EXPLORING THE RELATIONSHIP BETWEEN THE TEACHERINSIGHT SCORE
AND THE TEACHER GROWTH INDEX

DISSERTATION
SUBMITTED TO
THE DWIGHT SCHAR COLLEGE OF EDUCATION
ASHLAND UNIVERSITY

In Partial Fulfillment of the Requirements for
The Degree
Doctor of Education in Leadership Studies
Thomas Fry Jr., B.A. Ed., M.Ed.

ASHLAND UNIVERSITY
ASHLAND, OH
2013
A Dissertation

titled

Exploring the Relationship between the TeacherInsight Score

and the Teacher Growth Index

by

Thomas Fry Jr.

In Partial Fulfillment of the Requirements for

The Degree

Doctor of Education in Leadership Studies

Carla Edlefson, Ph.D., Committee Chair

Date

James Olive, Ph.D., Committee Member

Date

Paul Milton, Ph.D., Committee Member

Date

Judy A. Alston, Ph.D., Chair, Leadership Studies

Date

James Van Keuren, Ph.D., Dean, Dwight Schar College of Education

Date

John Moser, Ph.D., Interim Director, Graduate School

Date
In an attempt to ensure the hiring of the most effective teachers from the candidate pool, many school districts rely on teacher selection instruments such as the TeacherInsight™ (TI) from The Gallup Organization. Teacher-level value-added data are now available to measure the effectiveness of teachers. The purpose of this research study was to determine how well teacher-screening instruments predict the highest quality teachers from the candidate pool. The study used a sample of teachers for whom TI scores as well as a value-added measure, the teacher growth index (TGI), had been recorded. The analyses found a very weak positive relationship between the TI score and TGI score for English/reading teachers. There was no relationship between the TI score and TGI score for mathematics teachers. Until there is additional research into the effectiveness of the TI score to predict effective TGI scores, school systems will need to be cautious with using only the TI score to screen applicants.
DEDICATION

I dedicate this dissertation to my wife and family. Without their understanding, support, and love I would not have even taken on this incredible endeavor. Cassy, you mean everything to me. I appreciate your willingness to read and reread numerous iterations of chapters. Your patience with my frustrations and time away from the family can never be overstated. My son Hayden, your love of reading provided me with motivation to continually ‘read-on’. My daughter Taryn, watching your determination in gymnastics was inspirational and kept me moving forward. Dad, your work ethic is incredible and I aspire to match it. Mom, your caring heart knows no bounds. Seth, you give of yourself without expectation of anything in return. Beth, your sense of humor always brings a smile to my face. Tracy, you have been supporting me since birth. I strive daily to make all of you proud.
ACKNOWLEDGMENTS

It seems like such a long time ago that I began this journey. I would not have completed it without countless friends and colleagues who provided support, encouragement, and even the occasional fire. There are no words to express my gratitude to all of you.

Dr. Jenny Hooie, you were always there to listen and prod me forward whenever I hit an obstacle. Dr. Ian Wei, you have been a great friend and resource throughout the entire process. Anytime I needed to bounce an idea or meet to talk through a struggle you were there.

I would like to thank my dissertation committee, Dr. Carla Edlefson, Dr. James Olive, and Dr. Paul Milton. I appreciate your guidance through the entire process.

Thanks to all of my friends and cohort mates at Ashland University, especially the Columbus car-pooling group. You were all willing to assist in removing obstacles or just listen to me vent. To all others who helped along this journey I thank you.
Table of Contents

CHAPTER I. Introduction ...................................................................................................1
Conceptual Framework ...............................................................................................2
Identification of the Problem .....................................................................................5
Overview of Methodology .........................................................................................6
Significance of the Study ............................................................................................7
Limitations ..................................................................................................................8
Definitions ..................................................................................................................9
Summary .....................................................................................................................12
Organization of the Dissertation .................................................................................13

CHAPTER II. Review of Related Research ................................................................15
Brief History of Value-Added .....................................................................................16
Various Methods of Identifying Teacher Effectiveness ...........................................21
Defining Teacher Quality ..........................................................................................24
Ohio Teacher Evaluation System .............................................................................26
Brief History of Teacher Interview Process .............................................................31
Teacher Perceiver Interview ....................................................................................32
Caveats of Value-Added ............................................................................................36
Attribution Error .......................................................................................................37
CHAPTER III.

Research Methodology .................................................................52

Purpose ....................................................................................52

Instrumentation and Variables .................................................53

Teacher Growth Index ...............................................................55

Research Design ........................................................................58

Procedures ................................................................................59

Research Questions and Related Hypotheses ..............................60

Participants ...............................................................................61

Sample Size ..............................................................................62

Research Approval ....................................................................63

Analysis of Data ........................................................................64

Summary ..................................................................................66
CHAPTER IV.

Results of Data Analysis........................................................................................................68
Research Questions ..................................................................................................................68
Descriptive Statistics .............................................................................................................69
  Independent Variables ........................................................................................................70
  Dependent Variable ...........................................................................................................73
Data Analysis ..........................................................................................................................74
  Preliminary Analysis .........................................................................................................74
  Question 1.1 .......................................................................................................................77
  Question 1.2 .......................................................................................................................79
  Question 2 ..........................................................................................................................80
Summary of Findings .............................................................................................................81

CHAPTER V.

