TO WHAT EXTENT DOES EMPOWERING TEACHERS WITH STUDENT STANDARDIZED TEST DATA IMPROVE STUDENT LEARNING?

by
Barbara J. Light

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF EDUCATION SPECIALIST IN EDUCATION AT NORTHERN MICHIGAN UNIVERSITY

December 9, 2009

APPROVED BY: Derek L. Anderson, Ed.D.

DATE: December 10, 2009
Table of Contents

Abstract ...................................................................................................................................... 3

Chapter I:  Introduction

Statement of Problem ............................................................................................................. 5
Research Questions .................................................................................................................... 6
Definition of Terms ................................................................................................................... 6

Chapter II:  Review of Literature

To What Degree Do Standardized Tests Measure Student Learning? ................................... 9
How Much Impact Do Teachers Have on Student Test Scores? ......................................... 13
To What Depth Can Teachers Understand and Analyze Data? .............................................. 16
What Conditions Empower Teachers to Make the Changes Indicated by Data? .................... 19
To What Extent Does Empowering Teachers with Data Improve Student Learning? ............ 22

Chapter III:  Results and Analysis Relative to the Problem

To What Degree Do Standardized Tests Measure Student Learning? ................................. 28
How Much Impact Do Teachers Have on Student Test Scores? ........................................... 30
To What Depth Can Teachers Understand and Analyze Data? .............................................. 30
What Conditions Empower Teachers to Make the Changes Indicated by Data? .................... 30
To What Extent Does Empowering Teachers with Data Improve Student Learning? ............ 30

Chapter IV:  Recommendations and Conclusions

Recommendations ..................................................................................................................... 32
Areas for Further Research ..................................................................................................... 35
Summary and Conclusion ....................................................................................................... 37
References .................................................................................................................................. 39
Abstract

Any study regarding the effect that empowering teachers with student standardized test data has on student learning must be built upon foundational research showing that standardized test data measures student learning, that the actions of teachers can impact student learning, and that teachers can use data and initiate change. While more research must be directed at the construct validity of standardized testing as well as the overall effect that empowering teachers with data has on student learning, the reviewed literature offered the requisite research foundation for further investigation. Additionally, a few case studies suggested a beginning framework for the actualization of empowering teachers with data to improve student learning.
Chapter I: Introduction

The field of education is moving in the direction of standardization. Even before the No Child Left Behind Act of 2001 (NCLB) was enacted, state content standards and criterion referenced standardized testing were replacing local curriculums and norm referenced data analysis (Bond, 1996; Popham, 2008). The title of NCLB (2001) mandated educators “to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind.” More definitively, NCLB, in rewriting Title I of the Elementary and Secondary Education Act of 1965, stated that all children should have “a fair, equal, and significant opportunity to obtain a high quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments” (NCLB, 2001). The phraseology of “no child left behind” offered a captivating political platform and a laudable goal. However, the reality of offering every American student a uniformly competent education is an arduous task that requires states and schools to be regulated by student academic achievement data.

In a discipline in which professionals have historically enjoyed a great deal of freedom about how the business of educating students is accomplished, the new era of accountability is difficult. Yet, student academic achievement data have the power to validate, as well as chastise, the efforts of teachers. Most importantly, the data are believed by many to have the power to drive school reform toward improving education for all students and closing the gaps between subpopulations of students (Boudett et al., 2005; Bray, 2005; Gamble-Risley, 2006; Lachat, Williams, & Smith, 2006).

To lessen academic disparity between subpopulations, a six step process has been at least partially implemented in American education. The first three steps must occur at a national or state level in order to standardize education on a broad scale. First, curriculum content standards
must be established. Second, standardized assessments that test the curriculum content standards need to be created and administered to students to generate academic achievement data. Third, the academic achievement data must be gathered and organized. The first three steps are required by NCLB and have been completed by every state (Education World, 2009; Vu, 2008).

The final steps directly involve classroom teachers and therefore must be accomplished at the district and individual classroom level. For step four, educators require access to the standardized test data. Fifth, the educators need to be able to identify gaps in achievement from the data, which may require some professional education to understand and analyze academic data. Sixth, the educators must be able to make changes that will improve instructional practice and student achievement. Implementation of NCLB to the classroom level of data based decision making requires tools and time that have not yet materialized in all schools. Yet, education is placing a great deal of emphasis on teachers using data to reform schools and closing the achievement gaps. Many leaders, both political and educational, have voiced faith in the future of data based education (Popham, 2008).

Statement of the Problem

NCLB directed states to use data from the 2001-2002 school year as a starting point for student achievement measurement. The starting point was to be the higher student proficiency of the following two options: the proficiency of the lowest achieving student subpopulation in the state or the proficiency of students in the school in the state that ranked at the twentieth percentile based on enrollment. From the starting point data, states were to establish criteria and a timeline for adequate yearly progress toward closing the achievement gaps, with the ultimatum that all students be proficient by the 2013-2014 school year (NCLB, 2001).
Many schools are nearing the timeline set by states for demonstrating that they no longer have any academic achievement gaps between subpopulations of students. The alignment of curriculum and standardized test questions to the content standards set by the state initiated the process of equalizing the educational opportunities of every child. The remaining gaps must be addressed by classroom teachers. The topic of equipping teachers to become users of data and finish the process of closing the gaps is the focus of this paper.

Research Questions

The overall research question for the literature review is: To what extent does empowering teachers with student standardized test data improve student learning? The jump from providing teachers with test data to using those test data to increase student learning is large and can be broken down into several smaller questions that must not be overlooked or assumed: To what degree do standardized tests measure student learning? How much impact do teachers have on student test scores? To what depth can teachers understand and analyze data? What conditions empower teachers to make the changes indicated by data? In order to effectively address the overall research question, each break-out question must be examined with educational research to avoid conclusions predicated on assumptions.

