

THE RELATIONSHIP BETWEEN TEACHERINSIGHT™ SCORES AND
PROFESSIONAL DEVELOPMENT AND APPRAISAL
SYSTEM DOMAIN SCORES

Michael T. Novotny, B.A., M.Ed.

Dissertation Prepared for the Degree of
DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2009

APPROVED:

Jane B. Huffman, Major Professor
Robin Henson, Minor Professor
Bill Camp, Committee Member
Carol Wickstrom, Interim Chair of the
Department of Teacher Education
and Administration
Jerry Thomas, Dean of the College of
Education
Michael Monticino, Interim Dean of the
Robert B. Toulouse School of
Graduate Studies

Novotny, Michael T. *The Relationship Between TeacherInsight™ Scores and Professional Development and Appraisal System Domain Scores*. Doctor of Education (Educational Administration), May 2009, 100 pp., 5 tables, references, 123 titles.

Many school districts and alternative certification programs use standardized interviews such as the TeacherInsight developed by the Gallup Organization. The TeacherInsight is a Web-based interview consisting of multiple choice and Likert-style items that produces a score between 0 and 100. The Gallup Organization claims that it helps hire the best teachers.

The study analyzed the relationships between the TeacherInsight scores and the eight Professional Development Appraisal System (PDAS) domain scores for 527 teachers.

The TeacherInsight scores produced a statistically significant correlation with only one of the eight PDAS domain scores. However, even that correlation ($r = 0.14$) was weak. All eight PDAS domain scores were only able to account for an additional 1.9% of the variance of TeacherInsight scores, above and beyond what was explained by the teachers' age, gender, years of experience, and highest degree earned. Another finding was that 47.4% of the teachers hired had TeacherInsight scores below the district recommended cutoff score of 67.

The findings do not support the ability of the TeacherInsight to identify more effective teachers, based on Professional Development Appraisal System scores. The findings also cast doubt on the extent of consideration that principals in this district give the TeacherInsight scores during the selection process. Recommendations for future studies are provided.

Copyright 2009

By

Michael T. Novotny

ACKNOWLEDGMENTS

First and foremost, I thank God. Everything I accomplish, including completing this dissertation, is only possible through Him.

I am very thankful for the support of my family. My three boys, Andrew, Ryan, and Thomas, and my beautiful wife, Laura, have demonstrated incredible patience, support, and love. I dedicate the completion of this project and degree to you. I also appreciate my parents for teaching me the value of education and for never giving up on me.

I would like to thank Dr. Jane Huffman, my major professor, for your patience, support, and timely feedback. Thanks to Dr. Robin Henson, my minor professor, for your expertise in research and statistics. Thanks also to Dr. Bill Camp for his feedback and for challenging me to develop a quality dissertation study. In addition to your help during the dissertation process, I also appreciate the three of you for everything I learned from you during my doctorate coursework.

I am grateful for the support of several employees of my former colleagues including Dr. Doug Otto, superintendent, Jeff Bailey, deputy superintendent, Tamira Griffin, executive director of human resources, Becky Wussow, director of human resources, James Ashby, director of assessment and accountability, and Jun Melvin, director of compensation and diversity. You are all outstanding educators, and it was a privilege and honor to work with you.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	vii
Chapter	
1. INTRODUCTION	1
Background	1
Statement of the Problem.....	7
Purpose.....	8
Research Questions.....	8
Methodology.....	9
Limitations	9
Definition of Terms	9
Significance of the Study.....	13
Organization of the Study.....	14
2. REVIEW OF THE LITERATURE	16
Introduction.....	16
Student Achievement and Accountability	18
Texas Accountability Rating System	20
Federal Adequate Yearly Progress.....	27
Teacher Effectiveness and Student Achievement.....	32
Teacher Evaluation	34
Professional Development and Appraisal System (PDAS).....	36
Teacher Selection	41
Standardized Interviews	44
Interactive Computer Information System-Urban (ICIS-Urban)	44
Principal Perceiver Interview (PPI)	45
STAR Teacher Selection Interview	45

Teacher Perceiver Instrument (TPI)	47
Urban Teacher Perceiver Interview (UTPI)	49
TeacherInsight.....	50
Summary	52
3. METHODS AND PROCEDURES	53
Purpose	53
Research Questions	54
Context	54
Study Participants	55
Research Design.....	55
Instrumentation	56
Data Collection.....	58
Data Analysis	58
Summary.....	59
4. PRESENTATION AND ANALYSIS OF THE DATA	60
Introduction	60
Descriptive Data	60
Description of the Population	60
TeacherInsight	61
Professional Development and Appraisal System.....	62
Data Analysis	63
Summary.....	72
5. SUMMARY, DISCUSSION, RECOMMENDATIONS, AND SUGESTIONS FOR FUTURE RESEARCH.....	74
Introduction	74
Background.....	74
Overview of the Study	76
Review of the Methodology	76
Summary of the Findings	76
Description of the Population	76

TeacherInsight (TI).....	78
Professional Development and Appraisal System (PDAS).....	78
Research Question 1.....	78
Research Question 2.....	79
Research Question 3.....	79
Research Question 4.....	80
Discussion of the Results	80
Recommendations.....	84
Suggestions for Further Studies	85
Conclusions	87
APPENDIX	89
REFERENCES.....	91

LIST OF TABLES

Table	Page
1. PDAS Domain Scores Descriptive Data.....	62
2. Correlation Matrix PDAS Domain Scores and TeacherInsight Scores	64
3. Correlation Matrix Teacher Characteristics and TeacherInsight Scores....	65
4. Correlation Matrix PDAS Domain Scores and Teacher Characteristics	68
5. Hierarchical Multiple Regression of PDAS Domain Scores on TeacherInsight Score	71

CHAPTER 1

INTRODUCTION

Background

Student achievement is affected by many factors. Educators cannot control many of these factors such as nutrition (Victora, Adair, Fall, Hallal, Martorell, Richter, & Sachdev, 2008), family's socioeconomic level (Duncan & Magnuson, 2005), student mobility (Alexander, Entwisle, & Dauber, 1996; Kerbow, 1996; Nelson, Simoni, & Adelman, 1996), mother's educational expectations (Bennerm & Mistry, 2007), neighborhood influence (Sampson, Sharkey, & Raudenbush, 2007) and parental involvement (Barron, 2007; Desimone, 1999). Barton and Coley (2007) found that approximately two-thirds of the large variance between states in eighth grade reading National Assessment of Educational Progress (NAEP) can be accounted for by the following four factors: single-parent families, parents reading to young children every day, hours spent watching television, and the frequency of school absences. Other factors that affect achievement that schools do not have control over include parent participation, birth weight, lead poisoning, and parent availability (Barton, 2003).

However, educators can influence many achievement factors including school size (Howley & Bickel, 2000), teacher experience (Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005; Kane, Rockoff, & Staiger, 2006), class size (Krueger, 1999) and curriculum and instructional strategies (Marzano, 2003). Other school-related factors that correlate with student achievement include rigor of curriculum, teacher preparation, teacher experience and attendance, class size, technology-assisted instruction, and school safety (Barton, 2003).

Researchers have determined that teacher effectiveness is one of the most significant factors related to student achievement (Goldhaber, 2002; Rivkin, Hanushek, & Kain, 2005; Sanders & Rivers, 1996). It has also been found that disadvantaged students are much more likely to be taught by less qualified teachers (Langford, Loeb, & Wyckoff, 2002).

Thus, teacher recruitment and selection is one of the most important responsibilities for school administrators. The process of teacher selection often involves several steps which include reviews of application, resume, portfolio, and transcripts, and references checks. Interviews by human resources staff, principals and other campus staff are also critical to the hiring of qualified teachers. This process is often very time-consuming for principals and human resource staff members. The challenge to hire quality teachers is amplified by the nation-wide shortage of teachers that has been reported for many years (Hussar, 1999).

In this study, the researcher investigated the predictive validity of the TeacherInsight™ (<http://www.gallup.com>). The TeacherInsight (TI) is a Web-based interview developed by the Gallup Organization that consists of multiple choice and Likert-scale questions. Based on the responses to the questions, applicants receive a score between 0 and 100. The Gallup Organization claims that a higher TI score is predictive of a more effective teacher (Gallup, 2008a).

Selecting effective teachers has become even more important and challenging in recent years. Federal and state laws are holding schools and districts more accountable for student achievement. The Federal No Child Left Behind (NCLB) Act requires students to be tested annually and for schools and districts to make Adequate Yearly

Progress (AYP) through increases in the results of these tests (Public Law 107-110, 2001). Many states also have their own accountability systems for student achievement. The Texas Education Agency rates schools and districts in one of four categories (Exemplary, Recognized, Academically Acceptable, and Academically Unacceptable) based on the Texas Assessment of Knowledge and Skills (TAKS) test scores and other factors such as attendance rate and completion rate. Students are required to pass the TAKS tests in 3rd, 5th, and 8th grades to be promoted to the next grade level. Students also need to pass all four 11th grade exit level TAKS tests to graduate. Thus, effective teachers are critical to enable students to be successful and to ensure that schools and districts meet state and federal accountability requirements.

There is often a shortage of teacher applicants in many districts and states across the country. More importantly, even when there is a sufficient supply of teacher applicants, there may be a shortage of quality teacher applicants. College graduates with higher ability, based on test scores and class rank, are less likely to go into teaching (Guarino, Santibanez, Daley, & Brewer, 2004).

The NCLB Act requires teachers of core subject areas to be *highly qualified*. To meet the definition of highly qualified, teachers need to have a bachelor's degree, state certification, and demonstrate proficiency in the subject area they teach. To demonstrate proficiency, teachers need to pass a competency exam or demonstrate their qualifications through a combination of college credits in the subject area, years of teaching experience, and professional development. Whether highly qualified teachers are more effective than those that are not is debatable as a study by Dee and Cohodes (2008) found mixed results. Regardless, this requirement has further exacerbated the

shortage of teacher candidates because many current teachers and new applicants are not able to meet the highly qualified definition in a particular subject area.

The teacher selection process varies considerably among districts and schools. Many districts use a centralized model in which teachers are hired by the personnel department and assigned to campuses. Other districts follow a more decentralized process in which principals have more authority over which teachers are hired for their campuses. Some districts follow a balanced approach that involves both the personnel department and the principals (Liu, 2002).

The hiring process can be very subjective on the part of the administrator making the decision. One administrator may like a particular answer to an interview question, while another administrator may not. Also, different administrators look for different qualities in teachers and use a variety of methods to evaluate those qualities. Inevitably, some administrators are more successful at selecting effective teachers than other administrators.

The recruitment, selection, and training of new teachers requires a significant financial investment on the part of school districts. Barnes, Crowe, and Schaefer (2007) found that the average cost of replacing a teacher varied from \$4,366 to \$17,872, depending on the district. These include the payroll costs of the human resources and professional development staff, traveling to job fairs, mentoring programs, new teacher training and workshops. This money is a great investment when successful teachers are hired who will stay in a district for years to come. However, the money is not well spent when ineffective teachers are selected or when teachers do not stay in the district long.

There have been efforts to standardize the teacher selection process to make it

more effective and objective and less time consuming. Standardized interview tools have been used by many school districts for decades. Some examples include the Haberman Star Teacher Evaluation, Mesa Educator Perceiver Interview (MEPI), Emphasizing More Personalized Attitudes Toward Helping Youth (EMPATHY), Gallup's Teacher Perceiver Interview (TPI), and Gallup's Urban Teacher Perceiver Interview (Urban TPI).

The Gallup Organization developed the TI tool to assist school districts in selecting teachers from their applicant pools. It is a web-based assessment instrument consisting of multiple choice and Likert items that requires approximately 30 minutes to complete. The TI tool produces a score from 0 to 100 based on an applicant's responses. The Gallup Organization says the TI provides the following benefits:

TeacherInsight helps you hire the best teachers -- fast. Based on more than 30 years of research into the talents of the very best teachers, TeacherInsight provides a quick, effective way to source and assess a large volume of applicants. This innovative selection tool allows you to:

- Identify more teacher candidates like your best teachers
- Increase the speed of assessing applicants
- Reduce staff time spent interviewing applicants – and the associated costs
- Focus valuable district staff time on recruiting candidates
- Keep your HR office open 24/7

Moving beyond knowledge and skills, TI assesses the talents that result in teacher excellence that are difficult or nearly impossible to teach. The

TeacherInsight assessment comes from qualitative and quantitative approaches

with outstanding teachers and follow-up quantitative predictive validity studies.

Consisting of multiple choice and Likert items in a Web-based assessment, TI complies with all Equal Employment Opportunity (EEO) guidelines.

The TI assessment requires approximately 30 minutes to complete and is available 24/7. Results are based on the applicant's responses and include a score that is predictive of an applicant's potential for teaching success based on his or her talent. Once an applicant has completed TI, access to the results is nearly immediate. Gallup instantly reports scores to districts through its Web-based reporting site, Gallup Online. Districts can also automate applicants accessing TI and recording the results into existing applicant tracking systems. Gallup provides a seminar for principals, human resource specialists, and other district personnel in districts using TI. This seminar helps participants understand the research behind the TI System and provides a foundation for interpreting the TI scores. The seminar also introduces participants to the Teacher FIT (Further Insight into Talent) interview, a supplemental set of questions that principals or other staff may use with applicants. (Gallup, 2008a, p.1)

The Gallup Organization (2008b) claims that the TeacherInsight “ensures the depth of teaching talent throughout the district” (p.1) Gallup also says that the TI “consistently measures the talent of every teacher you hire” (Gallup, 2008b, p.1).

Many school districts across the country use the TI as part of their teacher application and selection process. Some school districts use a cut score and do not consider applicants that fail to achieve or exceed that minimum score on the TI. Other districts do not use a cut score and instead use the TI score as one source of

information to be considered in the selection process. Regardless of which method a district utilizes, the TeacherInsight score inevitably affects applicants' chances of obtaining a position.

Some alternative certification programs also use the TI as part of their selection process. Thus, the TI not only can affect an individual's chances of being hired by some school districts, it can also determine whether someone is accepted into an alternative certification program.

Statement of the Problem

It is critical that schools and districts identify highly effective, highly qualified teachers to raise student achievement. School districts have limited resources such as time, money, and manpower to achieve this task. If standardized interview tools such as the TI are effective at identifying better teachers, the time and money spent on them are worthwhile. However, if these tools are not effective, then the time and money spent could be better utilized elsewhere. Furthermore, if the TI does not effectively identify better teachers it could be preventing good candidates from being hired or from being accepted into alternative certification programs.

Thus, the problem for educators is how to identify and hire effective teachers. Research conducted for this study examines the relationship between TI scores and teacher effectiveness. The results of this research provide school district administrators with correlations between teachers' TI interview tool scores and each of the eight domain scores of their Professional Development and Appraisal System (PDAS) evaluations. Districts and schools may use this information to review their teacher

selection methods. Improvements in the teacher selection process can increase the effectiveness of the teachers hired, and as a result, increase student achievement.

Purpose

The TI is being used by many school districts across the nation as part of their hiring process. It is also used by many alternative certification programs as part of their admissions process.

The Gallup Organization charges \$17 per applicant for the administration of the TeacherInsight. However, the district being studied paid an annual fee of \$42,000 for an unlimited number of administrations of the TI. If the TI is able to identify more effective teacher applicants, that time and money is well spent. However, if the TI does not reliably predict teacher effectiveness, then it is actually doing a disservice as it is wasting limited resources and, worse yet, potentially resulting in some less effective teachers being hired. Only one independent, published study could be found regarding the TeacherInsight interview tool. One additional published study conducted by Gallup was found. Additional research could benefit school administrators who are using, or considering the use of, the TI as part of the teacher selection process. Thus, the purpose of this study is to determine the relationship between TI scores and teacher effectiveness, as measured by the Professional Development and Appraisal System.

Research Questions

This study addressed the following research questions:

1. What are the relationships between teachers' TeacherInsight scores and each of the eight domain scores of their PDAS evaluation ratings?

2. What are the relationships between teachers' TeacherInsight scores and their age, years of teaching experience, gender, and highest degree earned?
3. What are the relationships between teachers' eight PDAS domain scores and their age, years of teaching experience, gender, and highest degree earned?
4. To what extent do the eight PDAS domain scores predict the variability in the TeacherInsight scores over and above that explained by teacher age, gender, years of experience, and highest degree earned?

Methodology

This study involved teachers hired into a North Texas school district for the 2006-2007 school year. Those teachers' TI scores were collected, along with the eight domain scores of their PDAS evaluation scores. The teachers' demographic information including age, years of teaching experience, gender, and highest degree earned were also collected. The relationships between the teachers' TI scores, eight domain PDAS evaluation scores, age, years of teaching experience, gender, and highest degree earned were studied through correlational and linear regression statistical methods.

Limitations

The data concerns only one North Texas school district. While the TI tool is used by districts and schools across the country, specific interpretations of the results may not be comparable to other states that do not use the PDAS teacher appraisal tool.