Discussion ................................................................................................................................83
  TI Scores as a Predictor of Teacher Effectiveness .........................................................86
  Summary of Results for Question 1.1 .................................................................86
  Summary of Results for Question 1.2 .................................................................88
  Summary of Results for Question 2 .................................................................91
Recommendations and Implications ...............................................................................92
  Limitations .......................................................................................................................93
  Delimitations ..................................................................................................................94
List of Tables

Table 3.1 Interview Status Based on TI Level ................................................................. 54, 90
Table 3.2 Teacher Effectiveness Ratings ...................................................................... 56
Table 3.3 Variables of Study ....................................................................................... 57
Table 3.4 Assessments for TCAP Analysis .................................................................. 57
Table 4.1 Frequency: Number of Teachers by Gender .................................................. 71
Table 4.2 Descriptive Statistics for English/reading Teachers ........................................ 72
Table 4.3 Descriptive Statistics for Mathematics Teachers ............................................. 73
Table 4.4 Descriptive Statistics for TGI Score of English/reading and Math Teachers .... 73
Table 4.5 Pearson Correlation of Continuous Variables for English/reading Teachers .... 75
Table 4.6 Pearson Correlation of Continuous Variables for Mathematics Teachers ...... 75
Table 4.7 Independent Samples T-Test for English/reading Teachers ............................. 77
Table 4.8 Independent Samples T-Test for Mathematics Teachers ................................. 77
Table 4.9 Regression Results for English/reading Teachers ........................................... 79
Table 4.10 Regression Results for Mathematics Teachers .............................................. 80
Table 4.11 Block 2 Summary for English/reading and Mathematics Teachers .......... 81
List of Figures

Figure 1.1 Conceptual framework .................................................................................. 3
Figure 2.1 OTES evaluation framework ........................................................................ 28
Figure 2.2 OTES evaluation matrix .............................................................................. 29
Figure 3.1 Study district application process ................................................................ 53, 85
CHAPTER I

Introduction

Teachers are the most important factor in student achievement (The Gallup Organization, 2011; Goe, Bell, & Little, 2008; Stronge, Ward, & Grant, 2011; Wright, Horn, & Sanders, 1997). The problem is that many of the factors currently used to identify the most effective candidates for a teaching position do not predict accurately those who will be the most effective teachers. These characteristics include teaching credentials, both advanced degrees and certifications, years of teaching experience, passing scores on content knowledge exams, and formal observations. Each characteristic has more or less weight in the hiring process, depending on the school district.

As a result, some school districts incorporate the use of an online teacher screening instrument called the TeacherInsight (TI). This instrument, created by The Gallup Organization, helps district administrators identify teaching candidates that have the strengths of effective teachers (The Gallup Organization, 2007). The use of a teacher-screening instrument does provide some consistency in the hiring process, as every applicant receives the same screening questions. Since the creation of the TI, additional measures of teacher effectiveness have been developed in Ohio and across the nation, including Value-Added data. School districts continue to purchase the TI screening instrument from the Gallup Organization at a time when budgets are extremely thin. The use of commercial teacher-screening instruments has increased across the country, but there is little evidence of how accurate these instruments are at predicting teacher effectiveness (Jacob, Kane, Rockoff, & Staiger, 2009, p. 13) If school districts are going to dedicate a portion of their limited financial resources to teacher screening instruments
such as the TI, they will want some assurances that it accurately predicts future teacher effectiveness as measured by student academic progress.

**Conceptual Framework**

In order to provide a foundation for understanding the empirical findings of this study, it is necessary to consider both how school districts currently make selection decisions for hiring teachers and how teacher effectiveness is defined by Teachers Connecting Achievement and Progress (TCAP) reporting. The TCAP reports contain classroom level value-added data for individual teachers. I will describe the use of the TI scores and how district administrators in the study school district interpret these scores for the purpose of teacher selection. Additionally, I will explain the formal model of using teacher-level value-added data for identifying effective teachers in English/reading and mathematics.

The conceptual framework of the study is shown in Figure 1.1, which illustrates the hypothesis that TI scores are related to a teacher growth index (TGI) scores for English/reading and mathematics. TI scores are reported to identify the highest quality teaching candidates from the applicant pool (The Gallup Organization, 2007). As such, it is hypothesized that teacher quality is positively related to teacher effectiveness in English/reading and mathematics (Hypothesis 1.1). Teacher demographics (gender and years of experience) are considered covariates for the study and are hypothesized to have no influence on teacher effectiveness (Hypothesis 1.2). As TI scores increase it is hypothesized that TGI score will also increase. For the purpose of this study, TGI scores in mathematics and reading will only be collected for grades 3 through 8 and high school
courses of English 9 through 12 and Algebra I, Algebra II, Geometry, and Pre-Calculus. Although mathematics and English/reading are both subject areas, they are not considered to be the same. Mathematics is considered a more concrete subject area, while reading is considered more of a process. Due to this difference in subjects, it is hypothesized that the TI score will predict effective mathematics teachers better than English/reading teachers (Hypothesis 2).