Definition of Terms

- Empowerment in education is a school culture in which participative decision making becomes the normal mode of problem solving (Goyne, Padgett, Rowicki, & Triplitt, n.d.; Short & Greer, 1997). Teacher autonomy was discussed as the key component to empowerment (Goyne, et al., n.d.).

- Student achievement data can and should be from a variety of assessment sources including teacher observations, student questioning, teacher-created formative and
summative assessments, standardized tests, and projects. However, for the literature review, the sole source of student achievement data will be standardized test scores.

- Brown (2006) described data warehousing and explained its role in education. Data warehousing was first used in business and industry. Data warehouses are repositories of information that can integrate information from various sources. Data warehouses are tools that can be used with data analysis tools to provide educators with timely access to demographic and academic data.

- Rubadeau, Garrett, & Rubadeau (1990, p. 12) offered the following definition: “A standardized test is a measuring instrument which must be administered under a standard set of conditions and scored in a predetermined manner. The main purpose of a standardized test is to make it possible to compare or rank students in terms of the specific behaviors sampled by the test.”

- Klein et al. (2009, p. 5) defined test reliability as score consistency and explained that, “Reliability is reported on a scale from 0.00 to 1.00, where higher values indicate greater reliability.” Standardized tests must first be verified as reliable before being evaluated for validity (Brown, 2000).

- Test validity, as defined by Brown (2000, p. 8), “was traditionally subdivided into three categories: content, criterion-related, and construct validity.” Content validity measures “the degree to which a test is a representative sample of the content…the test was originally designed to measure.” (Brown, 2000, p. 8) Criterion validity is demonstrated by correlating one test with another test with established validity. Construct validity is an evaluation of whether the test measures the material knowledge it was designed to measure. Brown (2000) further explained that the field of psychometrics was moving
toward a definition of test validity in which the three types of validity were considered different aspects of construct validity.
Chapter II: Literature Review

Ideally, the literature review of the overall research question relating empowering teachers with test data to student learning would include a plethora of multi-year quantitative studies about student academic achievement under teachers with access to academic achievement data compared to student achievement under teachers without access to academic achievement data. Additionally, in the perfect literature review of the topic, a wealth of qualitative studies examining the framework that enables teachers to understand and use the academic data would be available. The literature to review on the topic of empowering teachers with data is neither abundant or of ideal design to effectively address the overall research question presented in this paper. However, each break-out question has been more comprehensively studied and must be addressed to prevent assumptions from tainting research of the overall question.

To What Degree Do Standardized Tests Measure Student Learning?

If standardized tests are not capable of measuring student learning, the question of empowering teachers with data is premature. The only student academic achievement data that can empower teachers to improve student learning is accurate data. If the data does not measure what the educators understand it to measure, the decisions made on the basis of the data will be faulty. The specific psychometric analysis that assesses whether a test measures what it is designed to measure is construct validity. Brown (2000, p. 9) defined a construct as “an attribute, proficiency, ability, or skill that happens in the human brain” and characterized construct validity as “the experimental demonstration that a test is measuring the construct it claims to be measuring.” Construct validation of a test shows what the test measures as well as what the test does not measure (Brown, 2000; Nolet & Tindal, 1990). Brown (2000) additionally discussed the
dependence of test validity on test reliability; without demonstrated test reliability, test validity cannot be evaluated.

While each standardized test used in elementary and secondary education must be analyzed individually for construct validity, verifying that standardized tests are capable of actually assessing student learning is the crucial point to the overall research question. Nolet and Tindal (1990) researched the construct validity of thirteen commonly used standardized tests including the Iowa Test of Basic Skills, the California Achievement Tests, Metropolitan Achievement Tests, Stanford Achievement Test Series, and the Woodcock-Johnson Psychological-Educational Battery. The assessments used for the study were in the subject areas of reading, language arts, listening comprehension, and mathematics. The test items from the various assessments were qualitatively classified into nine categories by data coders who were university graduate students in education. The inter-rater agreement was above .90 and the correlation of coding of items into the categories of intended measurement was .96. The tests were then categorized and examined for patterns of inter-correlation. After the qualitative work of categorization and pattern identification, the data was quantitatively subjected to correlation analysis.

The mixed method research found twelve of the thirteen tests examined by the categorization process to assess the subject matter the test was designed to test at a level of .90 or above, as in a mathematics test assessed mathematics constructs. Many of the tests earned a perfect 1.00 score of correlation between stated and qualitatively identified subject matter. The methodology used by Nolet and Tindal for the study was well designed for the limited focus of the study and the relevance of the work to the question of whether standardized tests measure student learning is twofold. First, the research dispelled the view sometimes voiced by educators
that all standardized tests are really reading tests. Secondly, the work demonstrated that objective coders agree that standardized tests can be designed to evaluate specific content areas as planned by the creators of the tests.

In a corollary study, Klein et al. (2009) conducted the Test Validity Study to quantitatively compare various standardized assessments used in postsecondary education. Thirteen different tests were administered to 1,051 college freshmen and seniors from thirteen colleges. The students volunteered, and although the schools were encouraged to seek a sample representative of the general student body, most used a first come, first serve procedure. The students were each given a $150 gift card to Amazon.com for their participation, with financial support provided by the Fund for the Improvement of Postsecondary Education. The selection and incentive processes may have skewed the representativeness of the test group as decisiveness and interest in online shopping may have been favored traits. The grade point average of the sample group was slightly above the average grade point average of the student bodies at the involved colleges. The researchers hypothesized that high achieving students tended to volunteer for optional programs such as the Test Validity Study. The gender and ethnic distribution of the tested students aligned with the distribution of the general student body at each school. Although it would be very challenging to choose thirteen colleges that formed a sample representative of all colleges, the schools “varied in size, average college admission test scores, geographic region, control (public or private), and other characteristics” (Klein et al., 2009, p. 2).