Definition of Terms

Academic Excellence Indicator System (AEIS): pulls together a wide range of

information on the performance of students in each school and district in Texas every year. This information is put into the annual AEIS reports, which are available each year in the fall. The reports provide extensive information on school and district staff, finances, programs and student demographics. Performance on each of the following indicators is shown disaggregated by ethnicity, sex, special education, low income status, limited English proficient status, and at risk status (district only):

- Results of Texas Assessment of Knowledge and Skills (TAKS*); by grade, by subject, and by all grades tested;
 - Participation in the statewide assessment programs (TAKS/TAKS(Accommodated)/TAKS-M/TAKS-Alt);
 - Exit-level TAKS Cumulative Passing Rates;
 - Progress of Prior Year TAKS Failers;
 - Results of Student Success Initiative;
 - Attendance Rates;
 - Annual Dropout Rates (grades 7-8, grades 7-12, and grades 9-12);
 - Completion Rates (4-year longitudinal);
 - College Readiness Indicators;
 - Completion of Advanced / Dual Enrollment Courses;
 - Completion of the Recommended High School Program or Distinguished Achievement Program;
 - Participation and Performance on Advanced Placement (AP) and International Baccalaureate (IB) Examinations;
 - College-Ready Graduates;
 - Texas Success Initiative (TSI) – Higher Education Readiness Component;
- and

- Participation and Performance on the College Admissions Tests (Scholastic Aptitude Test [SAT] and American College Testing Assessment [ACT]) (Texas Education Agency, 2008a, p.1).

Accountability Ratings: are assigned to every school and district annually by The Texas Education Agency (TEA). The different ratings include Exemplary, Recognized, Academically Acceptable, and Academically Unacceptable. These ratings are determined by three base indicators: TAKS and TAKS (Accommodated) scores, completion rate for grades 9 through 12, and annual dropout rate for grades 7 and 8.

Adequate Yearly Progress (AYP): a federal accountability system mandated under the NCLB Act. There are many similarities and differences between the Texas Academic Excellence Indicator System (AEIS) and federal AYP accountability systems.

Economically Disadvantaged: one of the subgroups analyzed in the AEIS and AYP systems. Students that qualify for free or reduced lunch based on their family household income are included in this category.

Gallup Organization: a company that has studied human nature and behavior for more than 70 years. This company developed the TeacherInsight tool and its predecessor, the Teacher Perceiver.

Highly Qualified (HQ): part of the NCLB Act of 2001. The law requires that all teachers of core academic subjects in the classroom be highly qualified. This is determined by three criteria: (1) attaining a bachelor's degree or better in the subject taught; (2) obtaining full state teacher certification; and (3) demonstrating knowledge in the subjects taught.

No Child Left Behind (NCLB) Act (Public Law 107-110, 2001): signed into law on January 8, 2002. The legislation is based on four major areas or pillars: accountability

for student achievement, focus on what works, flexibility, and parental involvement.

There are five goals as part of this Act:

- By 2013-2014, all students will reach high standards, at a minimum attaining proficiency or better in reading/language arts and mathematics;
- All limited English proficient students will become proficient in English and reach high academic standards, at a minimum attaining proficiency or better in reading/language arts and mathematics;
- By 2005-2006, all students will be taught by highly qualified teachers;
- All students will be educated in learning environments that are safe, drug-free, and conducive to learning;
- All students will graduate from high school.

The Act requires schools or districts that fail to make adequate yearly progress (AYP) to school improvement interventions. The level of school improvement depends on how many years the school or district has failed to make AYP.

Professional Development and Appraisal System (PDAS): the State's approved instrument for appraising its teachers and identifying areas that would benefit from staff development. Cornerstones of the process include a minimum of one 45-minute observation and completion of the Teacher Self-Report form. PDAS includes 51 criteria within eight domains reflecting the Proficiencies for Learner-centered Instruction adopted in 1997 by the State Board for Educator Certification (SBEC) (Region XIII Education Service Center, 2008).

TeacherInsight (TI): a web-based assessment developed by The Gallup Organization, consisting of multiple choice and Likert items that requires approximately 30 minutes to complete. Results are based on the applicant's responses and include a score that Gallup claims is predictive of an applicant's potential for teaching success

based on his or her talent. Once an applicant has completed TeacherInsight, access to the results is nearly immediate. Gallup instantly reports scores to districts through its web-based reporting site, Gallup Online. Districts can also automate applicants accessing TeacherInsight and recording the results into existing applicant tracking systems.

Texas Assessment of Knowledge and Skills (TAKS): measures the statewide curriculum in reading at Grades 3-9; in writing at Grades 4 and 7; in English Language Arts at Grades 10 and 11; in mathematics at Grades 3-11; in science at Grades 5, 8, 10, and 11; and social studies at Grades 8, 10, and 11. The Spanish TAKS is administered at Grades 3 through 6. Satisfactory performance on the TAKS at Grades 3, 5, and 8 is a prerequisite to promotion to the following grade level. Satisfactory performance on the TAKS at Grade 11 is prerequisite to a high school diploma.

Texas Education Agency (TEA): guides and monitor activities and programs related to public education in Texas. It is comprised of the commissioner of education and agency staff.

Significance of the Study

The significance of this study was to determine the relationship between TeacherInsight scores and teacher effectiveness, as measured by the Professional Development and Appraisal System. Hiring effective teachers who are able to raise student achievement is a critical issue for all schools and districts across the nation. The NCLB Act is annually increasing the passing rates required for all students and subgroups of students for schools and districts to meet adequate yearly progress. This

Act will require all students to demonstrate proficiency on state tests by the 2013- 2014 school year. Schools and districts that do not meet adequate yearly progress are subject to additional requirements and sanctions such as tutoring paid for by the district and the opportunity for students to transfer to another school or district.

During the 2006 special session, the Texas State Legislature increased the authority of the Commissioner of Education to take over schools that have been rated academically unacceptable for two consecutive years. The Commissioner now has the authority to replace the entire staff at such campuses. The principal would automatically be removed and a campus intervention team would be appointed by the commissioner to determine which other employees would be removed (Stutz, 2006).

These high levels of state and federal accountability are just two of the factors that demonstrate the importance of selecting effective teachers. Other factors include the shortage of applicants in many districts around the nation and the highly qualified requirements of the NCLB Act, as well as the high cost of recruiting, selecting, and training new teachers.

Organization of the Study

This study is organized into five chapters. Chapter 1 includes an introduction, background, statement of the problem, research question, research hypothesis, limitations, definition of terms, significance of the study, and the organization of the study. Chapter 2 includes a review of relevant literature from journal articles, dissertations, and books concerning teacher selection and teacher effectiveness. Chapter 3 outlines the methodology of this study and will include the purpose, context,

study participants, research questions, research design, data collection, data analysis, and a summary. Chapter 4 includes the presentation and analysis of the data. Chapter 5 provides a summary, discussion of the results, recommendations, and suggestions for further studies.

CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

One of the most important roles of a school administrator is to hire effective teachers and staff members. There is a great need for effective teachers to be assigned to every classroom and every student across the nation. Education is the key to a successful future and effective teachers are essential to a quality education. Lack of education can have a significant impact on individuals and on society. Approximately 1.2 million students do not graduate on time every year (Alliance for Excellent Education, 2008a). Only 60% of low income students graduate from high school, only a third enroll in college, and less than half of those that enroll in college will earn a bachelor's degree (Bedsworth, Colby, & Doctor, 2006; Conley, 2005). Almost half of African American, Hispanic, and Native American students drop out of school before graduating from high school (Orfield, Losen, Wald, & Swanson, 2004). To address these issues, this study reviewed literature in the following areas: student achievement and accountability, teacher effectiveness and student achievement, teacher evaluation, and teacher selection.

The American Civil Liberties Union (ACLU) filed a lawsuit against the Palm Beach County school district claiming they are not providing a "uniform, efficient, safe, secure and high quality education" (Jordan & DeNardo, 2008, p. 1) as evidenced by the low graduation rates. Chris Hansen, the senior staff attorney with the ACLU, stated that "Graduating from high school is virtually the minimum requirement for success. A large percentage of the students are being essentially written off" (Jordan & DeNardo, 2008,

p.1).

The 2005 average annual income for a high school dropout is almost \$10,000 less than a high school graduate and over \$35,000 less than those with a bachelor's degree (U.S. Census Bureau, 2008). Approximately 70% of those incarcerated in the United States (U.S.) are high school dropouts (Hodgkinson, 2000).

The lost earnings from the high school graduating class of 2007 equates to \$329 billion (Alliance for Excellent Education, 2008b). If this trend continues, the more than 12 million dropouts during the next decade will cost the nation more than \$3 trillion in lost earnings.

Even those who graduate from high school are often not well prepared for higher education. According to the National Center for Education Statistics (2004), 40% of college students take at least one remedial course. Cost estimates for this remedial education range from \$1 billion per year (ACT, 2005) to \$1.4 billion per year (Alliance for Excellent Education, 2008c). Taking remedial coursework in college delays graduation and increases the probability of dropping out (Adelman, 1999; National Center for Education Statistics, 2004). The lost earning potential from college dropouts equates to \$2.3 billion per year (Alliance for Excellent Education, 2008c).

Researchers have found that teacher effectiveness is the most significant factor of student achievement that schools can control (Darling-Hammond & Bransford, 2005; Darling-Hammond, 2000; Wilson, Floden, & Ferrini-Mundy, 2001). Thus, teacher selection is an extremely important responsibility of school administrators. The process of selecting and hiring teachers requires significant amounts of our two most limited resources – time and money. Standardized interview tools attempt to help

administrators identify the most effective teachers. Many of these interview tools are lists of questions and scoring rubrics that are conducted in traditional interview settings – one on one with an interviewer and an interviewee. However, in an attempt to save interviewer's time and be more convenient for interviewee's, some of these standardized interviews are now administered via the internet. One of these Web-based interview tools is the TeacherInsight, which was developed and is sold by the Gallup Organization.

The need for highly effective teachers continues to increase along with student achievement standards under federal and state accountability. The federal No Child Left Behind (NCLB) Act requires all public schools and districts achieve adequate yearly progress as demonstrated by student test scores. Many states also have state-level accountability systems in place as well. Texas has the Academic Excellence Indicator System (AEIS) which rates all public schools and districts based on student test scores, attendance rates, drop out rates, and completion rates. These federal and state ratings are very significant as they are publicized extensively via the media, websites, and campus report cards. They are more than just a label as they can impact property values for the attendance zone for that school or district. Low ratings can also have significant consequences for schools or districts as they will be required to develop corrective action plans and make changes to attempt to earn a higher rating the following year.

Student Achievement and Accountability

Hiring effective teachers is even more important as a result of the increasing

federal and state accountability. Districts and schools in Texas receive ratings from both the federal level through the No Child Left Behind Act (NCLB) and from the state level through the Academic Excellence Indicator System (AEIS). These ratings are based primarily on student test scores so districts and schools are striving to find ways to increase student achievement. The strong connection between teacher effectiveness and student achievement has demonstrated the need to hire effective teachers that can help students achieve the passing test scores required by NCLB and AEIS.

Testing has been a method to measure student achievement or aptitude for hundreds of years. However, district and school accountability based on those test results are a much more recent phenomenon. Although tests were given more frequently in the 1930s and 1940s, students were often not held accountable for their performance as the result of a trend that is now known as social promotion (Ravitch, 2002). The Equality of Educational Opportunity report by James Coleman (1966) may have been the catalyst for the accountability movement. This study analyzed the differences in educational resources and student achievement scores. Prior to this report, student test scores were only used for individual student accountability and educational reform focused on resources, not on results (Ravitch, 1983).

The amount of standardized assessment data has increased significantly over the past few decades. Examples include state-level testing, the National Assessment of Educational Progress, and international tests of math and science (Ravitch, 2002). This data has also become much more accessible through the internet and other technological advances and its availability have prompted more comparisons of results between schools, districts, states, nations and student subgroups.

Finn (2002) identified several versions of accountability, including compliance, professional norms and expertise, standards-based reform, the marketplace, and several combinations of two or three of those types. Accountability systems frequently include goals, standards, measurement, reporting, and consequences (Hanushek & Raymond, 2002).

Legislators and policy-makers have become more focused on student achievement and educational outcomes and less focused on resources and educational inputs (Ravitch, 2002). Twenty-seven states require students to pass standardized tests in order to receive a high school diploma (Hammond, 2008). However, these exams are contrary to research by Rob Warren and Eric Grodsky which found that state graduation exams do not increase achievement, employment, or earnings, but these tests do increase the dropout rate (Morrison, 2008).

The cost of state accountability systems in the U.S. was over \$234 million in 2001 (Hoxby, 2002). The amount of spending varied considerably by state, from \$1.79 per student in South Carolina to \$34.02 per student in Delaware. This continued increase in testing and accountability is in spite of evidence of negative effects of high-stakes testing (Nichols & Berliner, 2007; Jones, Jones, & Hargroves, 2003; Orfield & Kornhaber, 2001).

Texas Accountability Rating System

The State of Texas has an accountability process for students, schools, and districts called the *Accountability Rating System*. While it is based primarily on student performance as demonstrated on the Texas Assessment of Knowledge and Skills

(TAKS) test, it also include factors such as elementary school attendance rate, middle school dropout rate, and high school completion rate. Not only are the results of all students at a school or district evaluated, but subgroups such as white, African American, Hispanic, and economically disadvantaged students are also analyzed. This system assigns one of the following four ratings to each public school and district in the state: exemplary (highest), recognized, academically acceptable, and academically unacceptable.

The origins of this system started back in 1980 with the implementation of the Texas Assessment of Basic Skills (TABS). This test covered reading, writing, and math and was administered to students in grades three, five, and nine. There was no student, school, or district accountability based on this test; but the state did start requiring ninth grade students that did not pass it to re-take the exam every year they remained in school. However, students were not denied high school diplomas if they were not able to pass the TABS tests (Texas Education Agency, 2008b).

The TABS was replaced by the Texas Educational Assessment of Minimum Skills (TEAMS) during the 1985-1986 school year. The same subjects (reading, writing, and math) were tested but the tests were more rigorous and tested more grade levels (one, three, five, seven, nine, and eleven). Student-level accountability began with this test as the state started requiring students to pass the eleventh grade exit-level tests to be able to graduate.

The Texas Assessment of Academic Skills (TAAS) replaced the TEAMS in 1990. The TAAS still tested the same subjects (reading, writing, and math), but it was more rigorous than the TEAMS and was administered in grades three, five, seven, nine, and

eleven. Over the next thirteen years of the TAAS, there were some changes made, including adjusting the grade levels tested and the addition of two subjects, science and social studies. The test was moved from the fall to the spring and the exit level test was moved from eleventh grade to tenth grade. Students still needed to pass the exit level tests in order to receive a diploma.

School and district level accountability was first implemented in 1991 with the start of the Academic Excellence Indicator System and the Accountability Rating System. The legislation mandating this system was Texas House Bill 72 in 1984 which mandated that an accountability system be established that focused on student achievement instead of merely following state rules and regulations. Over the past seventeen years, the system has been updated many times with new state tests, new accountability standards, and new ratings.

The TAKS replaced the TAAS in 2003. It is the most rigorous Texas assessment yet and tests students in reading, math, science, social studies, English/language arts, and writing in grades three through eleven. Math and reading or English/language arts are tested every year and the other subjects are only administered at certain grade levels. For example, the science TAKS tests are administered in grades five, eight, ten, and eleven.

Students need to pass all four eleventh grade exit-level TAKS tests (English/language arts, mathematics, science, and social studies) in order to graduate from high school. However, the exit level TAKS tests are no longer the only Texas exams that are considered high stakes. Beginning in 2003, third grade students were required to pass the reading test to be promoted to fourth grade the following year. That

same cohort of students were required to pass the fifth grade reading and math tests in 2005 to be promoted to sixth grade. The same class was required to pass the eighth grade reading and math tests in 2008 to be promoted to ninth grade. However, students are provided three opportunities to pass the tests required for promotion and even those students that don't pass could still be promoted if the teacher, parent, and principal unanimously agree to do so. A survey of teachers found that only four percent of teachers support the requirement that students pass TAKS in grades three, five, and eight to be promoted to the next grade level (Johnson, Rice, Sullivan, Henderson, & Beard, 2008).

Students that are Limited English Proficient (LEP) can take the Linguistically Accommodated Testing (LAT) version of the TAKS tests or they can be exempted from the TAKS tests and take the Texas English Language Proficiency System (TELPAS) tests instead. However, LEP students cannot be exempted from the exit level tests. The Language Proficiency Assessment Committee (LPAC) decides which assessments are appropriate for each LEP student.

Students that receive special education services can be exempted from the TAKS tests by an Admission, Review, and Dismissal (ARD) committee and take one of three different versions: TAKS (Accommodated), TAKS (Modified), or TAKS (Alternate). Students receiving special education services are the only students that can be exempted from the exit level TAKS tests.

The TAKS tests are scored using a scaled score. The number of correct raw items is converted to the scale score for each TAKS test. A passing scale score is 2,100 and a *commended* scale score is 2,400. The only subjects tested through TAKS that

have consecutive administrations each year are mathematics and English/reading. The state of Texas analyzes the average scale score increase or decrease for each campus and district in mathematics and English/reading and awards *Gold Performance Acknowledgments* to those that fall within the top quartile of campuses or districts for scale score increase.