Figure 1.1 Conceptual Framework

In Ohio, school districts utilize the Educational Value-Added Assessment System (EVAAS) developed by the SAS® organization. This system uses estimated means to provide an accurate measure of a teacher’s effectiveness. The EVAAS® methodology utilizes up to five years of standardized test scores for individual students and analyzes all subjects at the same time to improve the precision of the effectiveness estimates (SAS®, 2012). Unlike the Value-added models (VAM) in Colorado, Massachusetts, and Indiana,
EVAAS® does not require vertically scaled assessments (Buzick & Laitusis, 2010; SAS®, 2012). This methodology is sensitive enough to measure growth of both gifted learners and students with disabilities (Battelle for Kids [BFK], 2011). SAS® identifies the following three criteria necessary for this sensitivity:

1. The end-of-year assessment must have enough academic stretch to differentiate the achievement of high- and low-achieving students.

2. The test must have appropriate levels of reliability.

3. The tests must be highly correlated with the curriculum taught by the teachers (BFK, 2011, p. 5).

It is important to note that almost all commercial and state-mandated standardized tests meet these criteria (Wright, White, Sanders, & Rivers, 2010). The EVAAS® system calculates and reports standard error to provide a confidence band for the teaching effectiveness estimate, therefore providing some context for interpreting the results. The methodology provides estimates of teacher effectiveness, but does not suggest any particular instructional methods for improving student growth. SAS® purports that students perform well on standardized tests as a result of good teaching based on sound instructional decisions promoting scholarship daily (SAS®, 2012). There is a variety of instructional methods a teacher can use to facilitate effective academic growth.

In a press release, Secretary of Education Spellings (2007) announced the approval of the EVAAS® value-added model for Ohio. The only caveat for the state was to adopt a minimum group size for all student subgroups. The student subgroups for Ohio include the seven major ethnicities, students with disabilities, gifted, limited English proficient, and economically disadvantaged students. Ohio adopted a minimum group
size of 30 students for accountability purposes and the equivalent of six full-time students for calculation of a teacher-level value-added report. There was a rigorous peer-review process, including academia, private organizations, and state and local education agencies, to ensure a fair and transparent selection of each growth model the U.S. Department of Education approved. Ohio’s model was one of ten high-quality growth models used in the pilot (Spellings, 2007). The Department of Education’s approval lends credibility to the use of the EVAAS® model for estimating teacher effectiveness.

Identification of the Problem

Building administrators are responsible for hiring teachers to fill vacancies. It is important for school administrators to be confident in their identification of the most effective teachers from the applicant pool (Cicotte, 2009; Novotny, 2009). How can a school district administrator be certain that the candidate will be an effective English/reading or mathematics teacher in the classroom? There is currently little independent research to support the relationship between TI scores and teacher effectiveness in English/reading and mathematics. This study furthers the knowledge base relating to teacher-level, value-added data and teacher screening instruments that purport to identify quality candidates. In Ohio, value-added information is provided to school districts in TCAP reports.

Three research questions created in the study will review the relationship between TI scores and teacher effectiveness. Recognizing the challenges of school districts to hire the most effective teachers possible from the candidate pool, the research questions and hypotheses are:
1.1: Is there a relationship between the TI scores and teacher effectiveness – as a proxy of student progress – as measured by the teacher growth index (TGI) score for English/reading, controlling for teacher demographics (years of experience and gender)?

1.1H₁: Teachers with a higher level of TI will have a higher TGI score in English/reading.

1.2: Is there a relationship between the TI score and teacher effectiveness as measured by the TGI score for mathematics controlling for teacher demographics (years of experience and gender)?

1.2H₂: Teachers with a higher level of TI will have a higher TGI score in mathematics.

2. Does the TI score predict effective English/reading teachers better than effective mathematics teachers as measured by the TGI score?

2H₃: The TI score will predict effective mathematics teachers better than it will predict effective English/reading teachers.

**Overview of Methodology**

This quantitative study used correlational research as its main research design to explore the relationship between the variables and the predictive value of TI scores. A Pearson-Product Correlation determined the relationship between TI scores and TGI scores for English/reading and mathematics teachers. Hierarchical multiple regression (HMR) was used to examine the predictive value of TI scores in relation to TGI scores of English/reading and mathematics teachers. HMR allows all variables, including the covariates (years of experience and gender), to be tested individually. This permits the
researcher to control for and explore the contribution of each variable to the predictive value of the model. There are several teacher quality estimates used in the study. TI scores are from a teacher-screening instrument, supposed to identify high-quality teaching candidates. TGI scores are estimates of teacher effectiveness in English/reading and mathematics.