The authors of the Test Validity Study had a vested interest in the research work, as each worked for a publishing company that produces commonly used standardized tests. The three companies represented both by the authors’ affiliations and by the provision of the tests for the study were ACT Incorporated, Council to Aid for Education, and Educational Testing Service.
The researchers from the three companies “worked collaboratively and collegially in designing the study, analyzing the data, and interpreting the results” (Klein et al., 2009, p. 2).

Nine of the standardized tests used in the study were multiple choice format and four tests were composed of constructed response questions. The content areas of the thirteen tests were: two reading, two mathematics, one science, four writing, and four critical thinking. All thirteen tests were used at all thirteen schools, with each student taking a package of tests over three days. The total time invested by each student was approximately six hours. The order of the tests was controlled for by the research design which started one third of each group of students on a different test.

The results were analyzed for test reliability and validity using both student-level data and school-level data. The mean reliability was .87, with the lowest value at .75. The mean school-level correlation for the multiple choice tests was .92 and the mean school-level correlation for constructed response tests was .84. The overall mean school-level correlation for all tests was .85. Higher levels of statistically significant correlation existed between same format tests of the same construct, such as two multiple choice mathematics tests. The basic conclusion from this data was that tests in the study demonstrated an appropriately high degree of reliability and strong validity to each other regardless of the test publisher, the format, or even the subject matter.

The scores of the freshmen and seniors were contrasted to examine construct validity. The seniors had higher mean scores on all tests but one mathematics test. When the outlying mathematics test was excluded and the score distribution among the test was controlled for, the seniors outscored the freshman by nearly .5 standard deviation unit. The higher scores achieved
by the upperclassmen support the validity of the testing because more college coursework resulted in higher test scores, therefore the students must be learning material in their courses.

Although the Test Validity Study examined postsecondary education, rather than elementary or secondary education, the strength of the validity measurement provides evidence that some standardized tests measure student knowledge. The Test Validity Study was recent work, conducted in fall 2008 and published in September, 2009. The recentness and breadth of the work enhances the relevance to current research regarding standardized test data.

**How Much Impact Do Teachers Have on Student Test Scores?**

One assumption commonly made by proponents of empowering teachers with data for increased academic achievement is that individual teachers can actually make a marked difference in student achievement. Nearly any educator, parent, or student would state that some teachers are better than others; the idea that teachers make a difference is a bedrock belief to a teacher’s very identity. However, what if student achievement is due solely to the students’ actions and decisions rather than the teacher’s? What if anyone with a basic understanding of the content could teach and have students produce the same results on standardized tests? As counterintuitive as this line of questioning is to educators, the research on the impact of teachers on student achievement is contradictory to the current practice of salary steps in education.

Rockoff (2003, p.1) noted, “School administrators, parents, and students themselves widely support the notion that teacher quality is vital to student achievement, despite the lack of evidence linking achievement to observable teacher characteristics.” Hanushek (2003) found that teacher education is rarely positively correlated at a level of statistical significance with student performance. Studies correlating teacher experience positively with student performance show much less frequent statistical significance than insignificance. Hanushek and Rivkin (2006, p.
11) further stated, “Perhaps most remarkable is the finding that a master’s degree has no systematic relationship to teacher quality as measured by student outcomes.”

Rockoff (2003) studied the impact of individual teachers on student achievement by analyzing twelve years of elementary standardized test scores in two New Jersey school districts. The participants were teachers at the elementary schools in the two districts. The total number of teachers or students analyzed for each teacher is never given in the research report and the researcher cited reasons of confidentiality for the lack of details. The median student in the study was tested three times with standardized testing during the course of the study and over 25% of the students were tested five times during the study. The median teacher in one district was assessed on three years of student data and the median teacher in the other district was assessed on five years of student data. The selection of the school districts was based on the commitment of these districts to avoid tracking students by ability or achievement, which was critical to an examination of teacher impact on student achievement. The school districts were described as having slightly above average socioeconomic status, approximately one third of students qualified for free or reduced lunch, and the per pupil spending was near the state average for the two districts. The teachers who were the subjects of the study were purposefully chosen to allow the question of teacher impact on student achievement to be examined without obvious flaws such as ability tracking or multiple teachers per student per year, as in typical secondary education. Despite the thoughtful selection of study subjects, Rockoff seemed to aim for teachers at an average school in New Jersey. All teachers of reading and mathematics in elementary schools in the two studied districts were included in the data and participation was involuntary. Given the probable number of teachers and students represented in the data and the basically average status of the school, according to the meager information supplied by Rockoff, the
subjects seemed representative of the general population of elementary teachers in New Jersey. If anything was remarkable about either school district or any part of the studied population of teachers, Rockoff did not include it in the research article.

The study was quantitative in design. The test data of four subject areas on a national standardized test were used. The subject areas were reading vocabulary, reading comprehension, mathematics computation, and mathematics concepts. Student scores were reported on a normal curve equivalent scale. The student scores were linked to their teacher. Student information such as school, grade, gender, ethnicity, special education classification, and primary language, along with teacher information such as highest degree earned, teaching experience and age were put into an equation on one side with the student test score on the other side. Rockoff tested for assumptions and collinearity prior to performing extensive regression analysis.

The results showed statistically significant differences among teachers. To calculate teacher quality, Rockoff used the differences between the 75th and 25th percentile teachers on the national achievement distribution of scores for the standardized tests. He noted that the difference between the percentile teachers was one normal curve equivalent, which equated with a .05 standard deviation. Rockoff noted that several measures of teacher quality have been used in educational research and his method was based on one used by Bertrand and Schoar (2001) and was chosen for its simplicity and transparency. By applying the method to his data, Rockoff concluded that one standard deviation increase in teacher quality correlated with a .20 standard deviation increase in student reading test scores and a .24 standard deviation increase in student mathematics test scores. There was no statistically significant correlation between teachers having a master’s degree and student achievement on the standardized tests.
Rockoff concluded that teachers differ in quality. However, traditional measures to improve teacher quality such as graduate education and teaching experience do not seem to work. In the end, Rockoff does not concretely link teacher quality to any specific factor measured in his research study. While this study was steeped in high level statistics and lacked definite conclusions regarding increasing teacher quality, the demonstration that teachers differ in quality and certain teachers correlate to higher student achievement is important. The assumption that teachers make a difference is at the center of teacher empowerment with data leading to increased student achievement. Rockoff (2003) established that teachers can impact student achievement at a statistically significant level, thus removing the assumption and grounding the work on teacher data empowerment on previous research.