The expectation level for students has increased over the years these tests have been administered. As previously explained, as the state tests changed from TABS to TEAMS to TAAS to TAKS, the level of rigor on the test increased with each change. Furthermore, when the TAKS test was implemented, the standard required for each student to pass was increased annually during the first few years of the TAKS. The first year of the TAKS the cutoff score for passing was two Standard Errors of Measurement (SEM) lower than the state panel's recommendation. The second year that standard was increased to only one SEM below the panel's recommendation. The third year the standard was increased again to the panel's recommendation. The only exception to this is the EXIT level TAKS tests which remained at two SEMs below for the first two years, one SEM below for the third year, and at the panel's recommendation starting in the fourth year.

In addition to the level of rigor increasing for the students through the state testing changes, the accountability system has also raised the bar for schools and districts. The first year the TAKS was implemented, there was no accountability ratings issued to schools and districts. Over the past four years that schools and districts did receive ratings, the TAKS passing rate standard to receive a recognized rating has increased from 70% to 75%. The TAKS passing rate standard to receive an

academically acceptable rating has increased for science from 25% to 45%, mathematics from 35% to 50%, social studies and writing from 50% to 65%, and reading/English language arts from 50% to 70%.

The TAKS is already scheduled to be replaced at the high school level in 2012 by 12 end-of-course exams: English I, English II, English III, algebra I, geometry, algebra II, biology, chemistry, physics, world geography, world history, and U.S. History. Students will be required to earn an average score of 70 or higher in each of the subject areas (English, math, science, and social studies) in order to graduate. Thus, a student with a low score on some of the end-of-course exams could still graduate as long as the student does well enough on the other exams to average a 70 for each subject.

A frequent complaint about the Texas Accountability Rating System is that the ratings are based on the lowest subgroup score on a test. A school or district could have very high scores for almost all tests and subgroups, but if there is one low score that is the one that determines the rating. Another school or district could have more students fail the test but still have a higher rating depending on how those failures are distributed among tests and subgroups. For example, a high school of 1,000 students could receive an exemplary rating if 90% of the students in each group and for each test pass. That means up to 100 students could fail up to four tests each and the school could still be exemplary. However, that same school of 1,000 students could have just six students out of a subgroup of 50 students fail just one test and that would lower that subgroup's passing rate into the next lower (recognized) rating.

Another complaint about the system is that it does not consider other factors out of the school's control such as student ethnic demographics and socioeconomic ratings

when assigning the ratings. In fact, having a more diverse student body can often make it more difficult to receive a high rating for multiple reasons. Economically disadvantaged students, Hispanic students, and African American students, on average, are less successful on the TAKS tests. Furthermore, greater diversity will cause a school or district to be accountable for more scores because subgroups (African American, economically disadvantaged, Hispanic, and white) will meet the minimum group size. Finally, each student can count up to three times in the accountability system – the all students group, an ethnicity student group (African American, Hispanic, or white), and the economically disadvantaged group. A study done by the Dallas Morning News found that districts that received the highest rating (exemplary) tended to be small and rural or wealthy and suburban (Unmuth, 2007). While the ratings system uses 36 measures, most of the exemplary districts are only judged on 10 or fewer because their enrollment lacks the size and/or diversity to have more of their student groups count for rating purposes.

Interventions are required for schools or districts that receive an academically unacceptable rating or are deemed at-risk of a future academically unacceptable rating. These interventions begin with the assignment of a Technical Assistance Team (TAT) from the state and become more extensive with additional years of low ratings. The most severe intervention would be mandatory closure, which could be required after five consecutive years of academically unacceptable ratings. Conversely, schools and districts that receive an exemplary rating are exempted from some statutes and rules.

The Texas Legislature is considering changes to the state accountability system to a value-added approach that would consider student achievement growth from the

previous year and would consider the student demographics of the school and district. Educators and communicators advocated for a value-added system as they testified to the Select Committee on Public School Accountability during one of a series of hearings (Radcliffe, 2008). Barton (2007) advocates for a dual-approach of accountability that measures both student performance on end-of-year tests and the gain based on how far the students have come during the school year.

The Dallas Independent School District has been using a system like this for the past 16 years. They calculate a School Effectiveness Index for their campuses based on how far the students advance academically. They scale the score such that a score of 50 is average and any score above or below that represents above or below average student learning (Fischer, 2008).

Federal Adequate Yearly Progress

While Texas and most other states already had an accountability system in place by the late 1990s, some states did not. The NCLB Act was signed into law in 2002 and requires all states to administer student performance assessments and then assigns each district and school a rating based on their students' results. Similar to the Texas Accountability Rating system, the NCLB rating is based primarily on the state test passing rates, attendance rates, and graduation rates for all students and certain student subgroups.

While there are some similarities between the two systems, there are more differences. While the Texas Accountability Rating System has four levels for ratings, the NCLB only has two – met adequate yearly progress (AYP) and missed AYP. The

minimum sizes for student groups to count are different so a group may count under the Texas system, but not under the NCLB. While some of the subgroups are similar for both systems – white, African American, Hispanic, and economically disadvantaged – the NCLB also considers the performance of two additional subgroups of students. These two additional subgroups under NCLB are students receiving special education services and students that are limited English proficient. Neither accountability system includes Asian students as a subgroup. Representatives of the Asian American Legal Defense and Education Fund believe the NCLB law should require schools and districts to disaggregate the test scores of Asian students (Redondo, Aung, Fung, & Yu, 2008).

While the Texas Accountability Rating System counts the scores for all grade levels of TAKS, the NCLB does not count ninth and eleventh grade TAKS scores since the system is based on the system from 2001 and students in those grade levels were not tested at that time. The Texas system counts all subjects of the state test, however, the NCLB only counts the math and reading/English language arts tests.

The two systems vary regarding how they count the state exams for special education students and limited English proficient students that are exempted from the regular TAKS test. The state system does not limit the number of students receiving special education that are exempted from the TAKS and take a different exam. However, the NCLB only allows up to 3% of students to take a special education exam and any students beyond the 3% that take a different exam are automatically counted as failures under that system.

The NCLB system requires a minimum participation rate on the state tests. Thus, regardless of the students' performance on the tests, a school or district could be rated

as missing AYP because the percentage of students taking the exam was not high enough due to student absences.

As mentioned previously, the passing rate standards have increased for the academically acceptable and recognized ratings under the Texas Accountability Rating System. The passing rate standards are also increasing under the NCLB, but at different rates. The NCLB standards were 47% for reading/English language arts and 33% for mathematics in 2003, the first year of NCLB ratings. The NCLB requires incremental increases of these standards until 2014, when it is expected that 100% of students will pass the tests. Thus, the standards as of the 2008-2009 school year are 67% for reading/English language arts and 58% for mathematics.

The NCLB imposes sanctions on schools and districts missing AYP two or more consecutive years. The corrective actions that are required get more significant each additional year the school or district misses AYP and they can range from allowing students to transfer to other schools and providing additional tutoring to mandatory restructuring. There are 411 schools districts (Associated Press, 2008) and 4,509 schools involved in some level of NCLB intervention (United States Government Accountability Office, 2007).

Mandatory restructuring is the most severe sanction and is required after not meeting AYP for six consecutive years. This involves replacing teachers, a state takeover, or other major change in school governance. Approximately 1,299 schools were involved in restructuring during the 2006-2007 school year (Tomsho, 2008). That number is projected to increase to approximately 5,000 schools by 2009-2010 (Calkins, Guenther, Belfiore, & Lash, 2007).

However, approximately 40% of these schools did not actually implement the required restructuring (United States Government Accountability Office, 2007). The schools that are required to implement corrective action and restructuring have a much higher percentage of ethnic minority students (96%) and economically disadvantaged students (83%) when compared to the other schools with enrollments that are made up of 37% minority students and 54% economically disadvantaged students (United States Government Accountability Office, 2007).

Over six million students are eligible for school choice because their school is required to offer that as part of corrective action through NCLB. However, only one percent of these students transferred to another school under this program during the 2003-04 school year (Gill, McCombs, Naftel, Ross, Song, Harmon, & Vernez, 2008). Almost two million students are eligible for additional tutoring through a NCLB corrective action called supplemental educational services (SES), however, only 17% of those students participated (Gill, et al., 2008).

Many states oppose the accountability enforced by the NCLB. Proposed legislation passed through committees of the Virginia state legislature that would get the state out of the NCLB law (Mellott, 2008). Pulling out of NCLB would have significant financial implications for the state, as they would lose an estimated \$363.7 million in federal funding for 2008 alone.

There are inconsistencies between the NCLB accountability system and state accountability systems in how graduation rates and dropout rates are calculated. The NCLB allows each state to choose from dozens of different methods to calculate graduation rates. New Mexico calculates its graduation rate based on the percentage of

12th grade students who receive a diploma. This significantly over-inflates the actual graduation rate since the majority of students that drop out of school do so before ever reaching the 12th grade. Some states even use more than one method to calculate graduation rates, depending on to who they are reporting the information. For example, Mississippi education officials report an 87% graduation rate under the NCLB. However, other education officials from the same state report a 63% graduation rate, and the State Schools Superintendent claims that rate is more accurate (Dillon, 2008).

Pressure from states has prompted changes in the NCLB Act. The law is up for re-authorization in 2009, and many feel that it may not be re-authorized. Education Secretary Margaret Spellings is asking states to submit proposals for differentiated consequences for those schools and districts that fail to make adequate yearly progress, depending on how much they missed the goals. She will allow 10 states to implement these individual plans (Streich, 2008). Furthermore, the education department is considering additional changes to the NCLB Act:

- allowing the states to focus funding on the schools most in need;
- letting schools develop separate standards for students with disabilities;
- tracking how effectively English is being taught to foreign-language students;
- and letting states add additional charter schools to their current total (Streich, 2008, p.3).

Having both state and federal systems of accountability that are different provides additional pressure for schools and districts. It could be compared to serving two masters and having twice the chances to receive a low rating. Thus, school administrators are under increased pressure to identify methods and strategies to

increase student performance.

Teacher Effectiveness and Student Achievement

Many researchers have found a strong correlation between teacher effectiveness and student achievement. Miller, Murnane, and Willett (2007) found that students' mathematics achievement was reduced by 3.3 percent of a standard deviation with each 10 days of teacher absence. The quality of a teacher made the difference of a full year of learning growth in a study by Hanushek (1992). Several other studies (Rockoff, 2004; Rivkin, Hanushek, & Kain, 2005; Aaronson, Barrow, & Sander, 2003; Kane, Rockoff, & Staiger, 2005; Nye, Konstantopoulos, & Hedges, 2004) found that teacher effectiveness had a significant impact on student achievement.

Students in Los Angeles that were taught by teachers in the top quartile of effectiveness increased their achievement by five percentile points per year (Gordon, Kane, & Staiger, 2006). However, students taught by teachers in the bottom quartile dropped five percentile points per year. The achievement gap between white students and African American students could be eliminated if African American students were assigned to highly effective teachers four years in a row (Gordon, Kane, & Staiger, 2006).

Researchers in Illinois found and developed the Teacher Quality Index (TQI) based on the teachers' average American College Testing Assessment (ACT) composite score, teachers' average ACT English score, teachers' average undergraduate college competitiveness ranking, percent of teachers failing the Basic Skills Test on their first attempt, percent of teachers with emergency or provisional

certification, and the percent of teachers with three or fewer years of experience (Presley, White, & Gong, 2005). They found that TQI made a difference in school achievement, particularly at the high school level and at schools with a high percentage of economically disadvantaged and minority students. They also found negative correlations between TQI and the percentage of economically disadvantaged and minority students at a school. This means that, on average, schools with high numbers of these students also have less qualified teachers to work with them.

Clotfelter, Ladd, and Vigdor (2007) found that teacher effectiveness has a greater impact on student achievement than the combined influence of race and parent educational level. A study in Tennessee also determined that the impact of teacher effectiveness on student achievement was cumulative. In this study, an average student with highly effective teachers three consecutive years outscored an average student with highly ineffective teachers three consecutive years by 50 percentile points (Sanders & Rivers, 1996). Another study found that economically disadvantaged students with effective teachers five years in a row, on average, would make large enough gains to close the achievement gap with students from wealthier families (Hanushek, Rivkin, & Kain, 2005).

Several studies (Babu & Mendro, 2003; Darling-Hammond, 1995; Kain & Singleton, 1996; Langford, Loeb, & Wyckoff, 2002; Presley, White, & Gong, 2005; Shields, Esch, Humphrey, Young, Gaston, & Hunt, 1999) found that economically disadvantaged and minority students were, on average, taught by less qualified teachers. The United States has the fourth-largest gap in teacher quality for economically disadvantaged students out of 46 countries (Scholastic Administr@tor,

2008).

A study of New York Public Schools found that teacher turnover was higher in low-performing schools (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2007). Yet, the researchers determined teacher attrition was not always a bad thing as they also found that less effective teachers – based on a value-added model of student growth on state test scores – were more likely to leave. However, another study (Wright, 2006) found that schools with higher teacher turnover rates had produced students with lower passing rates on TAKS and lower graduation rates.

Teacher Evaluation

Bossidy and Charon (2002) state “a good, candid assessment talks about the things a candidate does well and the things he or she must do better” (p. 134). There are a wide variety of methods used to evaluate teachers. Stronge and Tucker (2003) described several different models of teacher evaluation:

- **Teacher-Trait Model** – This model is characterized by a checklist of desirable attributes for teachers, such as *enthusiastic*, *fair*, and *creative*. According to a 1988 publication by the Educational Testing Service, 32% of schools used this approach despite its emphasis on the pre-existing personality traits that may not be amenable to improvement efforts.
- **Process-Oriented Model** – This model is most familiar to educators because it focuses on the instructional “processes” taking place in the classroom that can be easily observed by supervisors/administrators. Frequently, observational data are organized by specific teaching methods that research has shown to be correlated positively with student achievement, such as the Hunter model.
- **Duties-Based Evaluation** – A duties-based approach to evaluation is based on specific tasks or requirements of the job. For example, one duty might be the frequent assessment of student learning; in this model, however, the evaluation criteria would not specify the precise strategies for student assessment.

- **Accountability** – An accountability approach to evaluation typically links judgment about teacher performance to student achievement of instructional objectives or other outcome measures. In 1988, the Educational Research Service found that this approach was used in 35% of the schools nationally, and we can only assume that this percentage has increased as a result of the accountability movement across the country.
- **Goals-Based Evaluation** – This evaluation approach is similar to the business model of Managing By Objectives (MBO) and is typically used by school systems in combination with other models. It is often viewed as especially appropriate for more experienced teachers who set their own goals for professional development and are then evaluated based on their goal attainment.
- **Professional Growth Model** – A professional growth model of evaluation shifts the focus to individual teachers and their development as professionals. Observers provide ongoing, formative feedback for improving teaching skills that are identified by the teacher as areas of interest or need.
- **Hybrid** – As previously noted, hybrid evaluation models are the most common because school systems typically do not use a pure form of any of the six models listed above, but rather a unique combination that integrates multiple purposes and methodologies. Often, more prescriptive models are used for less experienced teachers and more open-ended models for more experienced and accomplished teachers (p. 17).

Studies have shown that the evaluations of teachers conducted by administrators are effective at identifying teachers that are able to increase student achievement. Jacob and Lefgren (2005) found that principal assessments of teachers are significantly better at predicting future student achievement significantly than teacher education, experience, or compensation. Podgursky and Springer (2006) analyzed multiple studies regarding the relationship between teacher effects and evaluation by supervisors and concluded “that principal evaluations provide an important independent source of information on teacher productivity” (p. 25). Thus, while teacher evaluations can be subjective, research has demonstrated the relationship between these evaluations and gains in student test scores. Finally, the Alliance for Excellent Education believes that

effective, value-added teacher evaluation can be used to improve instruction, guide professional learning, determine tenure, implement merit pay programs, and ultimately increase student achievement (Alliance for Excellent Education, 2008d).

Professional Development and Appraisal System (PDAS)

The Professional Development and Appraisal System (PDAS) is the official evaluation tool for teachers in the State of Texas. It was first piloted during the 1996-1997 school year and implemented statewide during the 1997-1998 school year. The PDAS replaced the Texas Teacher Appraisal System (TTAS).

The goal of the PDAS is “to improve student performance through the professional development of teachers” (Texas Education Agency, 2005, p. 6). The PDAS includes the following eight domains:

Domain I: Active Successful Student Participation in the Learning Process

Domain II: Learner-Centered Instruction

Domain III: Evaluation and Feedback on Student Progress

Domain IV: Management of Student Discipline, Instructional Strategies, Time, and Materials

Domain V: Professional Communication

Domain VI: Professional Development

Domain VII: Compliance with Policies, Operating Procedures and Requirements

Domain VIII: Improvement of the Academic Performance of all Students on the Campus (Texas Education Agency, 2005, p. 17).

The main information sources for scoring the PDAS are classroom observation and the teacher self-report. Classroom observation could be a one-time observation or it

could be multiple observations, but it must be a minimum of 45 minutes total. The teacher self-report includes three parts. The first part of the teacher self-report includes a checklist of objectives in reading, writing, math, science, and social studies and the teacher is to check which objectives are covered in the class(es) taught by the teacher. It also asks “What processes do you use to assess the needs of your students with regard to academic skills (TEKS/TAKS objectives)?” (Texas Education Agency, 2005, pp. 56-59). Part one is to be turned in within three weeks of the teacher orientation.