**Significance of the Study**

The purpose of this research study was to determine how well teacher-screening instruments identify the highest quality teachers from the candidate pool. For the purpose of this study, effective teachers were identified by their ability to facilitate at least a years’ worth of student academic growth in English/reading and mathematics. A years’ worth of academic growth is calculated through a teacher growth index (TGI) score provided by the EVAAS®. The study investigated the relationship between a specific teacher-screening instrument, the TI, and eventual teacher effectiveness ratings for English/reading and mathematics. Teacher-screening instruments such as the TI are used by school districts nationwide to identify the strongest candidates for their hiring process. The study district expected that candidates who score higher on the TI will eventually be more effective English/reading and mathematics teachers in the classroom. Due to this, candidates who score well on the TI move into the interview process. The TI is intended to improve the efficiency of a school districts human resource department. Teacher-screening instruments have a significant financial cost, but can reduce time and effort of human resource departments if a positive relationship with teacher effectiveness exists.
The information from this study can inform school districts of the effectiveness of using the TI screening instrument in the hiring process.

In Ohio, the teacher pension system will experience a significant change in cost of living adjustments effective July 1, 2013 and an increase in active teachers’ contributions from ten percent to fourteen percent by July 1, 2016 (State Teacher Retirement System, 2012). Beginning February 1, 2013 educators wanting to purchase past leave of absence credit will pay for interest on both member and employer payments, which represents a significant change from the current practice. Additionally, districts will not be able to create a retirement incentive plan to entice early retirements of teachers after July 1, 2014 (State Teacher Retirement System, 2012). These significant changes will likely lead to an increase in teacher retirements over the next few years. If this occurs, school districts will need to replace these veteran teachers leaving the profession. Every school district will want to ensure they interview high-quality candidates that will likely become effective teachers.

**Limitations**

There were several limitations to consider with the study. First, the analysis only includes one large affluent, suburban school district in central Ohio. It would be difficult to generalize the results of this study to urban or rural school districts because the data are only from an affluent suburban school district. There are not many school districts in Ohio that both use the TI as a teacher-screening instrument and have TGI scores. Next, the study only uses data from teachers who taught English/reading and mathematics courses in grades three through eight and certain high school courses. The reason for this
decision is there are many more teachers of English/reading and mathematics than science or social studies having the necessary data for analysis. Because the analysis only includes English/reading and mathematics teachers, it will not be generalizable to other subjects or to grade levels outside of the analysis, or for school districts outside of Ohio. Finally, the data set only includes TI and TGI scores from the 2007 – 2008 school year to the 2011 – 2012 school year. Although the study school district has TI scores from the 2003 – 2004 school year they only began collecting TGI scores for English/reading and mathematics in the 2007 – 2008 school year as a part of Project SOAR with the Battelle for Kids organization.

The study school district uses TI scores to identify high-quality teaching candidates for the interview stage in the hiring process. It is unlikely that many teachers who score low on the TI will obtain an interview and thereby eventually attain a teaching position in the school district. Teacher effectiveness data can only be collected for teachers who are hired for a position within the school district. The study also only investigates the TI, which is one of several teacher-screening instruments available for school districts to utilize in their hiring process. Any relationship that exists between the TI scores and teacher growth indices does not guarantee a relationship with other teacher-screening instruments. This provides a limit to the range of scores within the study.

**Definitions**

Shared understanding of terminology is important for any organization to ensure effective change (Dufour & Marzano, 2011). Although VAM have been a part of the education system in Ohio for the past five years, it is still not well understood by many
educators. There are several key terms used in this study that may have alternative definitions. To alleviate confusion and better ensure a shared understanding, a list of those terms and their definitions for the purpose of this study are listed below.

*Attribution error:* Error associated with crediting the change in student academic performance (either positive or negative) to an inappropriate teacher.

*Criterion-referenced assessment:* An assessment with questions aligned to specific subject matter or criteria. The objective of this type of assessment is to determine how much of this subject matter the student has learned or mastered. The Ohio Achievement Assessment (OAA), Ohio Graduation Test (OGT), and ACT Quality Core End of Course Exams (EOC) are examples of a criterion-referenced assessment used in Ohio.

*Educational Value-Added Assessment System (EVAAS):* An assessment system created by SAS®. EVAAS® provides school districts with various value-added reporting utilizing a multivariate mixed model. EVAAS® is utilized by several states including Ohio, Tennessee, and North Carolina.

*Measurement error:* A form of error that skews testing results from actual performance. There are two main types of measurement error, random and systemic. These errors can skew scores positively as well as negatively.

*Multivariate Response Model:* a multivariate, longitudinal, linear mixed model used for calculating value-added reports in Ohio.

*Norm-referenced assessments:* An assessment that gives the relative position of a student’s performance as compared to that of a given population (e.g. national, state, or local). Although it provides context relating to the students’ position within the specific
population it does not identify if the student has mastery of specific subject level content.

The TerraNova is an example of a norm-referenced assessment.

*Random error:* A form of measurement error that is caused by unknown or unpredictable factors and skews the achievement scores of an individual.

*Standard error:* Allows the establishment of a confidence band around the mean score of a group. It provides a measure of uncertainty.

*Standardized Gain Model* (SGM): A value-added model that uses the difference between a student’s current and previous standardized test scale scores to determine his or her academic growth or progress. Tennessee currently uses a value-added model comparable to the SGM.