*To What Depth Can Teachers Understand and Analyze Data?*

Teachers are a diverse group of professionals. On one hand, nearly all teachers have certain experiences in common such as earning a bachelor’s degree in education, coursework in pedagogy, and professional status in the workplace and community. On the other hand, within the profession, teachers specialize in subject matter, in the age of the students they teach, and even in the roles they fulfill in the school district such as technology advocate or coach or student government advisor. Until data analysis becomes a required part of undergraduate coursework for educators, it should not be assumed that teachers understand and can analyze academic data, especially as a basis for instructional decision making.

Henning (2006) studied seventeen elementary teachers and seven middle school teachers who enrolled in a teacher leader program at the University of Northern Iowa. The teachers had to be recommended by their principal for the program and received graduate credit for the coursework that formed the basis of the study. All of the teachers were seasoned classroom
educators and most were identified building or district leaders. The stated focus of the study was to compare the efficacy of different data analysis strategies in terms of instructional improvement.

The study participants were not representative of the general population of all educators or even all classroom teachers. However, given that not all teachers are leaders and that it tends to be the teacher leaders that drive school improvement, the study participants may be representative of the population of teachers that typically does the majority of the work to improve instruction through data analysis. Interestingly, the study lacked high school teacher representation, which may also be a valid portrayal of teachers who typically spend time attempting to improve instruction through educational coursework such as the University of Northern Iowa program in the study, which did not lead to an administration degree. High school content teachers have the option of taking content specific coursework as opposed to elementary teachers who typically undertake pedagogical focused coursework.

The study methodology grouped the teachers by school, which resulted in the formation of eight teams. The selection of teachers for the work and the grouping of teams by school were qualitative in design. Henning and his colleagues educated the teachers in four basic academic data analysis techniques: comparing to the norm, analyzing trends, correlating data, and disaggregating data. The study required each team of teachers to choose three analyses to perform. The teachers were directed to use one data analysis to examine academic achievement in their school building, one analysis to compare student subgroups in the school, and one to focus on a specific content area. Beyond those instructions, the teams of teachers were free to choose the analysis technique. The data the teachers analyzed were their own school’s
standardized test scores from the Iowa Test of Basic Skills. The study generated 24 analyses, three from each of the eight teams.

The researchers overtly directed the teacher test subjects in the study. In teaching the course, Henning and his colleagues chose which data analysis methods to include as part of the course, how to present the data analysis lessons including the presentation of potential uses of the analysis methods, and the directives about the goals of the data analysis assignment required of the teacher test subjects. While the overt direction skewed the quantitative validity of the choice of data analysis methods by teachers, it was necessary to fulfill the purpose of the course for the teachers and the schools involved in the program.

The only analysis of the study data was to classify each of the 24 data analyses generated by the teams of teachers. The analyses were qualitatively categorized as one of the following: correlating, disaggregating, trend analysis, trend analysis compared to the norm, or trend analysis of disaggregated data. Three of 24 analyses conducted by the teams of teachers correlated data, three analyses disaggregated data, five analyses analyzed trends over two to four years, three analyses analyzed trends as compared to the norm, and ten analyses analyzed trends in disaggregated data. Eighteen of the 24 analyses involved data from more than one year, but zero used five years or more of data.

The study of how teams of teachers analyzed data was designed as a quantitative project, but appeared as a mixed method approach in the discussion section of the research paper. Henning presented the results quantitatively as described above, but the first sentence of the research paper discussion section stated, “The purpose of this study was to provide a description of how a group of teacher leaders analyzed standardized achievement test scores in order to improve instruction” (Henning, 2006, p. 735). Henning provided a numerical description
followed by qualitative understanding from the study such as “the insights provided by school achievement data are often revealed through visual display of data” and “engaging with statistical concepts relevant to comparing percentile rankings should be an important part of coursework” (Henning, 2006, p. 736). These conclusions were not linked to the results of the study in the research paper. Prior to the discussion section, the research findings were presented in strict quantitative format.

Henning had an ideal set up with the teacher leader data course for a qualitative study to examine the evolution of professional understanding of academic data, in which case he should have conducted surveys or interviews of the teacher participants throughout the yearlong project. Henning could have quantitatively studied the standardized test scores of the students of the teacher participants for the three years before and the three years after the course to examine the effect of equipping the teacher with data on student achievement. Either design modification could have produced a more powerful research study and data that is more relevant to the overall research topic of this paper. As performed, the merit of Henning’s work was described by Henning himself in the final sentence: “Each new description adds another model of practical application for the benefit of teachers, principals, and professors who are interested in making principled decisions based on standardized achievement data” (Henning, 2006, p. 736). The study conclusively showed that teachers can understand and analyze academic data to a degree that could be useful for data-based decision making.

What Conditions Empower Teachers to Make the Changes Indicated by Data?

Even if standardized tests measure student learning, the actions of teachers can significantly impact the test scores, and teachers can understand the data and decide to make changes based on the data, the whole process is meaningless if teachers do not have the power to
make the necessary changes they derive from the data. The question of teacher power and authority touches upon many facets of education. Some teachers may feel that the state content standards prescribe curriculum to the point that teachers are powerless to make instructional decisions. Other teachers may work in a school culture where teachers feel powerless to make any changes. The data may indicate changes that need administrative initiative or at least administrative support. The issues of teacher empowerment vary from situation to situation and even teacher to teacher. While some individuals may seek positive change by communicating with parents, students, and administrators, other teachers may shun the added work or avoid the potential conflict pursuing change can bring. The framework for teacher empowerment is an important concept strongly related to the work of this project.