Teacher self-report parts two and three are due at least two weeks before the annual summative conference and include the following questions:

1. Describe a specific instructional adjustment (e.g., materials, sequencing, etc.) which you have made based on the needs assessment of your students.
2. Describe the approaches you have used to monitor classroom performance and to provide feedback to students regarding their progress in academic skills (TEKS/TAKS objectives).
3. Describe how you assisted your students who were experiencing serious attendance problems.
4. Describe your approach in working with students who were failing or in danger of failing.
5. List or describe, in the space provided below, your professional development activities for the past year related to campus/district goals, assigned subject/content, needs of students, or prior appraisal performance in the following areas: in-service, team planning, mentoring, collaboration with colleagues, self-study, video coursework or distance-learning, university-level coursework, professional conferences, and other non-traditional activities.
6. As a result of your professional development activities described above, what have you been able to use in your classroom that has positively impacted the learning of students?
7. Be prepared to discuss three target areas for continued professional growth. In order to organize your thoughts, you may wish to make notes below, but it is not required (Texas Education Agency, 2005, pp. 60-61).

Each PDAS domain contains between three and nine criteria, for a total 51 criteria contained within the eight domains. Each domain is scored independently of the other seven domains. Each criterion can be scored in one of four categories: exceeds expectations, proficient, below expectations, or unsatisfactory. The point values for each of these ratings are five, three, one, and zero, respectively. The points for all criteria within each domain are totaled and equate to a rating of Exceeds Expectations, Proficient, Below Expectations, or Unsatisfactory for each domain.

Domain VIII factors in the campus performance rating of exemplary, recognized, academically acceptable, or academically unacceptable and provides additional points of four, two, one, or zero, respectively. That domain also considers whether or not the campus meets AYP under the NCLB law and provides one additional point for meeting AYP. These additional points for state and federal campus ratings do not apply for teachers new to the campus that year since the ratings would be based on student performance from the previous school year.

The complete list of PDAS domains and criteria are as follows:

- Domain I: Active Successful Student Participation in the Learning Process
 - I-1. Students are actively engaged in learning.
 - I-2. Students are successful in learning.
 - I-3. Student behaviors indicate learning is at a high cognitive level (e.g., critical thinking, creative thinking, problem solving, etc.).
 - I-4. Students are self-directed/self-initiated as appropriate to the lesson objectives.
 - I-5. Students connecting learning to work and life applications, both within the discipline, and with other disciplines.

- Domain II: Learner-Centered Instruction
 - II-1. Objectives and goals include basic knowledge/skills and central themes/concepts of the discipline.
 - II-2. Instructional content is learner-centered (e.g., relates to the interests and varied characteristics of students).
 - II-3. Instructional strategies promote critical thinking and problem

solving.

- II-4. Instructional strategies include motivational techniques to successfully and actively engage students in the learning process.
- II-5. Instructional strategies are aligned with the objectives, activities, student characteristics, prior learning, and work and life applications, both within the discipline and with other disciplines.
- II-6. The teacher varies activities appropriately and maintains appropriate pacing and sequencing of instruction.
- II-7. The teacher emphasizes the value and importance of the activity/content.
- II-8. The teacher uses appropriate questioning and inquiry techniques to challenge students.
- II-9. The teacher makes appropriate and effective use of available technology as a part of the instructional process.

Domain III: Evaluation and Feedback on Student Progress

- III-1. Academic progress of students is monitored and assessed.
- III-2. Assessment and feedback are aligned with goals and objectives and instructional strategies.
- III-3. Assessment strategies are appropriate to the varied characteristics of students.
- III-4. Student learning is reinforced.
- III-5. Students receive specific, constructive feedback.
- III-6. The teacher provides opportunities for relearning and re-evaluation of material.

Domain IV: Management of Student Discipline, Instructional Strategies, Time, and Materials

- IV-1. The teacher effectively implements the discipline management procedures approved by the campus. Any lack of effective implementation is rare, inadvertent, and does not seriously compromise the needs of students or the effective operation of the classroom or campus.
- IV-2. The teacher establishes a classroom environment that promotes and encourages self-discipline and self-directed learning.
- IV-3. The teacher interacts with students in an equitable manner, including fair application of rules.
- IV-4. The teacher specifies expectations for desired behavior.
- IV-5. The teacher intervenes and re-directs off-task, inappropriate, or disruptive behavior.
- IV-6. The teacher reinforces desired behavior when appropriate.
- IV-7. The instructional materials selected by the teacher are equitable and acknowledge the varied characteristics of all students.
- IV-8. The teacher effectively and efficiently manages time and materials.

Domain V: Professional Communication

- V-1. The teacher uses appropriate and accurate written communication with students.
- V-2. The teacher uses appropriate and accurate verbal and non-verbal communication with students.
- V-3. The teacher encourages and supports students who are reluctant and having difficulty.
- V-4. The teacher uses appropriate and accurate written communication with parents, staff, community members, and other professionals.
- V-5. The teacher uses appropriate and accurate verbal and non-verbal communication with parents, staff, community members, and other professionals.
- V-6. The teacher's interactions are supportive, courteous, and respectful with students, parents, staff, community members, and other professionals.

Domain VI: Professional Development

- VI-1. The teacher successfully engages in professional development activities that positively correlate with the goals of the campus and district.
- VI-2. The teacher successfully correlates professional development activities with assigned subject content and the varied needs of students.
- VI-3. The teacher successfully engages in professional development activities that positively correlate with the prior performance appraisal.
- VI-4. The teacher works collaboratively and constructively with colleagues and other professionals toward the overall improvement of student performance.

Domain VII: Compliance with Policies, Operating Procedures and Requirements

- VII-1. The teacher complies with all of the policies, operating procedures, and legal requirements (national, state, district, and campus). Any lack of compliance is rare, inadvertent, and does not seriously compromise the needs of students or the effective operations of the campus/district.
- VII-2. The teacher generally complies with all of the verbal and written directives. Any lack of compliance is rare, inadvertent, and does not seriously compromise the needs of students or the effective operations of the campus/district.
- VII-3. Apart from classroom responsibilities, the teacher generally contributes to making the whole school safe and orderly and a stimulating learning environment for all children.

Domain VIII: Improvement of the Academic Performance of all Students on the Campus

- VIII-1. The teacher aligns instruction to include appropriate TEKS/TAKS

- objectives to support student achievement in all assigned classes.
- VIII-2. The teacher analyzes TAKS performance data relevant to all students in assigned classes prior to beginning instruction.
 - VIII-3. The teacher adjusts the sequencing of classroom instruction to appropriately incorporate TEKS/TAKS objectives.
 - VIII-4. The teacher selects/adapts instructional materials and activities that are correlated with appropriate TEKS/TAKS objectives.
 - VIII-5. The teacher provides feedback to all students regarding their learning progress on appropriate TEKS/TAKS objectives.
 - VIII-6. The teacher monitors attendance of all students in assigned classes and contacts parents, counselors, and other school officials for students with serious attendance problems.
 - VIII-7. The teacher identifies and assesses the needs of assigned students who are in at-risk situations.
 - VIII-8. The teacher meets with students who are failing or in danger of failing and develops an appropriate plan for intervention.
 - VIII-9. The teacher modifies and adapts classroom materials and/or instruction for students in at-risk situations (Texas Education Agency, 2005, pp. 63-119).

Davis-Frost (2000) conducted a study regarding PDAS and found the following:

The analysis of data supports the precept that Texas' Professional Development and Appraisal System significantly impacts the classroom practices in the areas of learner centered instruction, classroom management, support for all students, professional growth, communication, TAAS improvement, and application of learning. These identified areas support and reflect the intentions of the instrument's designers. The teachers' years of experience and area of instruction had no significant effects on teachers' responses. (p. 58)

Teacher Selection

There are currently about 3.2 million teachers in public schools in the U.S., however, data of student enrollment growth and teacher turnover project that we will need to hire an additional 2.8 million teachers in the next eight years (Wallis, 2008). A recent survey of teachers in Texas found that 44% of teachers are thinking about quitting their jobs (Johnson, et al., 2008). Every year, approximately 14% to 15% of teachers leave their school (Ingersoll, 2003). About half of them leave teaching all

together while the other half move to another school or district. The Alliance for Education (2004) determined that teacher turnover costs public schools \$2.6 billion per year.

There are plenty of college students graduating with education degrees to fill these positions. However, more than half of them choose a different occupation or enter graduate school (Roukema, 2004).

Furthermore, studies have found that many schools are not hiring the best candidates. Levin and Quinn (2003) studied four large urban districts and found that they each had plenty of applicants – each had between six to 20 times more applicants than positions to fill. However, between 31% and 58% withdrew their applications before they were hired. The majority of them did so because of late hiring timelines – they were already hired by other, mostly suburban, districts. Those that withdrew had significantly higher college grade point averages and were much more likely to have a degree in their teaching field than those that were hired.

Multiple studies (Henke, Chen, Geis, & Knepper, 2000; Ballou, 1996; Hanushek & Pace, 1995) have determined that, on average, college graduates that went into teaching had lower measured ability than other college graduates. This could be a result of which students chose to major in education, the hiring practices of school districts, or both of these factors. Fortunately, more recent data compiled by Gitomer (2007) has shown that new teachers' average Scholastic Aptitude Test (SAT) verbal and math scores and grade point averages have increased from eight years ago.

Years of experience and education are two of the factors that are often considered by administrators when deciding which teachers to hire. However,

researchers have found that those factors are not very effective predictors of teachers' performance and their ability to raise student achievement. Schmidt (2004) found that the difference in job performance attributed to experience is only significant during the first five years. Thus, as far as experience impacts performance, a five-year veteran and a 25-year veteran are equal. Another study found that only three percent of the impact on student learning could be attributed to teacher experience and education (Goldhaber, 2002).

Districts and schools use a wide variety of methods and procedures to select teachers. The application and hiring process is becoming more technology-driven in most districts through district websites, on-line applications, and internet job postings (Keller, 2003). Common steps of the teacher selection process include applications, resume and credential review, criminal background check, interviews, and reference checks. Another technique that is less-commonly used is having the candidate demonstrate a sample lesson (Longo, 2003).

Liu and Johnson (2006) studied the hiring practices in California, Florida, Massachusetts, and Michigan through a survey of approximately 500 first and second year teachers. Approximately three-fourths of these teachers were hired through a decentralized process. However, most of them still only interviewed with the principal and not with department chairs, teachers, parents, or students. Thus, the schools are not reaping the potential benefits of a school-based approach by including more stakeholders in the process. Furthermore, over one-third of the new teachers hired in California and Florida were hired after the school year had started.

Adamowski, Therriault, and Cavanna (2007) surveyed principals and 100%

perceived hiring as very important to effectiveness as a school leader. However, when asked how much autonomy they currently have regarding hiring, the following was the distribution of responses: great deal of autonomy 27%, some autonomy 30%, not so much autonomy 33%, no autonomy 10%.

Thus, the process that many districts and schools use to select teachers is inefficient and also ineffective at identifying the best teachers. The TeacherInsight is a very fast screening process and, if it is effective, it could help identify better teachers and also prevent districts from losing applicants because of hiring delays.

Standardized Interviews

In an attempt to provide more validity and reliability to the interview process, several companies or individuals have developed standardized teacher interviews. While some of these standardized interviews are conducted in person, others are completed through a Web-based instrument.

Interactive Computer Information System-Urban (ICIS-Urban)

The Interactive Computer Information System-Urban (ICIS-Urban) is a revision of the original ICIS that was specifically designed for use by urban school districts. Springston (2006) found positive correlations between the ICIS-Urban and teacher effectiveness ratings by the administration. Statistically significant correlations were found between the administrator ratings and three of the five domains of the instrument. Significant correlations were found between the administrator ratings and these three domains: Aptitude/Suitability for Urban Teachers (UT), Knowledge of Teaching (KT) and

Knowledge of Students (KS). However, the sample size of this study was only 30 and there were weak correlations for Knowledge of Content (KT) and Working with Others (WO).

Principal Perceiver Interview (PPI)

The Principal Perceiver Interview (PPI) is similar to the Teacher Perceiver Interview except that it is intended to be used with administrator applicants instead of teachers. Redburn (1998) found no significant difference in principal effectiveness of those administrators selected using the PPI as part of the selection process and those that did not take the PPI as part of the selection process. Principal effectiveness was measured by their supervisors using the Principal Rating Scale (PRS). The research concluded that there is no compelling evidence to support the use of the PPI and the time and money required to do so.

STAR Teacher Selection Interview

Martin Haberman, professor of Education at the University of Wisconsin and founder of the Haberman Foundation, developed an instrument called the STAR Teacher Selection Interview in 1995. The tool is a 30 minute structured personal interview that evaluates the following ten aspects of teaching:

1. Persistence predicts the propensity to work with children who present learning and behavioral problems on a daily basis without giving up on them for the full 180 day work year.
2. Organization and Planning refers to how and why star teachers plan as well as their ability to manage complex classroom organizations.
3. Values student learning predicts the degree to which the responses reflect a

- willingness to make student learning the teacher's highest priority.
4. Theory to Practice predicts the respondent's ability to see the practical implications of generalizations as well as the concepts reflected by specific practices.
 5. At-Risk Students predicts the likelihood that the respondent will be able to connect with and teach students of all backgrounds and levels.
 6. Approach to Students predicts the way the respondent will attempt to relate to students and the likelihood this approach will be effective.
 7. Survive in Bureaucracy predicts the likelihood that the respondent will be able to function as a teacher in large, depersonalized organization.
 8. Explains Teacher Success deals with the criteria the respondent uses to determine teaching success and whether these are relevant to teachers in poverty schools.
 9. Explains Student Success deals with the criteria the respondent uses to determine students' success and whether these are relevant to students in poverty schools.
 10. Fallibility refers to how the teacher plans to deal with mistakes in the classroom (Haberman Foundation, 2008a).

The Haberman Foundation claims that the STAR Teacher Selection Interview identifies teachers “who will succeed in increasing student achievement, who will stay longer at your school, and do a better job” and claims a 97% accuracy rate (Haberman Foundation, 2008b, p. 1). Another document by the Haberman Foundation claims a “95% accuracy rate in predicting which teachers will stay and succeed and which ones will fail or quit.” (Haberman Foundation, 2008c, p. 1). However, a study by Sawyer (2005) found no significant relationship between the Star Teacher Selection Interview and first-year teacher success based on the principal’s evaluation, even after controlling for age and relevant work experience.

Teacher Perceiver Instrument (TPI)

The Teacher Perceiver Instrument (TPI) was created by The Gallup Organization when Gallup was a subsidiary of Selection Research, Incorporated. The TPI is often referred to as the *SRI Interview* as a result of its creation by the Selection Research, Incorporated. The first version of the TPI was 60 open-ended questions read by a trained evaluator and scored as correct or incorrect, producing a score ranging from zero to 60. An abbreviated version of the TPI was developed that was very similar to the original TPI except that it only contained 22 open-ended questions.

The questions in both versions of the TPI were developed around twelve different themes:

1. **Mission:** The belief that students achieve success through the contributions of the teacher and other significant people in the students' lives.
2. **Empathy:** The belief the students' feelings and attitudes are essential to understand as a teacher and to be able to understand the students' point of view.
3. **Rapport Drive:** The belief a positive relationship with a student is essential to maximizing the student's potential.
4. **Individualized Perception:** The belief that personalizing a student's learning experience comes from understanding each student's perception of themselves as learners.
5. **Listening:** The belief that in maximizing communication the teacher must be a responsive listener.
6. **Investment:** The belief that intrinsic value as a teacher is derived from student growth and the teacher's efforts to achieve this growth.
7. **Input Drive:** The belief the teacher must be resourceful in searching for ideas and strategies to maximize student learning.
8. **Activation:** The belief the teacher is the catalyst for student learning and their actions serve as an instigator in stimulating student learning.

9. Innovation: The belief the teacher must be creative and adaptive in finding effective ways to positively promote student learning.
10. Gestalt: The belief of task completion and the drive to succeed in the goal of student learning.
11. Objectivity: The belief the teacher must look at the whole before they can dissect and address the pieces of student learning.
12. Focus: The belief the teacher has a plan for their future actions and they center on the goal of student learning.

Antoline (2000) found that the TPI can discriminate between teachers that will receive high scores on principal evaluations and those that will receive poor scores on principal evaluations. This study also used ratings completed by the students and found that there were significant differences between the principal and student ratings of the teachers. Three of the 12 themes (Mission, Input Drive, and Individualized Perception) could explain a significant amount of the variability of principal ratings. The Empathy theme explained a significant amount of the variability of the student ratings. A significant amount of the variability in a job satisfaction survey completed by the teachers could be explained by the Individualized Perception theme. The teachers were divided into four categories: highly recommended, recommended, conditionally recommended, and not recommended. The research found a relationship between the rankings the teachers received from the principals and these four classifications.

A meta-analysis of 24 studies found a weak to moderate relationship ($r = 0.28$) between Teacher Perceiver scores and the teacher effectiveness based on student ratings, administrator ratings, classroom observations, student gain scores, and teacher attendance (Metzger & Wu, 2006). The correlations were stronger between the Teacher Perceiver scores and administrator ratings ($r = 0.31$) and teacher attendance ($r = 0.30$).

