look up even further to see bottom.

High-stakes mania threatens struggling, low SES schools and districts in another way. If schools and districts that tend to score poorly on the MEAP and on standardized tests (those with low SES) are punished with lack of accreditation, periodic traumatic administrative turnover, and even threat of closure, why would any teachers want to teach there? Why would administrators want to work there? Barring extreme, sacrificial altruism (it does exist), why would any educator accept employment there when the data tells us in advance that the school may be in line for state-defined failure?

Defense of the use of MEAP scores for high stakes purposes has often taken this form: “The ‘bar’ (passing the MEAP test) is set at the same height for all; there is no class, race or gender discrimination in

either the test itself or in scoring procedures.”²⁴ Using that high jumper analogy: how reliable are track meet results if some of the participants came to the bar with their feet tied together? How fair is it to judge the track coaches?

The current “choice fever” on the political right makes sufferers accept the assumption that the playing field is level, that is, all children come to school equally ready to learn, and with all teachers facing the same level of task difficulty, and those assumptions are simply false, as this study demonstrates. It is important to note that the results of this study do not allow educators to expect low achievement of individual students. This study simply implies that Cooley (1993) was correct: The educational task in disadvantaged districts is not impossible, simply harder.

What do the few outliers--districts with low SES but relatively high MEAP scores--tell us? Perhaps that the job of educating all students can be accomplished in low SES schools, if all the conditions are right. (See Recommendations for Further Study below.)

An alternate interpretation of the data presented in this study is that teachers and administrators in low-MEAP-score districts are simply not doing their jobs very well. That is a testable hypothesis (albeit testable with great difficulty), one which this study did not address. It is an unlikely one, however, given the strong SES/MEAP Score correlations found.

It is unlikely for another reason: if low achievement on high school MEAP tests in low SES districts is due to the consistent inferiority of

²⁴ Although MEAP test items are reviewed by several committees, including one charged with eliminating bias, the author of this study does not accept that the tests are unbiased. However, the validity of the MEAP tests--their freedom from bias and their ability to accurately tell us what students know and can do--was not addressed via this study.

staffs in those districts, it follows that there exists a high correlation between low district SES and the presence of inferior teachers and administrators. That is, either some unknown factor strongly tends to attract incompetent teachers and administrators to poorer districts, or, a nearly impossible mathematical coincidence has occurred (all the bad teachers and administrators have migrated to poor districts.) This is logically absurd.

It was stated earlier that the ability to develop an equation based on SES factors and predictive of MEAP scores would spotlight the illegitimacy of attaching high stakes to such scores. The existence of a predictor equation may provide some degree of guidance for policy and lawmakers who influence education resource allocation. The chief implication is this: Public policies that hold the most potential to decrease poverty, and child poverty in particular, may also positively impact student achievement, at least as indicated by student test scores. Early intervention is likely crucial: Jencks and Phillips found that one-half of the black-white academic gap in reading and mathematics in grade 12 was explained by differences in first grade scores (Jencks and Phillips, 1998).

It must again be emphasized that even a strong correlation does not prove causality. In this case, that means that although several SES factors were found to be associated with MEAP scores, it cannot be said that low SES causes low scores. It may be that one or more intervening variables are operative. Is it possible, for example, that an unidentified variable (to some degree) causes both poverty and low academic achievement of children of poverty?

Still, a reasonable hypothesis is that conditions often associated with lower SES--less nurturing of children, fewer positive adult role

models, less access to cultural enrichment, fewer books and less reading in poor households, for instance--indeed account for much of the typical lower academic achievement of children of poverty.

Recommendations for Further Study

Although the slow change in district SES and MEAP data implies that the equation developed by this researcher is likely to remain a strong predictor of MEAP scores, the equation's usefulness should be verified as new data become available.

The correlational scatterplots presented in Chapter Four show that a few school districts are "outliers"; that is, they have a lower socioeconomic status, but relatively high MEAP scores. Such schools and districts have been, and should continue to be, the objects of further study, with focus on the question "What are these districts doing right, and different?" Such study, which could be qualitative or quantitative, might be useful in identifying intervening causal variables, something this study did not do. A caution, however, must be made. Studying such outlier districts implies an assumption that the MEAP tests are good measures of student academic achievement. Identifying what those supposedly overachieving schools are "doing right" could conceivably be simply a matter of discovering which tricks students and teachers are using to score well. This researcher considers the question of whether MEAP tests are good indicators of true student achievement to be unsettled.