Lichtenstein, McLaughlin, and Knudsen (1991) studied the relationship between teacher empowerment and professional knowledge in a qualitative research project. The researchers noted that their sample pool of teachers was not representative of all teachers. The thirty participants were all secondary math teachers in California during a time when California was implementing a new state-mandated math curriculum and therefore teachers “interest in examining current practice and new ideas may have been artificially high.” Also, the teacher participants were chosen for their positive involvement in the Urban Mathematics Collaborative, which was an ongoing project to establish professional networking among mathematics educators.

The thirty teachers were interviewed about their perspectives on the sense of community and the evolution of empowerment due to the Urban Mathematics Collaboration. Excerpts of interview answers form the data presented in the paper. The interviews were examined for patterns of thought among the participants as well as unusual, but pertinent, reflections that
brought out different perspectives on empowerment. The interview responses revealed three categories of knowledge that teachers linked to empowerment. One conclusion of the researchers was that “knowledge is an elemental, irreducible aspect of teacher empowerment” (Lichtenstein, et al., 1991, p. 3). The three categories of knowledge were: knowledge of professional community, knowledge of education policy, and knowledge of subject area. Under knowledge of professional community, the teacher participants discussed the network of math teachers that the Urban Mathematics Collaboration fostered. Although the teachers in the study resided in California, they were networked with teachers from all over the United States. The network of other teachers provided a valuable resource of lesson strategy ideas and the network combated feelings of isolation for the study participants. Having other teachers to go to with questions and new ideas empowered the study participants to take new risks in instructional practices. Also, being asked for their opinions by other math teachers empowered the study participants by validating their expertise as math educators. One goal of the Urban Mathematics Collaboration was to educate teachers about policy changes such as the adoption of the new math curriculum in California. The study participants voiced appreciation for the notification of policy changes and furthermore, discussed educational policy with other teachers in their network. The study participants felt empowered by the knowledge of current policy and then felt further empowerment when they began to use their understanding of policy in influencing textbook selection at their schools. “The result of teachers’ policy knowledge is empowerment – a sense of authority and a belief that they can make a difference” (Lichtenstein, et al., 1991, p. 12). Knowledge of subject matter empowered teachers in their own classrooms. The teacher participants desired to be creative in reaching students with math curriculum and they identified
a “deep and broad knowledge of mathematics” as vital to attempting creative instructional strategies (Lichtenstein, et al., 1991, p. 13).

As with any interview process, data can be skewed by the questioning technique. If the teacher participant understood the focus of the research project from the questions asked, they may have tried to provide information they thought the researchers wanted. Lichtenstein, et al. (1991) demonstrated that teachers could be empowered and that knowledge, especially knowledge of professional community, knowledge of education policy, and knowledge of subject area, is a crucial source of empowerment.

Another conclusion of Lichtenstein, et al. (1991) was that structural reform in education hinges on the teachers’ abilities to adequately fulfill the responsibilities placed upon them by the reform effort. A study done by Kadel-Taras (1996) built upon that conclusion. Kadel-Taras observed and interviewed nine teachers with the intent of learning how the teachers made sense of the changes in instructional practices they had undertaken. The study was entirely qualitative, and the participants were not representative of all teachers since they all taught at the same high school. The conclusion of the work was that the teachers’ needs for their own learning and control in their classrooms must be met in order for change to occur. The focus of the study was related to the framework necessary for teacher empowerment. The nine participating teachers were seeking change in their own instructional practices of their own volition, rather than change mandated by an external authority. The teachers felt empowered to make changes as long as they devoted enough energy to getting their own needs met in the process.

To What Extent Does Empowering Teachers with Data Improve Student Learning?

The overall research question has also been examined in some studies. Bernhardt (2009) presented a school team approach to analyzing data. Six teachers, the principal, and the district
data analyst from Marylin Avenue Elementary School in Livermore, California attended a summer data institute in July, 2006. The eight educators were voluntary participants in the project. The summer data institute was apparently attended by other educators from other schools, but Bernhardt chose to focus the qualitative case study on the Marylin Avenue Elementary attendees. The participants are not intended to be representative of the general population of educators. The demographics of the students at Marylin Avenue Elementary include 66% Hispanic student population, nearly 76% of students receive free or reduced lunch, 507 total students in grades kindergarten through five, and a significant trend of increase in each of the aforementioned demographic categories over the past five years. Furthermore, the school had failed to make adequate yearly progress for four years and the institute attendees from the school “perceived that the school culture was not ready to change” (Bernhardt, 2009, p. 24). These factors may have contributed to choice of participants in this case study.

At the institute the study participants engaged in intensive data scrutiny to understand their current status, how the school achieved its current status, gaps demonstrated by data, and what was working and what was not working. The team self identified the issue of changing student demographics and stagnant school and instructional procedures from student demographic data. Academic data analysis, including extensive disaggregation, allowed the team to understand where instructional coherence was lacking and which specific student subpopulations were underachieving. The data understanding gained at the summer institute spawned widespread changes during the following school year including the creation of a vision shared by the entire staff, more data analysis done by the entire staff, professional development sessions to improve instruction and assessment, collaborative teams, and a school portfolio to house and track progress toward the vision. Essentially, the experience of the institute equipped
and empowered the eight educators to lead a comprehensive school reform initiative during the following school year. The team that attended the training institute directed the process back at school without input from trainers once the summer institute was completed. Therefore, the case study was designed to assess school academic achievement data, provide intense professional development to a team of educators, wait one school year, and then assess school academic achievement data again. When Marylin Avenue Elementary School showed tremendous gains in standardized test scores, the team that had attended the institute were contacted and interviewed for details about the improvement. The team was also invited back as featured speakers to the summer institute for the next summer and the project was written up as a case study.