*Student Growth Percentile Model* (SGPM): A growth model using quantile regression where student growth is calculated based on data from other students with similar achievement levels. Academic growth is above, at, or below that of their academic peer group. Colorado, Massachusetts, and Indiana use this growth model.

*Systemic error:* A form of measurement error that is usually caused by the measuring instruments and that skews achievement scores of an entire system.

*Teachers Connecting Achievement and Progress* (TCAP): In Ohio represents the name of classroom-level value-added reporting. This report provides a teacher with his or her effectiveness rating, standard error, teacher effect, number of students in the analysis, and a teacher growth index.

*Teacher Effect:* On a TCAP report, this is a conservative estimate of the student progress for a specific teacher for a certain grade and subject. It is expressed in either scale score points or State NCE’s.
**Teacher Growth Index (TGI):** A score represented by the teacher effect divided by his or her standard error on the TCAP report.

**TeacherInsight (TI):** Developed by The Gallup Organization, an online version of the Teacher Perceiver Interview that provides immediate teacher-screening results to school districts.

**Teacher Perceiver Interview (TPI):** Developed by The Gallup Organization, it is a teacher-screening instrument that estimates potential teacher quality based on 12 life themes.

**Univariate Response Model (URM):** Data model used for calculation of value-added reports in Ohio. This model is only used in Ohio for the creation of projections and when the MRM is not possible.

**Value-added model (VAM):** A means of measuring student academic progress from the beginning of the year to the end of the year. Value-added modeling is calculated with longitudinal standardized achievement data using a multivariate, mixed methods model.

**Summary**

According to Sanders (2000), the “cumulative and residual effects of teacher effectiveness on student academic achievement are measureable and huge” (p. 334). The success of a school district will depend greatly on the effectiveness of its teachers. School districts that are able to identify the highest-quality applicants are most likely to facilitate maximum learning for their students. Some school districts are currently utilizing teacher-screening instruments, such as the TI, to identify the best candidates. School
districts with numerous applicants need an effective and efficient means of identifying the candidates for the interview process. This study will provide information relating to the accuracy of the TI screening instrument to identify effective English/reading and mathematics teachers. School districts must know that the financial investment of the TI screening instrument produces results that improve student achievement.

**Organization of the Dissertation**

This report is organized into five chapters. The first chapter outlines the purpose of the study and identifies the research questions. Chapter II represents a review of related literature that supports the research questions of the study. A brief history of value-added methods and how they are incorporated into Ohio’s accountability and evaluation systems is included in this chapter. The TI teacher-screener instrument is explained as well as its predecessor the Teacher Perceiver Interview. Cautions about these data including several types of measurement and attribution errors are explored. Finally, advantages of the use of teacher-effectiveness data and enhancements that address measurement and attribution errors are explained.

Chapter III addresses the research methodology of the study including the design, research questions and hypotheses, statistical techniques, procedures for collecting data, and data analysis. The data collected include teacher growth indices in English/reading and mathematics, TI scores, years of experience, and gender. Demographic information of the study school district is provided. SPSS® statistical and data management package was used to determine relationships, significance, effect size, and descriptive statistics.
Details of TI categories, the teacher effectiveness rating, and teacher growth index calculations are explained.

The results of the statistical analysis of each question are described in Chapter IV. Tables with descriptive statistics, sample sizes, standard deviations, effect sizes, levels of significance, and coefficients are included. The discoveries of these analyses are explained in detail. A conclusion of the research study is the main focus of Chapter V including interpretations about the use of the variables and the relationships between the TI and teacher effectiveness. The chapter also includes recommendations for future research involving the TI scores and teacher effectiveness data.
CHAPTER II

Review of Related Research and Literature

This review of literature consists of four sections each focusing on a different aspect of value-added or teacher-screening instrument. Section one provides a brief historical overview of the development of value-added models (VAM). The section provides an overview of the various methods for VAM across the country and Ohio’s specific VAM. Additionally, it examines the changes in the complexity of VAM over the past few decades. Explanations on how VAM have become a part of the state accountability system for districts and schools in Ohio are included. Finally, it outlines how the Ohio Department of Education (ODE) is incorporating VAM of teachers into the Ohio Teacher Evaluation System (OTES) (Ohio Department of Education [ODE], 2012b).

Section two provides a brief history of teacher interview procedures as they relate to the interview process. The section will address the need for incorporating these procedures into school districts’ hiring process to help ensure the identification of the most qualified candidates and reduce the administrative time and effort associated with the interview process. Additionally, the Teacher Perceiver Interview (TPI), an interview instrument developed by Selective Research International/Gallup, and its online counterpart the TeacherInsight (TI) are described in detail.

Section three reviews the issues related to utilizing VAM for the purposes of teacher retention, promotion, evaluation, and determining teaching assignment. These issues include measurement error and attribution error. These errors include missing
student data, incomplete or inaccurate linkage of students to teachers, random error, and systemic errors. Other concerns presented relate to test construction and the lack of consistent annual standardized assessments, as well as bias from non-random assignment of students to classrooms.