One example is illustrative of that possibility. Figure 9 is a MEAP Mathematics Test item for grade four, released by the Michigan Department of Education. Standard Venn Diagram notation has not been used in this item: "Z" would typically represent all elements within the

rectangle (the “universal set”), not simply the “outside” area.

Conventional Venn Diagram notation does not include separate labeling of *parts* of ovals, such as is seen here. The atypical notation renders the item invalid and scores based on the item unreliable.

The results of this study and others, the MEAP test item concerns noted above, and the continued use of MEAP test scores for high-stakes purposes, suggest that further study of the validity, reliability, and usefulness of the MEAP is appropriate.

Figure 9. MEAP Sample Item.

References

Achieving equality through high standards and accountability. (2001). White House paper. Retrieved January 12, 2002, from <http://www.ed.gov/inits/ncib/part3.html>

Achilles, C., Finn, J., Gerber, S., & PannoZZo, G., (2001). *The power of small classes to reduce the test-score achievement gap.* Paper presented at the Mid-South Educational Research Association, Little Rock, AK.

Achilles, C., & Mitchel, C. (2001-2002). National impact: A challenge for educators in dealing with child poverty and punitive policy. *National Forum of Applied Educational Research Journal*, 5, (1).

Adams, C., & Singh, K. (1998). Direct and indirect effects of school learning variables on the academic achievement of African American 10th graders. *The Journal of Negro Education*, 67 (1), 48-66.

Adams, E. (1994). *The effects of cost, income, and socio-economic variables on student scholastic aptitude scores.* Ball State University. DAI, 55, no. 08A, 2276.

Anderson, G. (2001). Promoting educational equity in a period of growing social inequity. *Education and Urban Society*, 33 (3), 320-332.

Anspaugh, D. (1995). *Teachers' perception of instructional leadership and student achievement.* Western Michigan University. DAI, 56, no. 06A, 2049.

Beck, B. (1995). *State-mandated test score accountability: its usefulness and meaning for school districts.* The University of Utah. DAI, 56, no. 11A: 4370.

Berliner, D., & Biddle, B. (1995). *The manufactured crisis.* Addison-Wesley Publishing Co., Reading, MA.

Bernaureur, J., & Cress, K. (1997). How school communities can help redefine accountability assessment. *Phi Delta Kappan*, 79, 71-75.

Bobbett, G., French, R., & Achilles, C. (1993). *A study of various statistical analyses applied to school report cards.* Mid-South Educational Research Association, New Orleans, LA.

Bobbett, G., French, R., & Achilles, C. (1996). *How meaningful are Texas school district report cards?* American Association of School Administrators. San Diego, CA.

Burde, R. (1989). *A study of the relationship of class size and student achievement on the Michigan Educational Assessment Program fourth-grade test*. Western Michigan University. DAI, 50, no. 07A, (1989): 0257.

Bushweller, K. (1997) Teaching to the test. *The American School Board Journal*. 184, 20-26.

Carter, G., & Cunningham, W. (1997). *The American school superintendent*. San Francisco: Jossey-Bass.

The Center on Education Policy. (2001). *It takes more than testing: Closing the achievement gap*. Washington, D.C.

The Center on Education Policy and American Youth Policy Forum, (2000). *The good news about american education*. Washington, D.C.

Cooley, W.W. (1993). *The difficulty of the educational task*. Pittsburgh PA: University of Pittsburgh. Pennsylvania Educational Policy Studies. Paper #16.

Danielson, M. (1999). *How principals perceive and respond to a high-stakes accountability measure*. The University of North Carolina at Chapel Hill. DAI, 61, no. 03A (1999): 827.

Demars, C. (1998). *The impact of test performance and response format on performance*. Michigan State University. DIA 59, no. 10A, (1998): 3796.

Dorn, S. (1998). The political legacy of school accountability systems. *Educational Policy Analysis Archives*, 6 (1).