However, the correlations between Teacher Perceiver scores and student ratings ($r = 0.24$) and outside observer ratings ($r = 0.06$) were weaker. The correlation between Teacher Perceiver scores and student gain scores was $r = 0.29$, however, it was based on only one study and it was not statistically significant. The Teacher Perceiver correlations with overall teacher effectiveness were stronger with teachers that were already hired ($r = 0.35$) than with teachers that were hired using the Teacher Perceiver as part of the process ($r = 0.21$).

Urban Teacher Perceiver Interview (UTPI)

The Urban Teacher Perceiver Interview (UTPI) was created by The Gallup Organization as an updated version of the Teacher Perceiver specifically designed for urban districts. The questions in the Urban Teacher Perceiver Interview are based on eleven different themes:

1. Individual Perception
2. Empathy
3. Input Drive
4. Commitment
5. Dedication
6. Caring
7. Initiator
8. Stimulator
9. Concept
10. Involver

11. Positivity

Kanipe (1996) found a positive correlation between UTPI and principals' perceptions of teacher effectiveness, but the correlation was too small to be statistically significant. Chesley (2004) found a statistically significant relationship between UTPI scores and teacher evaluation scores on the Teacher Performance Assessment. There was no statistically significant relationship between the Teacher Performance Assessment and Praxis Series scores. However, the predictability of Teacher Performance Assessment scores was increased when using both the UTPI and the Praxis Series.

TeacherInsight (TI)

TeacherInsight (TI) is a Web-based assessment consisting of multiple choice and Likert items that requires approximately 30 minutes to complete. Results are based on the applicant's responses and result in a score between 0 and 100 which Gallup claims is predictive of an applicant's potential for teaching success based on his or her talent.

The only research on the TI published by Gallup that could be found was a study comparing the TI scores of certified and noncertified applicants. The study analyzed the TI scores of the 13,489 applicants that took the interview between March 2003 and June 2003. Of all the applicants, there were 10,937 applicants who indicated that they were certified and 2,552 applicants who indicated they were not certified. The study found that "The mean scores for the two groups were slightly different. The mean percent score for certified teachers was 63.85 and the mean percent score for noncertified teachers was 62.25. There is no meaningful difference, however, when

using these differences on the TeacherInsight scores” (Gordon, 2003, p. 3).

As a result of those findings, “Gallup believes that TeacherInsight is an appropriate tool to be used as part of the process to select individuals for alternative certification programs” (Gordon, 2003, p.2). The Texas Region 13 Education Service Center Alternative Certification program is one of the programs that uses the TeacherInsight as part of their selection process (Sanford, 2003).

The only independent, published research on the TeacherInsight that could be found was a dissertation study that analyzed the relationship between TeacherInsight interview scores and student performance (Koerner, 2007). Student performance was measured by the Texas Growth Index (TGI), a value-added measure of achievement on the TAKS tests in mathematics and English/reading.

The study found a statistically significant ($p < 0.01$) correlation between TI scores and Texas Growth Index scores in English/reading ($r = 0.045$). The correlation between TI scores and Texas Growth Index scores in math ($r = 0.042$) was also statistically significant ($p < 0.05$). However, while both of these correlations are statistically significant, “both of these values demonstrate positive but very weak relationships.” (Koerner, 2007, p. 57)

Koerner (2007) concluded in the abstract of his dissertation that “the findings provide little support to the validity of TeacherInsight in terms of its ability to predict student achievement scores and its usefulness as a tool for the selection of teachers by school systems. Until more extensive research is completed on the TI and its impact on student achievement, no definitive answers for school systems can be made.”

Summary

A quality education is critical for our students, our communities, our nation, and our future. Teacher effectiveness is the most significant variable of student learning that schools can influence. School and district accountability continues to increase at both the state and federal level. All of these are reasons why we need effective teacher selection and teacher evaluation systems. Commercial teacher interviews such as the TI could be very helpful in hiring if they are able to identify more effective teachers. However, if they lack predictive validity their use could be detrimental as a result of wasted time and money and, even more significantly, the hiring of less effective teachers.

Chapter 3 discusses the methods used to analyze the relationship between TI scores and teacher effectiveness. It includes an explanation of the statistical methods that were utilized to determine the correlation between teachers' TI scores, their eight domain scores of their PDAS evaluations, and their age, years of teaching experience, gender, and highest degree earned.

CHAPTER 3

METHODS AND PROCEDURES

This chapter discusses the methodology used in this study. It includes a restatement of the purpose and research questions and is followed by the context, the study participants, research questions, the methodology, and the analysis of the data.

Purpose

The TeacherInsight (TI) is being used by many school districts across the nation as part of the hiring process. It is also used by many alternative certification programs as part of the admissions process. Some of these districts and programs simply use it as one of many factors, while others have a cut score in which applicants must meet or exceed to have any chance of being selected.

The Gallup Organization charges \$17 per applicant for the administration of the TI. However, the district being studied pays a flat rate of \$42,000 for an unlimited number of administrations of the TI. If the TI is able to identify more effective teacher applicants, then the time and money is well spent. However, if it does not reliably predict teacher effectiveness, then it is actually doing a disservice as it is wasting limited resources and, worse yet, potentially resulting in some less effective teachers being hired. Very little independent research has been conducted on the TI interview tool. Additional research could benefit school administrators who are using, or considering the use of, the TI as part of the teacher selection process. Thus, the purpose of this study is to determine the relationship between TI scores and teacher effectiveness, as measured by the Professional Development and Appraisal System.

Research Questions

This study addressed the following research questions:

1. What are the relationships between teachers' TI scores and each of the eight domain scores of their Professional Development and Appraisal System (PDAS) evaluation ratings?
2. What are the relationships between teachers' TI scores and their age, years of teaching experience, gender, and highest degree earned?
3. What are the relationships between teachers' PDAS evaluation ratings and their age, years of teaching experience, gender, and highest degree earned?
4. To what extent do the eight PDAS domain scores predict the variability in the TI scores over and above that explained by teacher age, gender, years of experience, and highest degree earned?

Context

This study analyzed information from one school district located in North Texas. The district has a population of approximately 250,000 citizens, with approximately 53,000 students. The district has 68 schools including two alternative schools (one is kindergarten through 8th grade and the other is 9th through 12th grade), three early childhood schools (pre-kindergarten), 43 elementary schools (kindergarten through 5th grade), twelve middle schools (6th through 8th grade), five high schools (9th and 10th grade), and three senior high schools (11th and 12th grade).

The district requires all teacher applicants to complete the TI as part of the application process. Each applicant's TI score is available for principals to view and is one of the many factors considered when making hiring decisions. The district recommends a minimum score of 67 for an applicant to be hired. Principals that select an applicant who has a score below 67 are required to send an e-mail to human

resources explaining why they want to hire the person in spite of their low TI score.

The district allows three different appraisal options for teachers including the Professional Development and Appraisal System (PDAS), a three-year rotation, and alternative appraisal/goal setting. However, the three-year rotation and alternative appraisal/goal setting options are only available to teachers on a term contract. Thus, all teachers must use PDAS during their first year in the district since they are on a probationary contract.

Study Participants

The participants for this study were 527 teachers hired into the district for the 2006-2007 school year ($N = 527$). All of these participants completed the TI tool prior to being hired, per requirements set forth by the district's human resources department. All participants were also evaluated by an administrator using the PDAS. Permission to access teachers' TI scores, eight domain scores of their PDAS evaluations, age, years of teaching experience, gender, and highest degree earned was granted by the school district's superintendent's office, human resources department, and assessment and accountability department and the University's Institutional Review Board (IRB) (Appendix A).

Research Design

The study was quantitative in nature and analyzed the relationship between teachers' TI scores, their eight domain scores of their PDAS appraisal, and their age, years of teaching experience, gender, and highest degree earned. The statistical

methods used were correlational and linear regression. Correlations reflect the degree of linear relationship between two variables. The correlation ranges from +1, a perfect positive relationship, to -1, a perfect negative relationship. A correlation of 0 means there is no relationship between the two variables. Linear regression is a regression method that models the relationship between a dependent variable and independent variables.

The teachers' TI scores, eight domain scores of their PDAS ratings, age, years of teaching experience, gender, and highest degree earned were collected from the district's human resources department. The information was coded to protect the identities of the teachers. The data was entered into the SPSS™ (SPSS Inc., <http://www.SPSS.com>) program with the variables of TI score, eight PDAS domain scores, age, years of teaching experience, gender, and highest degree earned.

Instrumentation

The TI was the instrument used as the basis for this analysis; it is used by many school districts and alternative certification programs across the country to assist in the selection process. It is an online instrument developed by the Gallup Organization to analyze teacher *talent*. The scores on the TI are between 0 and 100.

The Gallup Organization has examined the relationship between TI scores and student and principal surveys and between TI scores and student achievement scores, however this research has not been published. The only independent, published research regarding the TI is an analysis of the relationship between TI scores and student performance as measured by the Texas Growth Index (Koerner, 2007). No

published study today addresses the relationship between TI scores and teacher evaluation ratings.

The teachers' PDAS scores were used to evaluate teacher effectiveness. The PDAS is completed by an administrator and produces numerical scores in eight different domains:

Domain I: Active Successful Student Participation in the Learning Process

Domain II: Learner-Centered Instruction

Domain III: Evaluation and Feedback on Student Progress

Domain IV: Management of Student Discipline, Instructional Strategies, Time, and Materials

Domain V: Professional Communication

Domain VI: Professional Development

Domain VII: Compliance with Policies, Operating Procedures and Requirements

Domain VIII: Improvement of Academic Performance of all Students on the Campus

Within each domain there are between four and nine evaluation criteria, for a total of 51 evaluation criteria within the eight domains. Each evaluation criteria is scored as exceeds expectations, proficient, below expectations, or unsatisfactory. Those ratings are assigned point values of five, three, one, and zero, respectively. The points are totaled within each domain and then the domains are also rated as exceeds expectations, proficient, below expectations, or unsatisfactory. All eight domains are rated independently and there is no composite or cumulative score or rating for PDAS.

Data Collection

The data for the teachers' TI and PDAS scores were collected from the district's human resources department. Permission was granted by the director of human resources and the director of assessment and accountability to complete the study. A coding system was used to preserve anonymity while collecting the teachers' TI scores, PDAS ratings, age, years of teaching experience, gender, and highest degree earned. Each participant was given a case number so participants' names did not appear on the recorded information.

Data Analysis

After the data was collected, correlational and linear regression research designs were conducted to determine the relationships between teachers' TI scores, eight domain scores of their PDAS evaluations, age, years of teaching experience, gender, and highest degree earned. These analyses determined the ability of the TeacherInsight to identify more effective teachers, based on their PDAS evaluations.

In order to answer the research questions, a correlation was calculated between each of the eight Professional Development and Appraisal System domain scores and the TeacherInsight scores. The same method was used to determine the correlation between teachers' TI scores and their ages, years of teaching experience, gender, and highest degree earned. Finally, the same method was also used to determine the correlation between the eight domain scores of the teachers' PDAS evaluations and their age, years of teaching experience, gender, and highest degree earned. Linear regression was conducted with TI score as the dependent variable and the eight PDAS

domain scores, age, years of teaching experience, gender, and highest degree earned as predictor variables.

Summary

This chapter provides an overview of the purpose, context, and participants of this study. It explains the research questions, research design, and the data collection and analysis that were used to compare teachers' TI scores, eight PDAS domain scores, age, years of teaching experience, gender, and highest degree earned.

The results of this study are important to districts as they try to hire the most effective teachers. As the pressure of accountability increases at both the state and federal level, schools and school districts need to find ways to increase their students' achievement levels. Locating and hiring the most effective teachers is a critical method to increasing student achievement. Standardized interview tools such as the TI need to be studied to determine if they are worth the time and money by whether or not they are able to identify more effective teachers.

Chapter 4 presents the results of the quantitative data analysis. The statistical analyses of teachers' TI scores, eight domain scores of their PDAS evaluations, age, years of teaching experience, gender, and highest degree earned are shared and the significance of the findings discussed.

CHAPTER 4

PRESENTATION AND ANALYSIS OF THE DATA

Introduction

The results from the analyses of the data presented in this chapter are divided into several sections. The first section consists of descriptive data of the population used in this study. The middle sections include information about TeacherInsight (TI), Professional Development and Appraisal System (PDAS), the presentation and analyses of the statistics as they relate to the four research questions. The final section of this chapter is the summary.

Descriptive Data

Description of the Population

The population studied was comprised of 527 teachers hired in a North Texas school district for the 2006-2007 school year. The majority, (82.9%) of the teachers were female and 17.1% were male. A slight majority (51.6%) taught at the elementary level, while 48.4% taught at the secondary level. There was a greater proportion (93%) of female teachers at the elementary level compared to the secondary level (72.2%).

Most of the teachers (75.9%) were white and the rest were Hispanic (12.3%), African American (6.5%), Asian (4.6%), or American Indian (0.8%). The elementary teachers were slightly more diverse (71% White, 16.9% Hispanic, 6.6% African American, 4.4% Asian, and 1.1% American Indian) than the secondary teachers (81.2% White, 7.5% Hispanic, 6.3% African American, 4.7% Asian, and .4% American Indian).

The teachers' ages ranged from 21 years to 68 years, with a mean of 31.8 years,

a standard deviation of 9.4 years, and a median of 29 years. The average age of elementary teachers was 30.9 years, with a standard deviation of 9.0 years. Secondary teachers' average age was 32.9 years, with a standard deviation of 9.8 years.

The teachers' years of teaching experience ranged from zero to 34, with a mean of 3.1 years, a standard deviation of 4.9 years, and a median of 1 year. The average number of years of experience for elementary teachers was 2.9 with a standard deviation of 4.6 years. The average number of years of experience for secondary teachers was 3.3 with a standard deviation of 5.2 years.

The highest degree earned for the majority (78.2%) was a bachelor's degree, while 20.9% had a master's degree and 0.9% had a doctorate. A smaller proportion of the elementary teachers had master's degrees (17.3%) than the secondary teachers (24.7%). All three of the teachers with doctorate degrees taught at the secondary level.

TeacherInsight

The teachers' TI scores ranged from a low of 43 to a high of 89, with a mean of 66.4, a standard deviation of 8.0, and a median of 67. Almost half (47.4%) of the teachers had TI scores below the cutoff score of 67 that is recommended by the district. Thus, the principals needed to send e-mails to the human resources department to justify the selection of 250 out of the 527 teachers.

The elementary teachers' TI scores ranged from 47 to 89, while the secondary teachers' TI scores ranged from 43 to 85. The average TI score for elementary teachers was 67.9, with a standard deviation of 7.4 and a median of 68. The secondary teachers' average TI score was 64.8, with a standard deviation of 8.2 and a median of 65. There

were a smaller proportion of elementary teachers (40.1%) than secondary teachers (55.3%) that had a TI score below the district recommended cutoff score of 67.

Professional Development and Appraisal System

Each domain of the PDAS contains criteria which are scored as zero, one, three, or five points. The scores for the eight domains of the PDAS vary considerably because they contain a different number of criteria. Thus, a better way to compare the scores of the different domains is to calculate the mean score per criteria for each domain. Table 1 displays the minimums, maximums, means, standard deviations, medians, number of criteria, and the means divided by the number of criteria for each of the eight domains of the PDAS scores for the 527 teachers in the study.

Table 1

PDAS Domain Scores Descriptive Data

	Min	Max	Mean	SD	Median	Criteria	Mean/Criteria
Domain I	5	25	19.1	3.3	19	5	3.8
Domain II	13	45	33.8	5.2	33	9	3.8
Domain III	8	30	23.0	3.8	24	6	3.8
Domain IV	6	40	30.2	5.4	30	8	3.8
Domain V	6	30	23.5	4.0	24	6	3.9
Domain VI	10	20	14.0	2.3	14	4	3.5
Domain VII	3	15	10.4	2.1	9	3	3.5
Domain VIII	19	45	34.0	6.1	33	9	3.8

Data Analysis

The four research questions stated previously were designed to provide information to determine the extent to which the TI was of value in predicting variability in the eight domain scores of the PDAS evaluations. Each research question is examined in this section with corresponding data charts presenting the statistical findings.

Research Question 1: What are the relationships between teachers' TeacherInsight scores and each of the eight domain scores of their PDAS evaluation ratings?

In order to answer the first research question, a Pearson's product-moment correlation was obtained between each of the eight domain scores of PDAS and the TeacherInsight (TI) scores. These are shown in Table 2.

The eight domain scores of PDAS had moderate to high correlations with each other with Pearson r 's ranging from 0.26 to 0.81 and all of them statistically significant at the 0.01 level. However, the correlations between the TI scores and seven of the eight domains of PDAS were not statistically significant at even the 0.05 level and the Pearson r 's ranged from only 0.05 to 0.08. The correlation between domain 5 PDAS scores and TI scores was statistically significant at the 0.01 level. However, that correlation was still relatively weak ($r = 0.14$) and so only two percent of the variance of domain 5 PDAS scores could be explained by the TI scores.

When analyzed separately by teaching level, none of the correlations between the eight PDAS domain scores and the TI scores were statistically significant at the 0.05 level for the secondary teachers or for the elementary teachers. The correlation between domain 5 PDAS scores and TI scores produced Pearson r scores of 0.11 for

both secondary teachers and elementary teachers; however, they were not statistically significant at the 0.05 level.