Section four investigates the advantages of using VAM to help inform decisions relating to teacher retention, promotion, and evaluation. The benefits of VAM over the use of only achievement measures and administrative observations for performance evaluations are explored. The benefits help mitigate some of the concerns relating to the use of VAM in Ohio as a means of estimating teacher effectiveness communicated in section three. Included are new software to improve student linkage to appropriate teachers, more complex statistical modeling, and improved standardized assessments.

**Brief History of Value-Added**

Value-added data are changing the face of public education. The introduction of student progress or value-added data into the accountability system is having a profound impact on school districts across Ohio. Value-added approaches have their limitations, but are more appropriate than the reporting of a school district achievement test averages, which include confounding factors outside of the school’s control (Amerein-Beardsley, 2008; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Sanders, 2000). School districts have used several different VAM over time. Each VAM has advantages and disadvantages.

One of the earliest versions of value-added analysis is the standardized gain model (SGM). To determine standardized gains school districts take the difference
between the current and previous standardized test scale scores. This method could work for vertically scaled or horizontally scaled standardized assessments. Vertically scaled standardized assessments, such as CTB-McGraw-Hill’s TerraNova have scale scores that increase with each grade level (CTB McGraw-Hill, 2012). A student who increases by the necessary scale score points to maintain their place in the distribution from one year to the next experiences a year’s worth of growth. For example, a student who achieves a scale score of 624 (which represents achievement at the 50th percentile) on the 3rd grade reading TerraNova would need to achieve a scale score of 637 on the 4th grade reading TerraNova to maintain achievement at the 50th percentile (CTB McGraw-Hill, 2012). Maintaining his or her place in the distribution requires a student to have progressed a year academically. Vertically scaled assessments do have limitations that can affect the validity of inferences drawn from them because content and statistical specifications of each assessment typically change across grade levels (Buzick & Laitusis, 2010).

A horizontally scaled assessment such as the Ohio Achievement Assessments (OAA) has scale scores that are comparable over each grade level (Ohio Department of Education [ODE], 2012c). A student who experiences a year’s worth of growth in a horizontally scaled assessment would maintain his or her scale score from one year to another. For example, a student who scores a 400 in 4th grade math needs to experience a year’s worth of content to score a 400 in 5th grade math thereby experiencing a year’s worth of growth. There is a high level of measurement error with this method of VAM (Wright, 2010). The fact that scale scores on achievement tests are heteroscedastic, meaning there is the greatest amount of variance for students scoring at the high and low extremes is one cause of measurement error (McCaffrey et al., 2004). This type of VAM
also tends to correlate more closely to school composition, meaning that high-poverty and high minority schools are more often rated as ineffective (Wright, 2010). The transparency for the end user is why this VAM is popular; however, this does not outweigh the bias of the measurement error (Wright, 2010). Tennessee’s value-added model uses this basic premise, although it is more sophisticated including the use of normal curve equivalents (Tennessee Department of Education, 2012).

Another VAM that is gaining in popularity is the student growth percentile model (SGPM), which is currently being used in Colorado, Massachusetts, and Indiana. It is similar to the standardized gain model above, except it utilizes conditional percentile ranks and multiple prior test scores when available (Wright, 2010). Although the SGPM is more complex than the SGM, it is still relatively easy to explain. In a SGPM, all available student test scores are included in the analysis. Student growth is calculated through a statistical model called quantile regression (Colorado Department of Education, 2012). SGPM allows a student’s growth to be compared to other students with similar scores to provide a context to their progress. The SGPM’s normative, value-added approach provides information as to whether the student’s progress is above average, typical, or below average as compared to an academic peer group (Colorado Department of Education, 2012). In this way all students, even low-achievers, are able to show high growth. The SGPM does have some concerns relating to its instability at the extremes (percentiles near one or 99). The instability is likely a cause of increased estimation error in the estimated regression coefficients. Estimation error will also cause the standard errors to be unstable (Wright, 2010). The standard error provides a confidence band with which we are able to designate the effectiveness of teachers at the classroom level. If
these standard errors are unstable, it greatly impairs the usefulness of this VAM to
differentiate the effectiveness among teachers. Wright (2010) notes that in high- and low-
scores on standardized assessments there is more measurement error than in the middle of
the distribution due to the heteroscedacity of scores. This leads to another concern for the
SGPM because it would mean more bias against high- and low-scoring students and
teachers with classrooms predominantly with these types of students.

In the mid-1990s, William Sanders worked with data across the state of
Tennessee to create a statistical model that was complex and able to address some of the
previous measurement error concerns. Sander’s value-added modeling began from an
education production function (Hanushek & Rivkin, 2010; Sanders & Rivers, 1996). The
multivariate mixed model incorporates instructional time and students’ previous
standardized test scores to create a prediction score for each student. Sander’s VAM uses
a student’s own academic performance as a basis for determining his or her growth or
academic progress (Battelle for Kids [BFK], 2012b). The model allows for the
incorporation of fixed and random effects and within-group error (Doran & Izumi, 2004;
Kupermintz, 2002; Sanders & Rivers, 1996).