Edelman, M. (1997). Leaving no child behind. *School Administrator*, 54, 14-16.

Educational Research Service. (1994). *NAEP test scores: Should they be used to compare and rank state educational quality?* Arlington, VA.

English, F. (1992). *Deciding what to teach and test*. Newbury Park: Corwin Press.

Epstein, J. (1998). *Power, politics and the National Assessment of Educational Progress*. Rutgers The State University of New Jersey, New Brunswick, NJ. DAI, 59, no. 05A (1998): 1445.

Farfsing, J. (1988). *A study of the correlation between key traits of effective school research and the Michigan Educational Assessment Program*. Central Michigan University. MAI, 26, no. 04, (1988): 0362.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28 (2), 465-498.

Finn, C. (1998). Revisiting the federal role: first, do no harm. *Education Week*, 17 (32), 52.

Fraenkel, J. & Wallen, N. (2000). *How to design and evaluate research in education*. McGraw-Hill, Boston, MA.

Fullan, M. (1999). *Change forces: The sequel*. London: Falmer Press.

Gay, L. & Airasian, P. (2000). *Educational research*. Merrill. Upper Saddle River, NJ.

Gill, S. (1997). *Educational expectations and school achievement of children in poverty*. The Pennsylvania State University. DIA, 58, no. 05A, (1997): 1579.

Goldstein, N. (1991). *Why poverty is bad for children*. Harvard University. DAI, 52, no. 03A, (1991): 1098.

Grinion, P. *Academic achievement and poverty: Closing the achievement gap between rich and poor high school students*. Spalding University. DAI, 60, no. 02A, (1999): 0386.

Haney, W., Fowler, C., Wheelock, A., Bebell, D., & Marlec, N. (1999). Less truth than error? An independent study of the massachusetts teacher tests. *Education Policy Analysis Archives*, 6, (4).

Haney, W. (2000) The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8 (41).

Hanley, D. (2000). *The changes teachers have made in their classroom practices to teach the Massachusetts frameworks*. University of Massachusetts Lowell.

Hardy, L. (2000). The trouble with standards. *The American School Board Journal*, 187 (9), 22-26.

Hart, R. (1999). *The impact of high stakes accountability on high school curriculum*. Texas A&M University. DAI 60, no. 11A (1999): 3854.

Hendrie, C. (1999). Poor districts fare worst on N.Y. assessment. *Education Week*, 18 (38).

Herman, J., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12, 20-25.

Herrick, M. (1999). *State-level performance assessments and consequential validity*. University of Maryland College Park. DAI 60, no. 04A (1999): 1096.

Hess, F., & Brigham, F. (2000). None of the above. *The American School Board Journal*, 187 (1), 26-9.

Hill, A. (1999). *Examining the effects of school climate, student background, teacher expectations, and school-community integration on academic achievement among high poverty urban schools*. University of Houston. DAI, 60, no. 04A (1999): 1018.

Hinojosa, L. (1997). *A comparative study of high-stakes testing (TAAS) curricular, instructional, and operational practices for Title I schools: Improving versus declining schools*. The University of Texas at Austin. DAI 59, no. 02A, (1997): 0384.

Hoff, D. (1998). Kentucky to include norm-referenced test in accountability plan. *Education Week*, 18 (8), 16.

Hoff, D. (2000). Key governors want NAEP given more often. *Education Week*, 19 (26), 18.

Hurwitz, N., & Hurwitz, S. (2000). Tests that count. *The American School Board Journal*. 187 (1), 20-25.

Illig, D. (1998). *Birth to kindergarten: The importance of the early years. A comprehensive review of the literature* (Report No. PS026313). (ERIC Document Reproduction Service No. ED417801)

Jadallah, E., & Partin, R. (1995). Educational trends at the turn of the millennium. *American Secondary Education*. 23, (3), 23-31.

Jencks, C. & Phillips, M. (1998). The black-white test score gap. *Education Week*, 18 (4), 45.

Jennings, J. (2000). *Presidential politics and education*. Center on Education Policy. Retrieved from the World Wide Web, January 2, 2002. www.ctrepol.org/articles/presidential_politics_education.