While quantitative data was used to select the school highlighted in the case study, the study itself was about the data based professional development of the educators and the actions that resulted from the professional development. Although many researchers may view the work of Bernhardt as simply sharing a success story rather than actual research, especially given the method of research subject choice, her article is truly a case study that shows one model for empowering educators with data. Bernhardt continues to follow the progress of Marylin Avenue Elementary staff and students. The report included two years of academic achievement data, which continued to increase, and staff actions, which continued to move toward their vision, following the pivotal summer institute experience. As with any interview data selected to show success, people tend to skew the information they provide toward the expected success. Only the staff and students at Marylin Avenue Elementary School know the changes in school culture that have occurred. Knowing that they had been chosen for their success may have enhanced the progress of positive change in school culture.
As with most qualitative studies and school success stories, school improvement can be attributed to one change such as data examination and the educators involved can link the improvement to that specific change, but in reality several changes occurred during the year at Marylin Avenue Elementary School. One could argue that the process of forming a shared staff vision or the formation of collaborative learning teams could have been the main cause of the improvement. Perhaps the data training was merely the spark that ignited the fuse for school improvement. Perhaps any intense experience that congealed those eight educators into a team would have led to the transformation of their school. A direct cause and effect relationship between data use and student academic achievement or even school improvement cannot be stated from Bernhardt’s case study.

Bernhardt concluded that educators need to examine all demographic and student achievement data to understand the big picture view of their school and that schools have the ability to improve through data analysis. Bernhardt’s small case study is one of the few research articles that link teachers’ use of academic achievement data to changes in instructional practices and increased student achievement.

In an earlier article Bernhardt (2003) shared two specific instances of high level data analyses by educators that revealed unusual patterns interfering with student achievement. At one school, 60% of third graders scored below proficient on a standardized reading test. In analyzing three years of disaggregated data, the teachers noted that one ethnic group scored consistently lower than other ethnic groups. Further analysis linked specific teachers to low scoring students of the ethnic group. In fact, “students who were of this particular ethnicity had never scored at or above the proficient level in three years” (Bernhardt, 2003, p. 28). Bernhardt went one step further and examined student perception data collected from surveys. Students of
the ethnic group with the specific teachers in question scored very low in their responses to questions about whether their teacher believed they could learn and be successful. The poor performance on a reading test was linked to the personal beliefs of teachers as perceived by their students. The school made personnel changes and adopted some new school procedures to align curriculum and measure progress objectively in classrooms. Whether the changes resulted in better reading test scores was not stated in the article.

Bernhardt’s second example of high level data analysis occurred at a high school where more than half of the ninth graders failed the state reading test. Educators chose to track student data back through elementary school to look for an early commonality among the failing students. The data revealed that many of the failing ninth graders had excessive absenteeism in first grade. The school district implemented extensive measures, including home visits and communication plans, to address the problem in early elementary. Again, the impact of the measures was not reported in the article, perhaps because it would take several years to see the results of the intervention. Both examples support the concept that educators can use data to make significant changes in education.

Two other data use stories echo the successful experiences described by Bernhardt (2003; 2009). Oklahoma’s Western Heights School District was featured in an article by Aarons (2009). The superintendent of the district established a longitudinal data system for student data in 2001. The article lacks enough description of the process to be analyzed as a case study, but the reported results include a decrease in dropout rate and teachers using the academic data to change instructional practices. Further support is offered by Anderson, MacDonald, and Sinnemann (2004) who examined educational data in British Columbia, Canada as a large scale case study. Unfortunately, in regard to relevance to teacher empowerment, the focus of the
reported results is a decrease in dropout rate among aboriginal students. While this is of limited relevance to the question of empowering teachers with data, the British Columbia case study is another example of data collection and analysis being correlated with student behavior by researchers.

The reviewed literature addressed the assumptions often made regarding empowering teachers with data. Klein et al. (2009) showed that standardized tests can have construct validity and can measure student learning to a useful degree for educators. Rockoff (2003) quantitatively supported the basic belief that teachers differ in quality. The work done by Henning (2006) demonstrated that teachers can understand and analyze data to a level that is useful for making decisions regarding the effectiveness of instruction. Empowering teachers to make changes that increase test scores is supported by the research of Aarons (2009), Anderson, et al. (2004), Bernhardt (2003; 2009), Kadel-Taras (1996), and Lichtenstein, et al. (1991).
Chapter III: Results and Analysis Relative to the Problem

The presumption of accountability advocates is that achievement data can and will increase student test scores. The logicality of understanding the current academic status of students as a first step to improving said academic status is a widely accepted strategy. Additionally, the contemporary availability of student data, especially post-analysis, due to technological advancements such as data warehousing, paired with the call to improve education, are the foundational pieces of the accountability and standardization movement enhanced by NCLB (2001). The overall picture of data based decision making may be appealing, but the transformation is a step by step process. Each and every step must be solidly supported by well designed research for the end result the NCLB sponsors envisioned.

To What Degree Do Standardized Tests Measure Student Learning?

The weak link in the research support chain leading to the overall research question of whether empowering teachers with data increases student learning seems to be the validity of the measurement of student learning. The research literature regarding construct validity is rife with disagreement and insufficiency. At the root of the issue is debate over the very definition of construct validity. Brown (2000), who provided the definition of construct validity for this research, also discussed a movement in psychometrics to use a broader view of construct validity as a measure of overall validity that includes content validity and criterion validity. Researchers agree that validity is complex (Burns, 1998; Nolet & Tindal, 1990) and can be measured in different ways (Brown, 2000).