The EVAAS® model for calculating student growth is actually multiple models
depending on the test data available (Wright et al., 2010). There are two general types of
VAM’s utilized in Ohio; multivariate response model (MRM) and univariate response
model (URM). The MRM is similar to a multivariate repeated-measures analysis of
variance (ANOVA) model. It uses all of the actual test scores for each student
simultaneously. The MRM does not represent the school within the model to allow for
comparisons of teachers at different buildings across the state. Teachers are also entered
as random effects to allow for the possibility of many teachers with very few students per teacher (Wright et al., 2010). The teacher effect is expressed as a deviation from the average gain for the entire pool, in this case the state of Ohio.

In certain situations, the data structures don’t allow for the use of the MRM (Wright et al., 2010). One example of this is the end of course exams used for calculating growth of high school and some middle school teachers. In this instance, the URM is used which is similar to an analysis of covariance (ANCOVA). In the URM, the student scores serve as the dependent variable and their previous scores serve as the independent variables. The categorical variable would be the teacher and is treated as random not fixed. The URM does not allow for the layering of MRM so instead a composite score is used in the ANCOVA model to capture prior schooling (Wright et al., 2010).

A common problem for many VAM’s is missing student data. This can occur due to illness on the test day, testing data being lost, or a student being new to the school district. EVAAS® does not impute missing data explicitly, but rather uses the correlation between current and prior scores in the non-missing data to estimate a mean for the prior and current score as if there were no missing data (Wright et al., 2010). The result is an estimate of the average gain for the group of students. It is more precise than the simpler methods of calculating the difference of the means or means of the differences (Wright, 2004).

Other models have complex statistical models for determining student growth; however, only Sanders’ model also has Bayesian shrinkage estimators used together with three-year averages of teacher-level growth data to help produce more precise estimates of teacher effectiveness (Doran & Izumi, 2004; Kupermintz, 2002; Sanders & Rivers,
The shrinkage estimator provides an additional adjustment for errors that other statistical models do not contain. Therefore, in practical terms, the shrinkage estimation makes it more likely for a teacher with a small class size to be rated as an effective teacher. Shrinkage estimation provides additional protection against an inaccurate estimate of a teachers’ effectiveness designation. Either by a false positive by rating the teacher inaccurately as effective or a false negative by rating the teacher inaccurately as ineffective (Wright et al., 2010). Sander’s model also uses all student standardized test data simultaneously to help decrease measurement error from any single test (BFK, 2011).

Various Methods of Identifying Teacher Effectiveness

Measuring teacher effectiveness is a complicated issue (Welsh, 2011). Historically, there have been various methods for determining the effectiveness of teachers. According to Welsh (2011), the difficulty is that there is more to an effective teacher than using a particular strategy or specific content knowledge. School districts typically use teacher inputs such as advanced degrees, certification, years of experience, teacher certification exams that measure content knowledge, teacher observations, or some combination of these characteristics to identify high-quality teacher candidates (Stronge, Ward, & Grant, 2011). Unfortunately, each of these teacher inputs has one or more fatal flaws when it comes to actually linking the method to improved student achievement. Below I will review the flaws associated with each of these proxies of teacher effectiveness.
One method of evaluating the effectiveness of teachers is through their attainment of advanced degrees including master’s degrees in education or National Board Certification (Amerein-Beardsley, 2008). National Board Certification represents an attempt to find a process that identifies the most effective teachers who are able to improve student achievement (Goldhaber & Anthony, 2007; Jacob et al., 2009). There has been some discussion as to whether National Board Certification actually equates to student academic gains (Goldhaber & Anthony, 2007; Harris & Sass, 2009; Sanders, Ashton, & Wright, 2005; Welsh, 2011). The effectiveness of NBCT are mixed. NBCT are more effective than non-NBCT during the year of their certification, but appear to be no more effective post-certification than non-NBCT (Goldhaber & Anthony, 2007). In some cases the NBCT is actually less effective than non-NBCT post-certification which is confounding.

Teaching credentials appear to be a less effective means of identifying teacher quality than measures related to student test scores (Rockoff, 2004). However, all fifty states require educators to have certification or licensure in order to teach. Interestingly, Walsh and Podgursky (2001) contended that there is no compelling evidence that teachers with appropriate teacher certification are of higher quality than those without certification when it comes to producing greater student achievement. Teacher certification exams only measure content knowledge and teaching pedagogy rather than actual instructional practice. Teacher certification exams also often assess whether the teacher has what Welsh (2011) calls the “minimum level of content knowledge required for effective instruction” (p. 764). Unfortunately, attainment of conceptual knowledge
does not ensure a teaching candidate is able to implement the knowledge and pedagogy effectively.