Johnston, R. (1997a). District push to get students to forgo test assailed. *Education Week*, 16 (2), 6.

Johnston, R. (1997b). Michigan schools warned not to excuse pupils from test. *Education Week*, 16 (4), 5.

Johnston, R. (1997c). Just saying no. Parents opt out of test in Michigan. *Education Week*, 16 (5), 6.

Johnston, R. (1999). Bush zeroes in on accountability for federal K-12 funds. *Education Week*, 19 (1), 5.

Jones, K., & Whitford, B. (1997). High stakes accountability and student performance assessment. *Phi Delta Kappan*, 79, 276-281.

Karier, C. (1972). Testing for order and control in the corporate liberal state. *Educational Theory*, 22, (2).

Kaufhold, J. (1995). Testing, testing. Why school boards shouldn't take standardized tests too seriously. *The American School Board Journal*, 182 (1), 41-2.

Keusenkothen, K. (1990). *Family environment and student achievement*. University of Missouri-Columbia. DAI, 52, no. 07A, (1990): 2492.

Kim, K. (1992). *The effect of poverty on children's academic performance*. The University of Texas at Dallas. DAI, 53, no. 06A, (1992): 2124.

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education*. 51 (4), 315-27.

Krathwohl, D. (1997). *Methods of educational and social science research: An integrated approach*. Addison-Wesley Publishing Co., Reading, MA.

Lawton, M. (1996). State test questions focus of new scrutiny. *Education Week*. 15 (1), 1.

Lawton, M. (1997). State boards' leaders call for assessments bearing consequences. *Education Week*. 17 (39), 7.

Lewis, A. (2000). Increasing federal control. *Phi Delta Kappan*, 82 (3).

MacEwan, A. (1997). Guest editorial: Bill Clinton's practical approach to public education. *Radical Teacher*, 51, 2-5.

Makki, S. (1992). *Identifying predictive variables in the success of high school students on the MEAP math test*. Wayne State University. DAI, 53, no. 12A, : 4188.

Madaus, G. (1988). The influence of testing on the curriculum. *Critical Issues in Curriculum*. Chicago: University of Chicago Press.

Manzo, K. (1996). Vermont to combine standardized tests with portfolios. *Education Week*. 16 (48), 3.

Markel-Fox, S. (1993). *The effects of race and poverty on preschoolers' performance in modified testing contexts*. North Carolina State University. DAI, 54, no. 12B, (1993): 6479.

McNeil, L. (2000). *Contradictions of school reform: The educational costs of standardized testing*. New York: Routledge.

Metcalf, S. (2002). Reading between the lines. *The Nation*. Retrieved from the World Wide Web, February 12, 2002. www.thenation.com/doc/Print.mhtml?i=2002018&s=metcalf

National Academy of Education. (1977). *Improving educational achievement*. New York: Author.

Payne, K., & Biddle, B. (1999). Poor school funding, child poverty, and mathematics achievement. *Educational Researcher*, 28, (4), 4-12.

Pedroza, B. (1997). *Consequences of high-stakes testing on historically disenfranchised students: Analysis of student outcomes*. The University of Texas at Austin. DAI, 58, no. 07A, (1997): 2615.

Polinard, J., Wrinkle, R., & Meier, K. (1995). The influence of educational and political resources on minority students' success. *The Journal of Negro Education*. 64, 463-74.

Popham, W. (1998). *Tests that help teaching: A personal odyssey*. Los Angeles: IOX Assessment Associates.

Popham, W. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56 (6), 8-15.

Popham, W. (2000). The score-boosting game. *The American School Board Journal*. 187 (6), 36-9.

Pre-election preview: presidential preferences on education. (2000). *The Education Digest*, 66 (2), 4-7

President Bush's education policy. (2001). Washington, D.C.: U.S. Department of Education. Retrieved from the World Wide Web, December 20, 2001. www.ed.gov/inits/ncib/part3.html

Price, W. (1990). Education research and the practitioner: the great divide. *The AASA Professor*. 19, (2), 18.

Prince, T. (2001). Education reform meets bipartisan support. *Business Education Forum*, 55 (4), 14-15.