Nolet and Tindal (1990) used qualitative methodology to show that standardized tests measured the subject matter they claimed to measure. Nearly twenty years later Klein et al. (2009) used quantitative methods to comprehensively demonstrate that standardized tests
measured student learning in specific content areas. Both studies claimed to assess construct validity and, technically, both accomplished the goal. Beyond the differences in approach, the significant difference between the studies was focus of the conclusion. Nolet and Tindal (1990) categorized test questions without including student test scores in their data. Therefore, the demonstration of construct validity was based solely on the wording of the questions. The conclusion of the work was that a test assessed the stated subject matter, such as a mathematics test measured mathematics ability, not language arts ability. The research was important to establish that standardized tests are not all reading tests. Many critics of standardized testing, especially teachers, have stated that such tests only assess reading ability (Visone, 2009). However, in Nolet and Tindal’s (1990) study content area and standardized testing are strongly correlated.

While Nolet and Tindal established the link between the constructs and the test, that is only one facet of construct validity. The other critical facet is connecting the test scores to student learning. The research of Klein et al. (2009) connected student learning and standardized testing as college seniors outscored college freshmen on standardized tests. Construct validity is “the experimental demonstration that a test is measuring the construct it claims to be measuring,” (Brown, 2000, p. 9). There are two components in the definition. The first component is that the test measures the construct it claims to measure; therefore a construct validation must be performed on the test itself. The second component is the experimental demonstration portion of the definition. In the case of educational standardized testing, a construct validation must be performed using student scores and some other achievement data to corroborate those students who score well know the construct and students who do not score well do not know the construct. Both aspects must be addressed to show construct validity of a test.
How Much Impact Do Teachers Have on Student Test Scores?

Rockoff (2003) demonstrated that teacher quality impacts student test scores. The description of factors that improve teacher quality is a topic for further research since the study showed that teacher experience and education are less correlated with student test scores than was expected. However, the actions of the teacher can impact the academic achievement of students on standardized tests.

To What Depth Can Teachers Understand and Analyze Data?

Henning’s (2006) study showed the ability of teachers to be trained in data analysis methods and to understand the meaning of student achievement data. Bernhardt (2009) further supported the conclusion that teachers can be effective users of data.

What Conditions Empower Teachers to Make the Changes Indicated by Data?

The two empowerment studies presented in the literature review chapter clearly indicate that teachers can be empowered to make changes. Lichtenstein, et al. (1991) researched teacher empowerment in mathematics teachers participating in a professional networking project and demonstrated knowledge of professional community, knowledge of educational policy, and content knowledge as conditions that empower teachers. Kadel-Taras (1996) studied teacher empowerment in high school teachers seeking to improve instruction on their own and showed that teachers must have their own needs met in order to feel empowered.

To What Extent Does Empowering Teachers with Data Improve Student Learning?

Each of the process research questions have been answered in the affirmative by research: Standardized tests can measure student learning (Klein et al., 2009). Teachers can impact test scores (Rockoff, 2003). Teachers can analyze and use data (Henning, 2006). Teachers can become empowered (Lichtenstein, et al., 1991). Therefore, theoretically,
empowering teachers with student data from standardized tests can improve student learning. Furthermore, educational leadership journals offer evidence of empowered teachers making changes and student learning increasing (Aarons, 2009; Bernhardt, 2003; Bernhardt 2009). The success stories tend to be case studies of one school or one district and, typically, an administrator, working with teachers, led the charge to improve. The research supports the verdict that empowering teachers with student standardized test data can improve student learning.
Chapter IV: Recommendations and Conclusions

Recommendations

Teachers or administrators who desire to use student standardized test score data as a basis for instructional reform would benefit from following the model offered by Bernhardt (2009) and supported by Henning (2006). To feel truly empowered to make decisions from data, teachers and administrators need knowledge (Lichtenstein, et al., 1991), so training in the understanding and analysis of data is the beginning step. The interested teachers and administrators in the Bernhardt (2009) and Henning (2006) models attended an intense data training session and worked in teams of educators from the same school. Because educators worked with the data generated by the students at their own school, the workshop was not only training in data understanding, but also an opportunity to analyze their own data in the presence of outside data experts who could be called upon for help as needed. In the case study (Bernhardt, 2009), the empowered educators returned to school motivated to work with the rest of the staff to create a vision, use collaboration as the mode for changes in school culture and establish a new focus on instruction and student learning. Administrative participation seems critical to the model. In the case study the school principal was a member of the team at training and strongly supported the data based reform from his leadership position.

The recent application of data warehousing to education provides data analysis tools for educators motivated to use student data. In Michigan large grants were provided to several regional educational agencies within the last few months to set up and train teachers and administrators to use an educational data warehouse. The process of setting up and training educators is just beginning in many areas of Michigan. Calhoun Intermediate School District and the Eastern Upper Peninsula Intermediate School District are the consortium leaders for the work.
of moving school districts into data warehousing because they have been using data warehousing for the past five years (M. Ribant, personal communication, October 16, 2009). Tools provided by data warehousing can disaggregate data by subpopulations, track multi-year trends in school data, and even show growth at the individual student level. Investing the time to learn a data management system would be very beneficial to educators who desire to regularly evaluate data.

The reviewed research indicates that leaders wishing to begin the process of empowering teachers with data should begin by identifying interested educators to form a data team. Ideally, the team would include three to six teachers and one administrator each with a strong interest in academic data. The team should undergo comprehensive training together to learn to use the data analysis tools that are available. The training activities should include an analysis of the schools own data and should be led by an expert in both education and data analysis. Training done by data warehouse sales representatives who lack a background in education is of limited value; similarly, training done by educational consultants who lack knowledge of the data management system used by the trainees is also of limited value.