Table 2

Correlation Matrix PDAS Domain Scores and TeacherInsight Scores

	1	2	3	4	5	6	7	8
PDAS Domain I (1)								
PDAS Domain II (2)	0.81**							
PDAS Domain III (3)	0.72**	0.76**						
PDAS Domain IV (4)	0.62**	0.65**	0.60**					
PDAS Domain V (5)	0.50**	0.56**	0.56**	0.56**				
PDAS Domain VI (6)	0.36**	0.38**	0.40**	0.31**	0.49**			
PDAS Domain VII (7)	0.26**	0.37**	0.36**	0.29**	0.47**	0.42**		
PDAS Domain VIII (8)	0.42**	0.50**	0.44**	0.37**	0.54**	0.51**	0.46**	
TeacherInsight (9)	0.06	0.05	0.06	0.06	0.14**	0.05	0.08	0.05

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

Research Question 2: What are the relationships between teachers' TeacherInsight scores and their age, years of teaching experience, gender, and highest degree earned?

In order to answer the second research question, a Pearson's product-moment correlation was obtained between the teachers' TI score and the continuous variables (age and years of teaching experience) and a Spearman's *rho* was calculated between TI scores and the non-continuous variables (gender, and highest degree earned). Table 3 lists the correlations and their statistical significance.

Highest degree earned and years of experience did not have statistically significant correlations with TI scores. The correlation between gender and TI scores

was statistically significant at the 0.05 level. On average, females scored slightly higher on the TI than males, however, this relationship was relatively weak with a Spearman's *rho* of only -0.12. Thus, gender accounted for 1.4% of TI score variance.

The correlation between age and TI scores was statistically significant at the 0.01 level. Younger teachers, on average, scored higher on the TI than older teachers, however, this relationship was relatively weak ($r = -0.154$). Thus, age was only able to explain 2.4% of the variance of TI scores.

Table 3

Correlation Matrix Teacher Characteristics and TeacherInsight Scores

	1	2	3	4
Gender (1)				
Age (2)	0.05			
Years of Experience (3)	0.05	0.55**		
Highest Degree Earned (4)	0.09*	0.25**	0.27**	
TeacherInsight (5)	-0.12**	-0.15**	-0.04	-0.01

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

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

Research Question 3: What are the relationships between teachers' eight PDAS domain scores and their age, years of teaching experience, gender, and highest degree earned?

In order to answer the third research question, a Pearson's product-moment correlation was obtained between the teachers' eight domain scores of their PDAS evaluations and the two continuous teacher variables (age and years of teaching experience) and a Spearman's *rho* was calculated between the eight PDAS domain

scores and the non-continuous variables (gender and highest degree earned). Table 4 lists the correlations and their statistical significance.

Teachers' gender had statistically significant correlations with four of the PDAS domain scores. The correlations between gender and domain II were significant at the 0.05 level and the correlations between gender and domains IV, V, and VIII were significant at the 0.01 level. Female teachers, on average, had higher scores than male teachers in those four PDAS domains; however, those correlations were relatively weak with Spearman *rho*'s ranging from -0.09 to -0.13. Thus, gender was only able to explain approximately one to two percent of the variance in each of those four PDAS domain scores.

Years of experience, was the teacher variable that produced the greatest number of statistically significant correlations with PDAS domain scores. The only PDAS domain that years of experience did not have a statistically significant correlation with was domain VII. The correlation between teachers' years of experience and domain V was significant at the 0.05 level and the correlations between years of experience and the other six PDAS domains were statistically significant at the 0.01 level. Teachers with more experience, on average, scored higher on those PDAS domains. However, those correlations were weak with Pearson *r*'s ranging from 0.09 to 0.21. Thus, the amount of teaching experience was only able to account for approximately one to four percent of each of the PDAS domain scores.

The correlation between highest degree earned and PDAS domain I was significant at the 0.05 level. Teachers with higher degrees, on average, scored slightly higher on this PDAS domain. However, the relationship was weak (Spearman's *rho* of

0.09) so the highest degree earned by the teachers was only able to less than one percent of the variance of the PDAS domain I scores. The correlations between highest degree earned and the other seven PDAS domains were not statistically significant. Finally, none of the eight PDAS domain scores had a statistically significant correlation with age.

Table 4

Correlation Matrix PDAS Domain Scores and Teacher Characteristics

	1	2	3	4	5	6	7	8	9	10	11
PDAS Domain I (1)											
PDAS Domain II (2)	0.81**										
PDAS Domain III (3)	0.72**	0.76**									
PDAS Domain IV (4)	0.62**	0.65**	0.60**								
PDAS Domain V (5)	0.50**	0.56**	0.56**	0.56**							
PDAS Domain VI (6)	0.36**	0.38**	0.40**	0.31**	0.49**						
PDAS Domain VII (7)	0.26**	0.37**	0.36**	0.29**	0.47**	0.42**					
PDAS Domain VIII (8)	0.42**	0.50**	0.44**	0.37**	0.54**	0.51**	0.46**				
Gender (9)	0.00	-0.09*	-0.06	-0.13**	-0.11**	-0.04	-0.07	-0.13**			
Age (10)	0.03	0.01	0.06	-0.01	-0.03	0.06	0.02	0.01	0.06		
Years of Experience (11)	0.21**	0.16**	0.18**	0.17**	0.09*	0.13*	0.01	0.14**	0.04	0.55**	
Highest Degree Earned (12)	0.09*	0.08	0.08	0.02	0.01	0.07	0.02	0.07	0.09*	0.26**	0.26**

* Correlation is significant at the 005 level 2-tailed.

**Correlation is significant at the 001 level 2-tailed.

Research Question 4: To what extent do the eight PDAS domain scores predict the variability in the TeacherInsight scores over and above that explained by teacher age, gender, years of experience, and highest degree earned?

In order to answer the research question 4, a regression analysis was conducted. TeacherInsight score was the dependent variable with the eight domain scores of PDAS as the independent variables. The variables for teacher age, gender, years of experience, and highest degree earned were controlled by entering them as the first step in a sequential regression model.

In the first step of the regression $F(4, 522) = 4.753, p < .01$ two of the four independent variables were statistically significant predictors. Gender ($\beta = -0.092, p < 0.05$) accounted for 26% of the predicted TI score. Age ($\beta = -0.188, p < 0.01$) explained 67.2% of this criterion variable. Highest degree earned ($B = 0.701, \beta = 0.036$) and years of teaching experience ($B = 0.072, \beta = 0.054$) did not reach significance and only accounted for 0.1% and 5.6% of the variance, respectively. However, none of the variables were very strong predictors as only 3.5% of the variance in TI scores was accounted for by all four independent variables in this regression model.

For the second step in this regression $F(12, 514) = 2.432, p < 0.01$ three of the independent variables were statistically significant predictors. Gender ($\beta = -0.09, p < 0.05$) accounted for 16.8% of predicted TI score. Age ($\beta = -0.18, p < 0.01$) explained 34.8% of the variance of this criterion variable. PDAS domain 5 ($\beta = 0.15, p < 0.05$) accounted for 33.6% of the variance. Education $\beta = 0.04$, Experience $\beta = 0.05$, PDAS domain 1 $\beta = 0.06$, PDAS domain 2 $\beta = -0.09$, PDAS domain 3 $\beta = 0.013$, PDAS domain 4 $\beta = -0.027$, PDAS domain 6 $\beta = -0.013$, PDAS domain 7 $\beta = 0.046$, and PDAS domain 8 $\beta = -0.039$ did not reach statistical significance and each only accounted for between

0% and 11.6% of the variance of the predicted TI score.

However, none of the 12 predictor variables were very strong predictors as they only accounted for 5.4% of the variance in TI scores in this regression model. The eight domain scores of PDAS were only able to explain 1.9% of the variance in TI scores over and above that explained by age, gender, experience, and highest degree earned in the first step of the regression model.

Table 5

Hierarchical Multiple Regression of PDAS Domain Scores on TeacherInsight Score

Variables	<i>B</i>	<i>SE B</i>	β	<i>t</i>	r_s	r_s^2	R^2	ΔR^2
Step 1							0.035	0.035
Age	-0.16	0.04	-0.19	-3.63**	0.82**	0.672		
Education	0.64	0.82	0.04	0.79	-0.03	0.001		
Experience	0.09	0.08	0.05	1.03	-0.24**	0.058		
Gender	-1.94	0.91	-0.09	-2.13*	-0.51**	0.260		
Step 2							0.054	0.019
Age	-0.153	0.044	-0.18	-3.46**	-0.59**	0.348		
Education	0.701	0.820	0.04	0.86	-0.02	0.000		
Experience	0.072	0.087	0.05	0.83	-0.16**	0.026		
Gender	-1.860	0.931	-0.09	-2.0*	-0.41**	0.168		
Domain I	0.153	0.186	0.07	0.82	0.22**	0.048		
Domain II	-0.131	0.132	-0.09	-0.99	0.16**	0.026		
Domain III	0.026	0.151	0.01	0.18	0.21**	0.044		
Domain IV	-0.040	0.091	-0.03	-0.44	0.22**	0.048		
Domain V	0.290	0.124	0.15	2.34*	0.58**	0.336		
Domain VI	-0.046	0.186	-0.01	-0.25	0.21**	0.044		
Domain VII	0.178	0.202	0.05	0.88	0.34**	0.116		
Domain VIII	-0.051	0.075	-0.04	-0.68	0.22**	0.048		

* $p < 0.05$; ** $p < 0.01$; $N = 527$

Summary

In the description of variables, it was noted that 47.4% of the 527 teachers had TI scores below the district recommended cutoff of 67. Differences were found in TI scores based on level of assignment, with elementary teachers scoring slightly higher than those teachers assigned to secondary schools.

Research question 1 examined the correlations between the eight domain scores of PDAS and the TI scores. Domain V was the only one of the eight PDAS domain scores found to have a statistically significant correlation with TI scores. However, that correlation was found to be relatively weak ($r = 0.14$). Thus, TI scores were only able to explain two percent of the variance of PDAS domain V scores. Domain V of PDAS is titled Professional Communication. The correlations between TI scores and the other seven PDAS domain scores ranged from 0.05 to 0.08 and were not statistically significant.

Research question 2 examined the correlations between the teachers' age, gender, years of experience, and highest degree earned and the teachers' TI scores. Experience and highest degree earned did not have statistically significant correlation with TI scores. The correlation between gender and TI scores was significant at the 0.05 level, but relatively weak ($r = -0.12$). The correlation between age and TI scores was significant at the 0.01 level, but also weak ($r = -0.15$). Thus, gender and age could each explain only approximately one to two percent of the variance of TI scores.

Research question 3 examined the correlations between teachers' age, gender, years of experience, and highest degree earned and their eight domain scores of the PDAS evaluations. Age did not have a statistically significant correlation with any of the

eight PDAS domains. The other three teacher variables (gender, years of experience, and highest degree earned) did have a statistically significant correlation with one or more of the eight PDAS domains. However, all of these correlations were relatively weak (Pearson r 's and Spearman ρ 's ranging from -0.13 to 0.21) and each accounted for only approximately one to four percent of the variance of the PDAS domains.

Research question 4 examined the amount of variation in TI scores that the eight domain scores of PDAS would be able to explain, above and beyond that explained by age, gender, years of experience, and highest degree earned. A regression analysis found that the four teacher variables were able to explain 3.5% of the variance in TI scores. When the eight PDAS domain scores were added to the second step of the regression model, a total of 5.4% of the variation in TI scores was explained. Thus, the eight domain scores of PDAS were only able to explain an additional 1.9% of the variance in TI scores.

Chapter 5 provides an interpretation of the results, along with discussion, recommendations and suggestions for future research.

CHAPTER 5
SUMMARY, DISCUSSION, RECOMMENDATIONS, AND SUGGESTIONS
FOR FUTURE RESEARCH

Introduction

This chapter presents the background, overview of the study, and a brief review of methodology. Next, a summary of the findings and discussion of results are presented. The final sections include recommendations, suggestions for future research, and the conclusion.

Background

Education is important to individuals and to society as a whole. Many studies have found the benefits frequently connected to a good education and the impact that lack of education has on individuals and society (ACT, 2005; Adelman, 1999; Alliance for Excellent Education, 2008a; Alliance for Excellent Education, 2008b; Alliance for Excellent Education, 2008c; Bedsworth, Colby, & Doctor, 2006; Conley, 2005; Hodgkinson, 2000; National Center for Education Statistics, 2004; Orfield, Losen, Wald, & Swanson, 2004; U.S. Census Bureau, 2008).

Researchers have found that teacher effectiveness is the most significant factor of student achievement that schools can control (Darling-Hammond & Bransford, 2005; Darling-Hammond, 2000; Wilson, Floden, & Ferrini-Mundy, 2001). Thus, teacher selection is an extremely important responsibility of school administrators. The process of selecting and hiring teachers requires significant amounts of our two most limited resources – time and money. Standardized interview tools attempt to help

administrators identify the most effective teachers. Many of these tools are lists of questions and scoring rubrics that are conducted in traditional interview settings – one-on-one with an interviewer and an interviewee. However, in an attempt to save time and be more convenient, some of them are now administered via the internet. One of these Web-based interview tools is the TeacherInsight (TI), which was developed and is sold by the Gallup Organization.

The need for highly effective teachers continues to increase along with student achievement standards under federal and state accountability. The federal No Child Left Behind Act (NCLB) requires all public schools and districts to achieve adequate yearly progress (AYP) as demonstrated by student test scores. Many states also have state-level accountability systems in place as well. Texas has the Academic Excellence Indicator System (AEIS) which rates all public schools and districts based on student test scores, attendance rates, drop out rates, and completion rates. These federal and state ratings are very significant as they are publicized extensively via the media, websites, and campus report cards. They are more than just a label as they can impact property values for the attendance zone for that school or district. Low ratings can also have significant consequences for schools or districts as they will be required to develop corrective action plans and make changes to earn a higher rating the following year. Educators strive to identify methods to increase student achievement to the levels mandated by NCLB and AEIS. Since teacher effectiveness is one of the most significant factors that affect student achievement, school leaders need to use effective hiring practices to select the best teachers.

Overview of the Study

This study involved an examination of the relationships between scores on the TI interview tool and the eight domain scores of the Professional Development and Appraisal System (PDAS). The study also analyzed the relationships between teacher age, gender, years of experience, and highest degree earned and TI scores. Also examined was the relationship of those same teacher characteristics and the eight PDAS domain scores. Finally, the ability of the PDAS scores to predict the variation of the TI scores, above and beyond the amount already explained by the teacher characteristics, was also analyzed.

Review of the Methodology

This study utilized archived data from a large, suburban school district in North Texas, which used the TI as a screening instrument for all teacher applicants. Information collected for 527 teachers hired for the 2006-2007 school year included demographic and qualification characteristics, scores on the TI, and the eight domain scores of their PDAS. Correlation and hierarchical regression analyses were conducted utilizing SPSS™ 16.0 to answer the four research questions.

Summary of the Findings

Description of the Population

The population studied was comprised of 527 teachers hired in a North Texas school district for the 2006-2007 school year. The majority (82.9%) of the teachers were

female and 17.1% were male. A slight majority (51.6%) of them taught at the elementary level, while 48.4% taught at the secondary level. There was a greater proportion (93%) of female teachers at the elementary level compared to the secondary level (72.2%).

Most of the teachers (75.9%) were white and the rest were Hispanic (12.3%), African American (6.5%), Asian (4.6%), or American Indian (0.8%). The elementary teachers were slightly more diverse (71% White, 16.9% Hispanic, 6.6% African American, 4.4% Asian, and 1.1% American Indian) than the secondary teachers (81.2% White, 7.5% Hispanic, 6.3% African American, 4.7% Asian, and .4% American Indian).

The teachers' ages ranged from 21 years to 68 years, with a mean of 31.8 years, a standard deviation of 9.4 years, and a median of 29 years. The average age of elementary teachers was 30.9 years, with a standard deviation of 9.0 years. Secondary teachers' average age was 32.9 years, with a standard deviation of 9.8 years.

The teachers' years of teaching experience ranged from zero to 34, with a mean of 3.1 years, a standard deviation of 4.9 years, and a median of 1 year. The average number of years of experience for elementary teachers was 2.9 with a standard deviation of 4.6 years. The average number of years of experience for secondary teachers was 3.3 with a standard deviation of 5.2 years.

The highest degree earned for the majority (78.2%) was a bachelor's degree, while 20.9% had a master's degree and 0.9% had a doctorate. A smaller proportion of the elementary teachers had master's degrees (17.3%) than the secondary teachers (24.7%). All three of the teachers with doctorate degrees taught at the secondary level.

TeacherInsight (TI)

The teachers' TI scores ranged from a low of 43 to a high of 89, with a mean of 66.4, a standard deviation of 8.0, and a median of 67. Almost half (47.4%) of the teachers had TI scores below the cutoff score of 67 that is recommended by the district. Thus, the principals needed to send e-mails to the human resources department to justify the selection of 250 out of the 527 teachers. The elementary teachers' average TI score was approximately three points higher than the secondary teachers and there was a smaller proportion of elementary teachers (40.1%) than secondary teachers (55.3%) that had a TI score below the district recommended cutoff score of 67.