Pruneau, D. (2001). Letter to Brighton Area Schools parents from superintendent's office, Brighton, MI, 1.

Ray-Taylor, R. (1991). *Teachers' interpretation and use of state achievement test results*. Michigan State University. DAI, 53, no. 01A, (1991): 0131.

Rhode Island investigates testing snafu. (1999). *The American School Board Journal*. 186 (5), 14.

Roscigno, V. (1996). *Race, place, and the reproduction of educational disadvantage: The black-white gap and local structures of opportunity*. North Carolina State University. DAI, 57, no. 10A (1996): 4552.

Rotberg, I. (2000). Campaign 2000: notes to the next president on education policy. *Education Week*, 19 (28), 72.

Rotberg, I. (2001), A self-fulfilling prophecy. *Phi Delta Kappan*. 83 (2), 170-171.

Rothman, R. (1996). Taking aim at testing. *The American School Board Journal*, 183 (2), 27-30.

Saks, J. (March, 1999). High stakes and high anxiety in Virginia - will your state be next? *The American School Board Journal*, 185, (2), 6-7.

Saks, J. (1999). High stakes and high anxiety in Virginia--will your state be next? *The American School Board Journal*. 186, (3), 6-7.

Scranton, J. (2001). Press release from state representative's office, Lansing, MI, 1.

Secumski-Kiligian, E. (1993). *The relationship between elementary school teachers' job satisfaction, principals' managerial styles and student achievement*. Wayne State University. DAI, 54, no. 05A, (1993): 1625.

Shepard, L. (1994). The challenges of assessing young people appropriately. *Phi Delta Kappan*, 76 (1), 206-212.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, (7), 4-14.

Shirley, D. (1997). *Community organizing for urban school reform*. Austin: University of Texas Press.

Siminski, R. (2000). *State funding, student performance, school effectiveness: A Massachusetts experience*. Johnson and Wales University. DAI, 61, no. 12A (2000): 4636.

Smith, M. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practices*, 10, 7-11.

Smith, M. (1999). *How the apportionment of district money is associated with differences in the grade 8 mathematics and reading achievement scores for minnesota students*. University of Minnesota. DAI, 60, no. 06A (1999): 1908.

Smith, S., & Jorgenson, O. (1997). Test prep 101. *The American School Board Journal*, 184 (9), 46-7.

Snyder, D. (1995). *A study of the collection between per pupil funding and student achievement in the state of Michigan*. Western Michigan University. DAI, 56, no. 04A, (1995): 1246.

Spady, W. (2001). *Beyond counterfeit reforms*. Lanham, NY: Scarecrow Press, Inc.

Tate, W. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: an update. *Journal for Research in Mathematics Education*, 28, 652-79.

Test results aren't comparable, panel says. (1998). *The American School Board Journal*, 185, (9), 14.

Thomas, V. (2001). High stakes testing: The tail wagging the dog. *The Phelps-Stokes Fund Dialogue*, 10, 11.

Trueba, H. (2001). Conclusion: Polar positions on the Texas Assessment of Academic Skills (TAAS): Pragmatism and the politics of neglect. *Education and Urban Society*, 33, (3), 333-344.

Tuch, W. (1996). *The impact of high stakes performance assessment on teachers and learning*. Arizona State University. DAI, 57, no. 10A: 4220

Tucker, J. (1996). *Public schools in Michigan: Impact of per-pupil revenue and other financial variables on student achievement as measured by the Michigan Educational Assessment Program*. Eastern Michigan University. DAI, 58, no. 10A, (1996): 3841.

U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. (1996). *Urban schools: The challenge of location and poverty*. Washington, D.C.

Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring*. Albany: State University of New York Press.

VanValkenburg, L. (1990). *Relationship between community wealth and student performance on the Michigan Educational Assessment Program*. The University of Toledo. DAI, 51, no. 09A, : 3050.

Viadero, D. (1998). Test data show disadvantaged pupils lag behind more affluent counterparts. *Education Week*, 18 (2), 5.

Vogler, K. (2000). *The impact of high-stakes, state-mandated student performance assessment on 10th grade English, mathematics and science teachers' instructional practices*. University of Massachusetts Lowell. DAI, 61, no. 02A: 487.