After the data team has the knowledge and experience of interpreting the data from their own school as provided by the training, the team should create a plan for educating the rest of the teachers and moving the school forward into data based decision making. Each school situation is unique and therefore each data team’s plan will be different. Typically data teams lead professional development sessions to either educate their colleagues in their own school data that has been analyzed by the data team or in the processes of using the data analysis tools. Some administrators have expectations that all teachers will become well versed in using data and data analysis tools, but that will not be realistic for all teachers. Schools that employ a collaborative team approach can include all instructional staff members in using data to make changes aimed
at improving student achievement, even if some staff members lack data literacy. The key to overcoming a lack of data literacy among staff members is to place a data enthusiast, ideally one of the initial data team members that attended the comprehensive training, on each professional learning team. Educators do not have to be data analyzers to be data based decision makers as someone can present post-analysis data to teachers. However, having someone on each professional learning team that can use the data tools to disaggregate data in various ways is a useful advantage. As with any educational initiative, the success or failure of the program rests on the communication skills and the efforts of the people leading the charge to change.

The recommended role of the school principal during the transformation to empowering teachers with data is twofold. First, the principal is responsible for upholding or creating a school culture that fosters teacher empowerment and instructional change if needed. Second, the principal should fully participate in the transformation as a member of a collaborative team. In some situations teachers may assume the lead role in the transformation process, but generally the principal participate in the leadership of the data based reform. Great educational leaders know when to lead and when to get out of the way. Once the teachers are empowered to make the instructional decisions needed to serve all students and close the gaps, educational leaders become resource people who support the work of the empowered teachers.

A reasonable timeline for implementation would be one semester to identify the group of teachers who are willing to be trained and then have them go through the training as a group. The next semester should be the introduction of the transformation to the staff and the work of training the entire teaching staff. This timeline is dependent on the availability and set up of the intense team training. Studies cited in the reviewed literature used training offered for a week in the summer, training offered over a university semester, and training done over a couple of years.
The timeline should enough time for teachers to absorb and reflect upon the information, but not so much time that the lessons from the data are of less use or the data team’s enthusiasm wanes. A one year focus on the training and implementation would begin the initiative and offer enough time for teachers to cognitively process the transformation. Once successfully implemented, school administrators must allow teachers the time to continue the process of data examination and instructional refinement each year. When the data begins to show improvement in student achievement, the success of the teachers and students should be celebrated.

Areas for Further Research

To further the academic body of knowledge regarding the relationship between student learning, standardized test data, and teacher practices, the areas of standardized test validation and teacher empowerment with data require more extensive research. Evaluating standardized tests for construct validity requires that the test itself be evaluated and then student test scores be examined to show that students who know the construct get significantly higher scores than students who do not know the constructs. Construct validation is time consuming, but must be done to insure the accuracy of the scores and the usefulness of the standardized test data to educators.

Currently each state may determine which standardized test to use. Michigan uses the Michigan Educational Assessment Program (MEAP) to assess students and provide accountability data as required under NCLB. Even though each test item on a MEAP test is based on the state content standard the item is intended to assess, construct validity should not be assumed. Burns (1998) reviewed the validity of the Michigan Educational Assessment Program (MEAP) and found the test lacking in construct validity. Ten years later, the Michigan Department of Education addressed the reliability and validity of the MEAP tests in the District
and Building Coordinator Handbook for the 2008-2009 Academic Year. While the authors of the handbook assured “the reliability of MEAP assessments and the inter-rater reliability of the scoring process meet high technical standards,” (Michigan Department of Education, 2008, p. 9), they did not state any assurances or measurement of validity. The handbook authors explained and discussed psychometric validity in some detail without providing actual information about the content, criterion, or construct validity of the MEAP assessments. Corwin (2001) commended the MEAP as “one of the most mature State assessment systems in the country” that exceeds the NCLB requirements.

If Michigan’s standardized testing program is one of the better state-run programs and construct validity is an issue, the construct validity of the tests used in many other states comes into question. While asking psychometricians to analyze the construct validity of all standardized tests used for NCLB accountability is a tall order, if educators are to base decisions on student standardized test scores, the tests must measure what they claim to measure. A test validity study similar to the one done by Klein et al. (2009) should be done to evaluate objectively standardized tests used in elementary and secondary schools.

The other vital area of further research is to address the overall research question. To date, studies regarding how empowering teachers with data affects student learning are sparse and are primarily case studies of successful school reform efforts. As stated earlier, well designed quantitative studies to show the difference in student test scores taught by teachers with access to standardized test data versus teachers without access to standardized test data would be an excellent area of further research. Perhaps more beneficial to the actual realization of the goal of teachers making data-based decisions about instruction, would be qualitative studies that provide information about the necessary framework for teachers to understand and use data. The
case studies done to date suggest that administrative support may be important, as noted in Bernhardt (2009) and Aarons (2009).

Summary and Conclusion

The future of education is standardization which ideally will eliminate academic achievement gaps in subpopulations based on ethnicity, gender, or special education status. The process has been initiated by states setting content expectations and requiring standardized testing of the content expectations. The current ideology embraced by many educators, which was also the overall research question addressed in this literature review, is empowering teachers by means of standardized test data with the expectation of improved student learning and gap elimination.

While the concept of teacher empowerment with data improving student learning is enticing, educational visionaries must not assume or overlook foundational pieces such as: To what degree do standardized tests measure student learning? How much impact do teachers have on student test scores? To what depth can teachers understand and analyze data? What conditions empower teachers to make the changes indicated by data? A review of literature for each break-out foundational question demonstrated that each break-out question is affirmatively supported by competent research. Since standardized tests can have the construct validity to show that student learning is measured, teacher quality can significantly impact student test scores, teachers can learn to perform in-depth data analysis and identify subpopulation gaps, and teachers can be empowered to make changes, the overall research question is theoretically supported by research. Additionally, the literature review included a few case studies that showed improvement in student learning with teacher data empowerment.
Much more research work is needed in the areas of standardized test validation and examining models of teacher empowerment with data for the purpose of increasing student learning. The construct validity of standardized tests is the weakest link in the chain of support for the overall research question of the literature review. Studies regarding the connection between teacher empowerment with data and student learning are primarily small case studies of successful implementation. Much more stringent and comprehensive research is needed to fully demonstrate that empowering teachers with data makes an academic difference for students.
References