Professional Development and Appraisal System (PDAS)

The scores in each PDAS domain varied considerably due to the different number of criteria in each domain. However, when analyzed by the mean score per criteria, the eight domains were relatively similar. Domain VII had the lowest mean score per criteria (3.5), while domain V had the highest (3.9). The mean criteria score in all eight domains of PDAS was 3.7.

Research Question 1: What are the relationships between teachers' TeacherInsight scores and each of the eight domain scores of their PDAS evaluation ratings?

Research question 1 examined the relationships between the eight domain scores of PDAS and the TI scores. Domain V was the only one of the eight PDAS domain scores found to have a statistically significant correlation with TI scores. However, even that correlation was found to be relatively weak ($r = 0.14$).

Research Question 2: What are the relationships between teachers' TeacherInsight scores and their age, years of teaching experience, gender, and highest degree earned?

Research question 2 examined the correlations between the teachers' age, gender, years of experience, and highest degree earned and the teachers' TI scores. Experience and highest degree earned did not have statistically significant correlation with TI scores. The correlation between gender and TI scores was significant at the 0.05 level. Female teachers, on average, scored higher on the TI than male teachers, and the correlation between age and TI scores was significant at the 0.05 level. Younger teachers, on average, scored higher on the TI than older teachers, and the correlation between age and TI scores was significant at the 0.01 level. However, even the two statistically significant correlations were relatively weak.

Research Question 3: What are the relationships between teachers' eight PDAS domain scores and their age, years of teaching experience, gender, and highest degree earned?

Research question 3 examined the correlations between teachers' age, gender, years of experience, and highest degree earned and their eight domain scores of the PDAS evaluations. Age did not have a statistically significant correlation with any of the eight PDAS domains.

Teachers' gender had statistically significant correlations with four of the PDAS domain scores. Highest degree earned had statistically significant correlations with one PDAS domain score. The teacher variable that produced statistically significant correlations with the greatest number (seven) of PDAS domain scores was years of teaching experience. However, while these 12 correlations were statistically significant,

the relationships were weak (Spearman's ρ or Pearson r ranging from -0.09 to 0.21), and each were only able to explain between one and four percent of the variation.

Research Question 4: To what extent do the eight PDAS domain scores predict the variability in the TeacherInsight scores over and above that explained by teacher age, gender, years of experience, and highest degree earned?

Research question 4 examined the amount of variation in TI scores that the eight domain scores of PDAS were able to explain, above and beyond that explained by age, gender, years of experience, and highest degree earned. A regression analysis found that the four teacher variables were able to explain 3.5% of the variance in TI scores. When the eight PDAS domain scores were added to the second step of the regression model, a total of 5.4% of the variation in TI scores was explained. Thus, the eight domain scores of PDAS were able to explain an additional 1.9% of the variance in TI scores.

Discussion of the Results

Each of the eight domain scores of PDAS all had moderate to high correlations with the other seven domains. The r values ranged from 0.26 to 0.81 and all of them were significant at the 0.01 level. Thus, teachers that did well on one domain of PDAS frequently also scored well on other domains. Conversely, teachers that did not do well on one PDAS domain often did not do well on other domains.

Female teachers, on average, scored slightly higher on the TI than male teachers. While the reason for this is unknown, one possible explanation is that the TI could be assessing personality characteristics more commonly found in females. For

example, some of the questions on the TI may evaluate the communication style of the teacher and females may more frequently respond with answers that Gallup considers correct. Younger teachers also tended to score higher on the TI than older teachers. One reason for this could be that younger teachers are more effective at completing a Web-based assessment. However, both of these correlations were weak and each only accounted for between one and two percent of the variance of TI scores.

More experienced teachers, on average, scored slightly higher than less experienced teachers on several of the PDAS domains. This could be evidence of experienced teachers being more effective at teaching. However, it could also be that experienced teachers are more effective at the PDAS evaluation itself. For example, experienced teachers may have a better understanding of what administrators will be looking for when completing PDAS observations. Experienced teachers may also be more effective at completing the teacher self-report forms. Regardless of the reasons, these correlations were relatively weak and each of the teacher variables was able to account for only between one and four percent of the variation of the PDAS domain scores.

Female teachers, on average, scored slightly higher than male teachers on some of the PDAS domains. The females in the population studied could be more slightly more effective teachers than males, or it could be that the PDAS evaluates areas of teaching in which females tend to be more effective. Regardless, the correlations were weak and gender only accounted for less than two percent of the variance of each PDAS domain score.

The TeacherInsight scores did not have a statistically significant correlation with

seven of the eight PDAS scores. The only domain that did have a significant correlation with TI scores was domain V, however, even that correlation was relatively weak ($r = 0.14$). Thus, TI scores could only account for two percent of the variation of the domain V scores. PDAS domain V is titled Professional Communication. All eight PDAS domains were only able to account for 1.9% of the variance above and beyond what was already explained by the four teacher variables.

Perhaps even more interesting than the weak relationships between the TI scores and the PDAS domain scores is the distribution of TI scores itself. One potential limitation of this study could have been that the distribution of TI scores would be limited and skewed due to the district recommendation of only hiring teachers with TI scores of 67 or higher. The population studied was teachers that were hired, so if principals always complied with that recommendation, then none of the participants would have TI scores lower than a 67. Thus, the population studied could have had TI scores in a limited range that was not representative of all the applicants that completed the TI.

However, this was not the case as the TI scores of the teachers hired ranged from a low of 43 to a high of 89. The average score was 66.4, with a standard deviation of 8.0. Thus, the average score was actually below the score cutoff recommended by the human resources department. In fact, almost half (47.4%) of the teachers had TI scores below the cutoff score of 67. Thus, the principals needed to send e-mails to the human resources department to justify the selection of 250 out of the 527 teachers.

The average score of elementary teachers' TI scores (67.9, with a standard deviation of 7.4) was higher than the average TI score of secondary teachers (64.8, with a standard deviation of 8.2). There were a smaller proportion of elementary teachers

(40.1%) than secondary teachers (55.3%) that had a TI score below the district recommended cutoff score of 67.

The large percentage of teachers hired with TI scores below the district recommended cutoff score of 67 casts doubt on the extent of consideration that principals give TI scores in the selection process. If principals believed that the TI is able to identify successful teachers then it is difficult to explain why 47.4% of the teachers they hired had TI scores below the district recommended cutoff.

The weak relationships between the TI scores and PDAS scores bring into question the predictive validity of the TI. These findings are consistent with those of Koerner (2007) that studied the relationship of TI scores and student performance. That was the only published study that could be found regarding the ability of the TeacherInsight at being able to discriminate between more effective and less effective teachers.

Thus, this study and Koerner's 2007 study both found very weak correlations between TI scores and measures of teacher effectiveness. While Koerner's study focused on the relationship between TI and student achievement scores, the new information gained through this study is the weak relationship between the eight domain scores of PDAS evaluations and TI scores. PDAS can be a very effective evaluation tool, as shown in a study by Davis-Frost (2000) which determined that "Texas' Professional Development and Appraisal System significantly impacts the classroom practices in the areas of learner centered instruction, classroom management, support for all students, professional growth, communication, TAAS improvement, and application of learning" (p. 58).

Recommendations

School districts that are using, or considering the use of, the TI should require the Gallup Organization to show them the research in support of the instrument. The lack of independent research on the TI and Gallup's unwillingness to publish their own research does not help support the credibility of the TI. If Gallup was to publish their own studies, then independent researchers could attempt to replicate those studies in order to either confirm or contradict their findings regarding the validity of the TI. Most large school districts have a research department with the capability of conducting such research. The methodology of the study should be based on what the district hopes the TI is doing, or will do, to improve their teacher selection process. For example, if the district's goal is simply to make the process more efficient and faster, the research should analyze the time spent selecting teachers with and without the TI. However, if the district's goal is to increase the quality of the teachers hired, the study should include teacher evaluations, student achievement data, and/or other criteria to determine teacher effectiveness. Studies by Jacob and Lefgren (2005) and Podgursky and Springer (2006) both found that teacher evaluations are effective predictors of student achievement.

Potentially more important, yet simpler, research for a district to conduct is regarding the confidence level the principals and human resource staff have in the TI. Even if the TI is highly effective in identifying better teachers, it will not be useful in a district if those doing the hiring do not have faith in the instrument and thus, do not consider the TI scores in the hiring process. This information could be obtained through anonymous surveys with the principals and other administrators that hire teachers.

Suggestions for Further Studies

One possible topic for further study could be the relationship between TI scores and the criteria within PDAS domain V. This was the only PDAS domain score that produced a statistically significant correlation with TI scores. While this correlation was still relatively weak ($r = 0.14$), one or more of the six criteria that comprise domain V may be more closely related with TI scores.

Domain V is titled Professional Communication and contains six criteria:

1. The teacher uses appropriate and accurate written communication with students.
2. The teacher uses appropriate and accurate verbal and non-verbal communication with students.
3. The teacher encourages and supports students who are reluctant and having difficulty.
4. The teacher uses appropriate and accurate written communication with parents, staff, community members, and other professionals.
5. The teacher uses appropriate and accurate verbal and non-verbal communication with parents, staff, community members, and other professionals.
6. The teacher's interactions are supportive, courteous and respectful with students, parents, staff, community members, and other professionals (Texas Education Agency, 2005, pp. 97-102).

Each criteria is scored as exceeds expectations (five points), proficient (three points), below expectations (one point), or unsatisfactory (zero points). Determining which criteria are more closely related with TI scores could help provide greater understanding to what the TeacherInsight evaluates and what information TI scores provide regarding applicants.

Another possible topic for further study is to assess the principals' attitudes and

confidence level regarding the TI. Even if the TI is extremely effective at identifying better teachers, it won't work unless the principals utilize the scores in the selection process. Examining principals' attitudes regarding the TI could help explain why 47.4% of teachers hired had TI scores below the district recommended cutoff score of 67. This could be studied through quantitative methods, such as surveys, and/or qualitative methods, such as interviews. Another source of information could be reviewing the e-mails sent by principals to the human resources justifying the hiring of teachers with TI scores below 67. The information gathered could then be analyzed in the context of the TI scores of the teachers hired by each of those principals.

An additional topic for further study would be to determine the relationship of individual items of the TeacherInsight with PDAS domain scores. While the overall TI score produced weak correlations with the PDAS domain scores, some of the individual items may be more closely related to TI scores. However, it would be difficult to gain approval from Gallup to conduct this study. Metzger and Wu (2006) studied Gallup's Teacher Perceiver Interview and wrote the following:

Gallup takes its proprietary rights over the TPI very seriously and required a signed agreement before sharing validity data with us. In May, 2004, Gallup agreed to our research protocol and gave us their available studies to include in our synthesis. Information about TPI training and instrumentation is not readily available to the public. (p. 19)

Further research regarding TI scores and their relationships with teacher evaluations and/or student achievement growth would also be beneficial. This research could be an extended time study in order to determine the ability of the TI to identify the effectiveness of teachers for multiple years after being hired by a district. Additional research could involve Texas-based measures (PDAS and TAKS), or it could involve

the teacher and student evaluation systems in other states. Henson (2006) noted, “Perhaps the single most critical weakness in much of social science research is the lack of emphasis on replication” (p. 621). Additional research could verify or contradict the findings of Koerner (2007) and this study.

Finally, additional research regarding teacher selection methods could be beneficial. There are a lot of other hiring processes and strategies being utilized by schools and districts, such as reviewing applications and resumes, reference checks, individual and group interviews, and background checks. Identifying best practices regarding the selection of effective teachers could benefit school districts and ultimately increase student achievement.

Conclusion

Educational leaders need to continue trying to identify and select the best people for their teaching positions. This is critical to the success of the schools, the districts, and the students. Douglas Gephart, retired superintendent and past president of the American Association of School Personnel Administrators, wrote the following:

The fundamental difference between a high-performing school district and one struggling to close the achievement gap can be found inside the doors of your human resource department. That’s where the all-important duty of recruiting, supporting and retaining high-caliber teachers and other staff members is a make-or-break proposition.

In Fremont, Calif., a highly diverse district of nearly 32,000 students whose families speak 114 different languages, we recognized the No. 1 factor influencing student achievement is directly related to the quality of instruction provided by the classroom teacher.

Decisions involving school district personnel should not be taken lightly. Every hiring decision will move the district one step closer to or one step further from the path of educational excellence.

To maximize the human resources department's contribution, district leadership must create a vision that aligns the interview and selection process with the desired knowledge, skills and attributes of school district staff (Gephart, 2009, pp. 24-25).

School districts need to make sure the selection strategies they are using are effective. The principals and other staff members in charge of the hiring also need to have confidence in the effectiveness of these strategies if they are to successfully implement them. Independent research of these strategies can enable school districts to determine the level of effectiveness and the consistency of implementation of their hiring practices. This will enable school districts to select more effective teachers that can raise student achievement to the levels mandated by NCLB and AYP.

APPENDIX
INSTITUTIONAL REVIEW BOARD (IRB)
APPROVAL LETTER



OFFICE OF THE VICE PRESIDENT FOR RESEARCH
Office of Research Services

November 26, 2007

Michael Novotny
Department of Teacher Education and Administration
University of North Texas

RE: Human Subjects Application No. 07-443

Dear Mr. Novotny:

In accordance with 45 CFR Part 46 Section 46.101, your study titled "Predictive Validity of Teacher Insight Screening Tool" has been determined to qualify for an exemption from further review by the UNT Institutional Review Board (IRB).

No changes may be made to your study's procedures or forms without prior written approval from the UNT IRB. Please contact Shelia Bourns, Research Compliance Administrator, ext. 3940, if you wish to make any such changes.

Sincerely,

Kenneth W. Sewell, Ph.D.
Chair
Institutional Review Board

KS:sb

CC: Dr. Jane Huffman

REFERENCES

- Aaronson, D. Barrow, L., & Sander, W. (2003). *Teachers and student achievement in the Chicago public high schools*. Working paper 2002-28. Chicago: Federal Reserve Bank of Chicago.
- ACT. (2005). *Crisis at the core: Preparing all students for college and work access*. Iowa City, IA: ACT, Inc.
- Adamowski, S., Therriault, S.B., & Cavanna, A.P. (2007). *The autonomy gap: Barriers to effective school leadership*. Retrieved May 12, 2008, from <http://www.edexcellence.net/doc/041107AutonomyGap.pdf>
- Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.
- Alexander, K., Entwisle, D., & Dauber, S. (1996). Children in motion: School transfers and elementary school performance. *Journal of Educational Research*, 90(1), 3-12.
- Alliance for Education. (2004). *Tapping the potential: Retaining and developing high quality new teachers*. Washington, DC: Author.
- Alliance for Excellent Education (2008a). *About the crisis*. Retrieved February 4, 2008, from http://www.all4ed.org/about_the_crisis
- Alliance for Excellent Education (2008b). *Impact*. Retrieved February 4, 2008, from http://www.all4ed.org/about_the_crisis/impact
- Alliance for Excellent Education (2008c). *Students*. Retrieved February 4, 2008, from http://www.all4ed.org/about_the_crisis/students
- Alliance for Excellent Education (2008d). *Measuring and improving the effectiveness of high school teachers*. Retrieved May 18, 2008, from <http://www.all4ed.org/files/TeacherEffectiveness.pdf>
- Antoline, C.M. (2000). A validation study of the Teacher Perceiver Interview. *Dissertation Abstracts International*, 61(09), 5035. (UMI No. 9988385).
- Associated Press. (2008). Schools face sanctions under landmark law. *MSNBC*. Retrieved May 13, 2008, from <http://www.msnbc.msn.com/id/24556075/>
- Babu, S., & Mendro, R. (2003, April). *Teacher accountability: HLM-based teacher effectiveness indices in the investigation of teacher effects on student achievement in a state assessment program*. Paper presented at the American Educational Research Association meeting, Chicago.

- Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, 111(1), 97-133.
- Barnes, G., Crowe, E., & Schaefer, B. (2007). *The cost of teacher turnover in five school districts*. Washington, DC: National Commission on Teaching and America's Future.
- Barron, J.B. (2007). *Improving Hispanic parental involvement in school*. Texas Study of Secondary Education Fall 2007, Vol. XVII, No. 1, TASSP.
- Barton, P.E. (2003). *Parsing the achievement gap*. Educational Testing Service. Retrieved May 18, 2008, from <http://www.ets.org/Media/Research/pdf/PICPARSING.pdf>
- Barton, P.E. (2007). *The right way to measure growth*. Educational Leadership December 2007/January 2008, (Vol. 65 No. 4), Association for Supervision and Curriculum Development.
- Barton, P.E., & Coley, R.J. (2007). *The family: America's smallest school*. Educational Testing Service. Retrieved May 12, 2008, from http://www.ets.org/Media/Education_Topics/pdf/5678_PERCReport_School.pdf
- Bedsworth, W., Colby, S., & Doctor, J. (2006). *Reclaiming the American dream*. Boston, MA: Bridgespan.
- Bennerm, A., & Mistry, R. (2007). Congruence of mother and teacher educational expectations and low-income youth's performance competence. *Journal of Educational Psychology*, 99(1).
- Bossidy, L., & Charan, R. (2002). *Execution: The discipline of getting things done*. New York, NY: Crown Business.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2007). *Who leaves? Teacher attrition and student achievement*. September 30, 2007.
- Calkins, A., Guenther, W., Belfiore, G., & Lash, D. (2007). *The turnaround challenge*. Boston, MA: Mass Insight Education and Research Institute. Retrieved May 18, 2008, from http://www.massinsight.org/resourcefiles/TheTurnaroundChallenge_2007.pdf
- Chesley, D. (2004). The validity of teacher employment success predictors. *Dissertation Abstracts International*, 65(03), 894. (UMI No. 3126191).

- Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). *Teacher credentials and student achievement in high school: A cross subject analysis with student fixed effects*. Working paper series No. 13617, Stanford: National Bureau of Economic Research.
- Coleman, J.S. (1966). *Equality of Educational Opportunity*. Washington, D C: Government Printing Office.
- Conley, D.T. (2005). *College knowledge: What it really takes for students to succeed and what we can do to get them ready*. San Francisco: Jossey: Bass.
- Darling-Hammond, L. (2000). *Teacher quality and student achievement: A review of state policy evidence*. Educational Policy Analysis Archives 8, no. 1. Retrieved April 28, 2008, from <http://epaa.asu.edu/epaa/v8n1/>
- Darling-Hammond, L. (1995). *The role of teacher expertise and experience in students' opportunity to learn*. In P. Brown (Ed.), *Strategies for linking school finance and students' opportunity to learn* (pp. 19-23). Washington, DC: Brookings Institution.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass.
- Davis-Frost, Diane (2000). *Intentions and Implementation of the Professional Development and Appraisal System in Texas*. Retrieved February 22, 2009 from <https://libproxy.library.unt.edu:2576/login?url=http://libproxy.library.unt.edu:2587/pqdweb?did=726305651&sid=2&Fmt=2&clientId=87&RQT=309&VName=PQD>
- Dee, T. S., & Cohodes, S. (2008). Out-of-Field Teaching and Student Achievement: Evidence from 'Matched-Pairs' Comparisons. *Public Finance Review* 36(1): 7-32.
- Desimone, L. (1999). Linking parent involvement with student achievement: do race and income matter? *The Journal of Educational Research*, 93(1).
- Dillon, S. (2008). States' data obscure how few finish high school. *The New York Times*, Thursday, March 20, 2008.
- Duncan, G., & Magnuson, K. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The future of children*, 15(1).
- Finn, C.E. (2002). *Real accountability in K-12 education*. Ravitch, D. Hoover Press: School Accountability.
- Fischer, K. (2008). AG: DISD can keep bonus data private. *The Dallas Morning News*, Friday, January 11, 2008.

- Gallup (2008a). *TeacherInsight™*. Retrieved November 2, 2008, from:
<http://www.gallup.com/consulting/education/22093/TeacherInsight.aspx>
- Gallup (2008b). *Selection*. Retrieved November 2, 2008, from:
<http://www.gallup.com/consulting/education/22267/Selection.aspx>
- Gephart, D. (2009). The route to achievement leads through the human resources door. *The School Administrator*, 66(1).
- Gill, B., McCombs, J.S., Naftel, S., Ross, K., Song, M., Harmon, J., & Vernez, G. (2008). *State and local implementation of the No Child Left Behind Act*, Volume IV – Title I School choice and supplemental educational services: Interim report. Retrieved May 18, 2008, from <http://www.ed.gov/rschstat/eval/choice/nclb-choice-ses/nclb-choice-ses.pdf>
- Gitomer, D.H. (2007). *Teacher quality in a changing policy landscape: Improvements in the teacher pool*. Princeton, NJ: Educational Testing Service. Retrieved May 13, 2008, from
http://www.ets.org/Media/Education_Topics/pdf/TQ_full_report.pdf
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2, 50-55. Retrieved on-line May 3, 2008, from
<http://www.hoover.org/publications/ednext/3368021.html>
- Gordon, G. (2003). *Alternative Certification and TeacherInsight*. Retrieved November 2, 2008, from
<http://media.gallup.com/EDUCATION/pdf/AlternativeCertificationAndTeacherInsight20040722.pdf>
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, D C: Brookings Institution.
- Guarino, C., Santibanez, L., Daley, G., & Brewer, D. (2004). *A Review of the Research Material on Teacher Recruitment and Retention*. Santa Monica, CA: Rand.
- Haberman Foundation (2008a). *Dimensions assessed*. The Haberman Educational Foundation, Inc. Retrieved May 4, 2008, from
<http://www.altcert.org/teacher/dimensions.asp>
- Haberman Foundation (2008b). *Need good teachers? Grow your own!* Retrieved April 28, 2008, from http://www.habermanfoundation.org/images/foundation_final.gif
- Haberman Foundation (2008c). *The science of teacher selection and interviewing*. The Haberman Educational Foundation, Inc. Retrieved May 4, 2008, from
<http://www.habermanfoundation.org/prescreeners.asp?test=teacher&page=On-Line%20Prescreeners>

- Hammond, B. (2008). Students will get high school diplomas only if they pass state tests, says board. *Oregon Live*, April 19, 2008.
- Hanushek, E.A. (1992). The trade off between child quality and quality. *Journal of Political Economy*, 100:84-117.
- Hanushek, E.A., & Pace, R.R. (1995). Who chooses to teach (and why)? *Economics of Education Review*, 14(2), 101-117.
- Hanushek, E.A., & Raymond, M.E. (2002). *Sorting out accountability systems*. Hoover Press: School Accountability.
- Hanushek, E.A., Rivkin, S.G., & Kain, J.J. (2005). Teachers, schools, and academic achievement. *Econometrica* 73(2), 417-58.
- Henke, R.R., Chen, X., Geis, S., & Knepper, P. (2000). *Progress through the teacher pipeline: 1992-93 college graduates and elementary/secondary teaching as of 1997*. Washington, DC: National Center for Education Statistics.
- Henson, R.K. (2006). Effect size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, 34, 601-629.
- Hodgkinson, H. (2000). *Secondary schools in a new millennium: Demographic certainties, social realities*. Reston, VA: National Association of Secondary School Principals.
- Hoxby, C.M. (2002). *The cost of accountability*. Hoover Press: School Accountability.
- Howley, C., & Bickel, R. (2000). *When it comes to schooling... small works: school size, poverty, and student achievement*.
- Hussar, W.J. (1999). *Predicting the need for newly hired teachers in the United States to 2008-09*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Ingersoll, R. (2003). *Is there really a teacher shortage? Center for the Study of Teaching and Policy*. University of Washington: Research report.
- Jacob, B.A., & Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education (Faculty Research Working Papers Series). Boston: Harvard University, John F. Kennedy School of Government. Retrieved May 18, 2008, from [http://www.ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP05-040/\\$File/rwp_05_040_jacob.pdf](http://www.ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP05-040/$File/rwp_05_040_jacob.pdf)

- Johnson, D.D., Rice, M.P., Sullivan, S.S., Henderson, D.L., & Beard, A. (2008). *Texas teachers, moonlighting and morale*. Texas State Teachers Association, April 2008.
- Jones, M.G., Jones, B., & Hargrove, T. (2003). The unintended consequences of high-stakes testing. Lanham, MD: Rowman and Littlefield.
- Jordan, D., & DeNardo, C. (2008). ACLU sues Palm Beach County school district over poor graduation rates. *PalmBeachPost.com*. Retrieved March 19, 2008, from http://www.palmbeachpost.com/localnews/content/local_news/epaper/2008/03/19/s1a_aclu_lawsuit_0319.html
- Kain, J.F., & Singleton, K. (1996, May/June). Equality of educational opportunity revisited. *New England Economic Review*, 87-114.
- Kane, T.J., Rockoff, J.E., & Staiger, D.O. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City. NBER Working Paper Number 12155, April 2006.
- Kanipe, M.L.S. (1996). A study of the predictive validity of the Urban Teacher Perceiver Interview in the Know County School System. *Dissertation Abstracts International*, 58(06), 2004. (UMI No. 9735805).
- Keller, B. (2003). Hiring headway in quality counts 2003: If I can't learn from you... ensuring a highly qualified teacher in every classroom. *Education Week*, January 12, 2003, 43-44.
- Kerbow, D. (1996). *Patterns of urban student mobility and local school reform* (Tech Rep. No. 5), Chicago: University of Chicago.
- Koerner, R. J. (2007). *The relationship between the TeacherInsight™ interview scores and student performance as measured by the Texas Growth Index*. Retrieved November 2, 2008 from <http://libproxy.library.unt.edu:2587/pgdweb?index=0&sid=5&srchmode=2&vinst=PROD&fmt=6&startpage=-1&clientid=87&vname=PQD&RQT=309&did=1436374121&scaling=FULL&ts=1235361064&vtype=PQD&rqt=309&TS=1235361071&clientId=87>
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* 114(2): 497-532.
- Langford, H., Loeb, S., & Wyckoff, J. (2002). Teaching sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37-62.

- Levin, J., & Quinn, M. (2003). *Missed opportunities: How we keep high-quality teachers out of urban classrooms*. The New Teacher Project. Retrieved May 10, 2008, from <http://www.tntp.org/files/MissedOpportunities.pdf>
- Liu, E. (2002). *New Teachers' Experiences of Hiring in New Jersey*. Harvard Graduate School of Education Project on the Next Generation of Teachers. Retrieved October 21, 2008 from http://www.gse.harvard.edu/~ngt/Liu_AERA2002.pdf
- Liu, E., & Johnson, S.M. (2006). New Teachers' Experiences of Hiring: Late, Rushed and Information-Poor, *Educational Administration Quarterly*, 42(3), 324-360.
- Longo, R.G. (2003). *Applying credibility to teacher hiring*. The School Administrator Web Edition. Retrieved May 10, 2008, from <http://www.aasa.org/publications/content.cfm?ItemNumber=4021>
- Marzano, R.J. (2003). *What Works In Schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mellott, J. (2008). Bills to pull Virginia out of No Child Left Behind move through committees. *Daily News Record Online*. Retrieved February 6, 2008, from http://www.dailynews-record.com/news_details.php?AID=14781&CHID=2
- Metzger, S.A., & Wu, M. (2006). *Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values*. Teacher Qualifications and The Quality of Teaching. Retrieved May 18, 2008, from <http://www.msu.edu/user/mkennedy/TQQT/Reports/MetzgerWu07-RER.doc>
- Miller, R., Murnane, R.J., & Willett, J.B. (2007). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. NBER Working Paper No. W13356.
- Morrison, D. (2008). A failing grade for exit exams. *UMN News*. May 16, 2008. Retrieved May 18, 2008, from http://www1.umn.edu/umnnews/Feature_Stories2/A_failing_grade_for_exit_exams.html
- National Center for Education Statistics. (2004). *Conditions of education 2004*. Washington, DC: U.S. Department of Education.
- Nelson, P., Simoni, J., & Adelman, H. (1996). Mobility and school functioning in the early grades. *Journal of Educational Research*, 89(6), 365-369.
- Nichols, S.L., & Berliner, D.C. (2007). *Collateral damage: How high stakes testing corrupts America's schools*. Cambridge, MA: Harvard Educational Press.
- Nye, B. Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis* 26(3): 237-57.

- Orfield, G., & Kornhaber, M.L. (Eds.) (2001). *Raising standards or barriers? Inequality and high stakes testing in public education*. New York: Century Foundation Press.
- Orfield, G., Losen, D., Wald, J., & Swanson, C.B. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis*. Cambridge, MA: The Civil Rights Project at Harvard University. Retrieved April 28, 2008, from http://www.urban.org/uploadedPDF/410936_LosingOurFuture.pdf
- Podgursky, M.J., & Springer, M.G. (2006, November). Teacher performance pay: A review [Working paper 2006–01]. National Center on Performance Incentives, Peabody College, Vanderbilt University. Retrieved February 25, 2008, from http://www.performanceincentives.org/data/files/news/PapersNews/Podgursky_and_Springer_2006_Revised.pdf
- Presley, J. B., White, B. R., & Gong, Y. (2005). *Examining the distribution and impact of teacher quality in Illinois*. Edwardsville: Illinois Education Research Council.
- Public Law 107-110 (2001). No Child Left Behind Act of 2001.
- Radcliffe, J. (2008). Educators want TAKS to count progress. *Houston Chronicle*, May 13, 2008. Retrieved May 13, 2008, from <http://www.chron.com/disp/story.mpl/metropolitan/5774485.html>
- Ravitch, D. (2002). *Testing and accountability, historically considered*. Hoover Press: School Accountability.
- Ravitch, D. (1983). *The Troubled Crusade: American Education, 1945-1980*, New York: Basic Books, 1983, 168.
- Redburn, M.K. (1998). A validation study of the SRI/Gallup Principal Perceiver Interview using the Principal Rating Scale. *Dissertation Abstracts International*, 60(05), 1414. (UMI No. 9931940).
- Redondo, B., Aung, K.M., Fung, M., & Yu, N.W. (2008). *Left in the margins: Asian American students & the No Child Left Behind Act*. New York, NY: Asian American Legal Defense and Education Fund. Retrieved May 18, 2008, from http://www.aaldef.org/docs/AALDEF_LeftintheMargins_NCLB.pdf
- Region XIII Education Service Center. (2008). *Professional Development and Appraisal System*. Retrieved May 10, 2008, from <http://www5.esc13.net/pdas/>
- Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2): 417-58.

- Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94(2): 247-252.
- Roukema, L. (2004). Keeping new teachers: The revolving door of the teaching profession. *Today's School*. August 2004.
- Sampson, R.J., Sharkey, P., Raudenbush, S.W. (2007). *Durable effects of concentrated disadvantage on verbal ability among African-American children*. Proceedings of the National Academy of Sciences of the United States of America. Retrieved May 12, 2008, from <http://www.pnas.org/cgi/reprint/105/3/845?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=Raudenbush&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT>
- Sanders, W., & Rivers, J. (1996). *The cumulative and residual effects of teachers on future academic achievement*. Memphis: University of Tennessee Value-added Research and Assessment Center.
- Sanford, B. (2003). Improving student performance, one teacher at a time. Retrieved November 2, 2008, from <http://media.gallup.com/EDUCATION/pdf/ImprovingStudentPerformanceOneTeacherAtATime20040422.pdf>
- Sawyer, G.B. (2005). A study using the Star Teacher Selection Interview to predict the successful performance of teachers in South Carolina's program of alternative certification for educators. *Dissertation Abstracts International*, 66(04), 1235. (UMI No. 3173188).
- Schmidt, F. (2004). 10,000 managers can't be wrong. *Gallup Management Journal*. Retrieved May 3, 2008, from <http://gmj.gallup.com/content/10882/10000-managers-cant-wrong.aspx>
- Scholastic Administr@tor (2008). *Fall into the achievement gap*. Scholastic Administr@tor, March/April 2008, 12.
- Shields, P.M., Esch, C.E., Humphrey, D.C., Young, V.M., Gaston, M., & Hunt, H. (1999). *The status of the teaching profession: Research findings and policy recommendations*. Santa Cruz, CA: Center for the Future of Teaching and Learning.
- Springston, S.M. (2006). A study of the reliability and concurrent validity of the ICS-Urban Teacher Interview Instrument. *Dissertation Abstracts International*, 67(02). (UMI No. 3207562).
- Streich, J. (2008). 'No Child Left Behind' still gets mixed grades. *Abc13.com*. Retrieved March 22, 2008, from <http://ww2.abc13.com/Global/story.asp?S=8033297>

- Stronge, J.H., & Tucker, P.D. (2003). *Handbook on Teacher Evaluation: Assessing and Improving Performance*. Larchmont, NY: Eye on Education.
- Stutz, T. (2006). School reform in state's hands. *Dallas Morning News*, July 10, 2006.
- Texas Education Agency (2008a). *Overview of the Academic Excellence Indicator System*. Retrieved December 4, 2008, from <http://www.tea.state.tx.us/perfreport.aeis.html>
- Texas Education Agency (2008b). *Timeline of testing in Texas*. Retrieved December 2, 2008 from <http://www.tea.state.tx.us/student.assessment/resources/studies/testingtimeline.pdf>
- Texas Education Agency (2005). *Professional Development and Appraisal System Teacher Manual*. Retrieved May 10, 2008, from <http://www5.esc13.net/pdas/docs/PDASTeacherManual.pdf>
- Tomsho, R. (2008). No child left behind lacks bite. *The Wall Street Journal Online*. May 13, 2008. Retrieved May 18, 2008, from <http://online.wsj.com/article/SB121063138674386483.html>
- United States Government Accountability Office (2007). *No Child Left Behind Act. GAO-07-1035*. Retrieved May 18, 2008, from <http://www.gao.gov/new.items/d071035.pdf>
- Unmuth, K.L. (2007). State's exemplary schools not judged on all criteria. *Dallas Morning News*, September 9, 2007.
- U.S. Census Bureau (2008) Retrieved February 4, 2008, from <http://www.census.gov/hhes/www/income/histinc/p16.html>
- Victoria, C.G., Adair, L., Fall, C., Hallal, P., Martorell, R., Richter, L., & Sachdev, H.S. (2008). *The Lancet*, Vol. 371, January 26, 2008.
- Wallis, C. (2008). How to make great teachers. *Time*, February 13, 2008.
- Wilson, S.M., Floden, R., Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Research report for the U.S. Department of Education. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Wright, B. (2006). *Can I get a teacher?* Fiscal Notes, January 2006.