Vornberg, J. and Hart, R. (2000). *Accountability and high stakes testing: Views from Texas schools*. Paper presented at National Council of Professors of Educational Administration conference, Ypsilanti, MI

Wasserman, S. (Sept. 2001). Quantum theory, The uncertainty principle and the alchemy of standardized testing. *Phi Delta Kappan* (83) 1, 28-40.

Wiggins, G. and McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

Williams, I. (1993). *The impact of test sophistication of tenth grade students' Michigan Educational Assessment Program (MEAP) scores*. Western Michigan University. DAI, 54, no. 03A, (1993): 0906.

Yarbrough, T. (1999). *Teacher perceptions of the North Carolina ABC program and the relationship to classroom practice*. The University of North Carolina at Chapel Hill.

Young, P. (1994). *Attitudes, perceptions, and beliefs of students and teachers regarding standardized tests*. The State University of New Jersey-Rutgers. DAI, 55, no. 06A: 1539.

Zlatos, B. (1995). "Don't test, don't tell" inflates test scores. *The Education Digest*, 60 (2), 4-9.

APPENDICES

APPENDIX A: SCHOOL DISTRICT DATA SOURCES

Michigan Department of Education website:

<http://www.state.mi.us/mde/reports/msr99/rptdef.htm>

1. K-12 Enrollment (1999)
2. Percent of District Students Eligible for Free and Reduced Price Lunches (1999)

Michigan Department of Education website:

<http://www.state.mi.us/mde/reports/B1014/terms.html>

1. State Equalized Homestead Valuation (SEV) Per State Aid Member (2000)
2. Pupil to Teacher Ratio (2000)

United States Census Bureau website:

<http://www.census.gov/hhes/www/saipe/school/sd97over.html>

1. Total District Population (1997)
2. Percent of District Children Poor, Ages 5-17 (synthetic estimate 1997)

Standard and Poor's website:

<http://www.ses.standardandpoors.com/>

1. Percent of District Lone Parent Households (2000)
2. Mean District Household Income (2000)
3. Median District Household Income (2000)
4. Annual Income Under \$30,000 (2000)
5. Composite MEAP Passing Without Social Studies (2000)

APPENDIX B: NOTES ON SCHOOL DISTRICT DATA SOURCES

K-12 Enrollment

The Michigan Department of Education defines “K-12 Enrollment” as the number of pupils in a school district. It includes alternative and special education pupils, but does not include adult education participants. District enrollments are taken from the audited fall state aid pupil membership counts in full-time equivalents (FTE) reported by the school districts.

Percent of District Students Eligible for Free and Reduced Price Lunches

The Michigan Department of Education defines “Free and Reduced Lunch” as the percentage of students who are eligible to participate in the school breakfast and/or lunch program at no cost or at a reduced cost. Eligibility is determined by applying federal guidelines through a review of a family's size and financial resources.

State Equalized Homestead Valuation (SEV) Per State Aid Member

This Michigan Department of Education figure represents a calculation made by dividing the taxable value of real estate and personal property (homestead assessed value as finally equalized by the State Tax Commission) in the district, by the number of pupils legally enrolled and reported to the MDE on the Fall DS-4061 report.

Pupil to Teacher Ratio 2000

This Michigan Department of Education figure is calculated by dividing the district's total K-12 enrollment at the Fall Pupil Count (not including Special Education) by the total number of K-12 teachers in the

district. This does not include support staff.

Total District Population, and the Percent of District Children Poor, Ages 5-17

Before 1999, Title I funds were distributed to counties, with individual states having responsibility to redistribute the funds from counties to school districts. The Improving America's Schools Act of 1994 directed the U.S. Department of Education to consider distributing Title I grants directly to school districts beginning in the 1999-2000 school year. In response, the Department of Education asked the U.S. Census Bureau to provide estimates at a school district level that would make this possible.

The U.S. Census Bureau developed the new poverty estimates (Percent of District Children Poor, Ages 5-17) using a "synthetic estimator." The estimate is the number of poor children as measured in the 1990 census (but for school districts defined by geographic boundaries for the 1999-2000 school year) multiplied by the proportional change in child poverty in the county in which the district is located. The school district estimates were then adjusted so that they summed to the county estimates, which were, in turn, controlled to sum to the state estimates. The county change is computed as the change from the 1990 census to the U.S. Census Bureau's county model estimates for 1997. The county estimation models use administrative data and data from the 1990 census and the Current Population Survey to estimate child poverty.

The "synthetic" estimates for school districts do not have the degree of precision normally associated with data published by the U.S. Census Bureau.

The U.S. Census Bureau constructed an evaluation file with the purpose of examining the precision of these estimates. It applied the estimation methods described above to the base data from the 1980 census, calculated "synthetic" estimates for 1990, and compared the results to the 1990 census. Districts were excluded from the evaluation file if they were coterminous with counties, did not cover the entire grade range, or were known to have changed geographic boundaries between 1980 and 1990. The remaining districts represented 61 percent of the total number of districts and contained 56 percent of school-age children.

According to the Census Bureau, the estimates proved reasonably accurate for large districts (with a population greater than 40,000) and for those coterminous with counties. These districts comprise 13 percent of districts but include 62 percent of poor children. However, there are potentially very large errors for small districts. The average absolute difference between the estimate of related school-age children in families in poverty and the same figure from the 1990 census is 60 percent of the census figure for school districts in the evaluation file. The results for large districts are similar to the county model results, which have a comparable figure of only 16 percent.

The Census Bureau has stated that the estimates for the number of school-age children and the total population are relatively more accurate. The estimate of the number of school-age children in the evaluation file differs from the 1990 census figure by 17 percent on average. For the estimate of total population, the difference is only 13 percent.

While the error relevant to the average school district is large, the error relevant to the average poor child is much less. The average absolute difference between the evaluation file estimate and the census

figure is only 22 percent of the average number of poor children per district.

Percent of District Lone Parent Households

The U.S. Census Bureau defines this figure as the proportion of households occupied by a single adult and one or more children. A house, an apartment or other group of rooms, or a single room, is regarded as a housing unit when it is occupied or intended for occupancy as separate living quarters. The figure is calculated by dividing the number of single-headed female and male households with children by the number of households in the district.

Mean Household Income

This U.S. Census Bureau figure is given as the typical income for a household in the district and is calculated by adding all household incomes in the district and dividing the sum by the total number of households.

Median Household Income

The U.S. Census Bureau defines “Median Household Income” as the point in a distribution of household incomes at which half of the incomes are below that amount, and half are above it. Household income encompasses all persons 15 years or older in the household, regardless if the individuals are related to one another. Income includes wages and salaries, self employment, interest, rentals and dividends, social security, public assistance, retirement, and disability.

Annual Income Under \$30,000

This figure was computed by adding together the percent of households with annual income less than \$15,000 with the percent of households with annual income between \$15,000 and \$30,000 from the Standard and Poor's website, which obtained the information from the U.S. Census Bureau.

Composite MEAP Passing Without Social Studies

This indicator broadly reflects the percentage of included MEAP tests that received passing scores. MEAP Composite Passing includes tests that "meet state standards" or "exceed state standards." For all five high school subject tests, this is defined as Level I or Level 2. The MEAP Composite Passing measure excludes MEAP High School Tests taken by students while in grade 10 and grade 12. These tests are excluded from the composite measure because when viewing a test cohort (e.g., grade 12 in 1999), Michigan is unable to identify which students may have taken the High School Test at some time other than the current year. In addition, about 95% of MEAP High School Tests are taken by students while in grade 11. Therefore, the grade 11 scores are used as the proxy for the High School Test when calculating the MEAP Composite Passing rate.

The MEAP Composite Passing measure excludes social studies tests, which Michigan introduced in 1999. This is significant in that the MEAP Merit Scholarship is awarded to high school students who pass four of the five high school tests, the social studies test being excluded. The indicator is calculated by dividing the total number of scores that met state standards in mathematics, reading, science, and writing by the total number of 11th grade scores. This data came from Standard and

Poor's, which obtained the data from the Michigan Department of Education K-12 Database.

APPENDIX C: HUMAN SUBJECTS COMMITTEE APPROVAL