Teacher experience is a characteristic that effectively identifies high-quality educators (Jacob et al., 2009; Koerner, 2007; Regan & Hayes, 2011). However, this is not a reliable characteristic to use for hiring decisions because it is not realistic for school districts to hire only teaching veterans (Jacob et al., 2009). Teaching experience is not possible for a new teacher and therefore does not provide a school district with a means of identifying if they will be effective in the classroom. Teacher content knowledge is another characteristic considered when identifying effective teaching candidates, but it too comes with concerns (Novotny, 2009). First, at best we can only identify estimated proxies of content knowledge. Teacher content knowledge is measured through subject-specific teacher exams. Second, it is difficult to identify these proxies for all content areas. Employment criteria of this type have produced mixed results (Jacob et al., 2009). Jacob et al., (2009) only found evidence of a positive relationship in mathematics with regard to content knowledge for teachers in first and third grades. Teacher beliefs and values as they relate to teacher self-efficacy is another means for school districts to identify a quality teacher. The problem is that studies have typically linked estimates of self-efficacy to supervisor evaluations. These evaluations are typically subjective and have small correlations to student achievement (Jacob et al., 2009).

Teacher observation is a common method for determining the effectiveness of an educator (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Welsh, 2011). Teacher observations have several issues with determining teacher effectiveness. Firstly, they do not assist a school district in hiring a pre-service teacher because they
typically do not have formal observations by an administrator in their student-teaching experience. For experienced teachers there are other problems relating to bias of the observation. Typically, a teacher has advanced notice of the formal observation, which allows the teacher to prepare for the formal observation and provide instruction that may not be typical on a daily basis. The advanced notice could be as little as a day or as much as a week prior to the formal observation. This phenomenon is the Hawthorne effect (Welsh, 2011). There is also the potential of bias from the observer commonly known as Halo effects. A halo effect refers to a principal’s taking into account the overall perception of the teacher by his or her peers or the community. Non-instructional factors, such as willingness to serve on committees or perform duties, may also influence the results of the formal observation. Another confounding factor is the teacher who is very organized or caring toward students, but not effective instructionally. Organization and caring for students are important, but do not necessarily affect the instructional effectiveness of the teacher (Welsh, 2011). Jacob and Lefgren, (2008) note that most administrators are only able to distinguish teacher effectiveness at the extremes through formal observations and not those in the middle. Finally, teacher observations do not typically differentiate the effectiveness of teachers, as most receive the highest ratings (The Gallup Organization, 2011; Kober, & Rentner, 2012; Weisberg, Sexton, Mulhern, & Keeling, 2009).

**Defining Teacher Quality**

The Gallup Organization created the TI because they believe teaching is an art and some candidates have a raw talent for the profession that can be identified early
Defining teacher quality is difficult because there are so many different meanings across the country. If a single method of identifying teacher quality existed every student would have one (Stronge, 2002). One definition of teacher quality includes those who have low absenteeism, facilitate student growth, and obtain positive ratings from colleagues and students via surveys (Smith, 2012). Goldhaber and Hannaway (2009) state that teacher quality is “the set of teacher skills, knowledge, personal attributes, and pedagogical abilities that yield desired student outcomes” (p. 30). Some of these teacher skills include professional preparation; affective characteristics like enthusiasm, respect, and caring for students; classroom management; organizing for instruction including time management, high expectations, and effective planning; instructional delivery; and monitoring progress. Ciotte (2009) defines teacher quality as a blend of both personal and professional characteristics. Personal characteristics include enthusiasm, positive attitude, creativity, respect, high expectations, and compassion. Professional characteristics include certifications, college degrees, and passing scores on teacher exams. Over 300 rural superintendents in Ohio believe that enthusiasm, creativity, and technology skills identify a quality teacher (Smith, 2012).

There is currently no consistent traits of teacher quality that are identified on one assessment (Goldhaber & Anthony, 2007). Without a standard definition of teacher quality, how can we begin to create an assessment to accurately measure teacher quality? There is also a difference between teacher quality and what the teacher actually does in the classroom known as teaching quality (Cicotte, 2009). Having an agreed upon standard definition for teacher quality would be beneficial for both school districts and organizations creating teacher screening instruments. For the purpose of this study,
teacher quality is defined exclusively as those who can facilitate a year or more of student academic growth based on the TGI score.

Ohio Teacher Evaluation System

Recently, school districts have been moving toward reviewing teacher products as a means for determining teacher effectiveness (Stronge, et al., 2011). Teaching products refer to student achievement and progress. Many states are moving toward using value-added metrics to identify quality teaching (Buzick & Laitusis, 2010; NCTQ, 2012). Ohio is one of these states with the creation of the Ohio Teacher Evaluation System (OTES) using student academic progress as fifty percent of the overall final summative evaluation rating (ODE, 2012b).

It is important that the evaluation instrument incorporate all available information relating to the effectiveness of a teacher, including value-added data (Glazerman, Loeb, Goldhaber, Staiger, Raudenbus, & Whitehurst, 2012; The Gallup Organization, 2011; ). The growth portion of the evaluation represents a three-year average of student progress when possible (ODE, 2012b). Incorporating a three-year average of student growth is important because value-added data are estimates. By using three-year averages instead of a one or two-year average there is more protection for teachers from measurement error that could misclassify a teacher’s effectiveness (BFK, 2012b). The OTES growth metric includes the student growth from all subjects taught by the teacher. The result is a longitudinal measure representing an overall composite of the teacher effect, which is more favorable than a one-year measure (Misco, 2008). It is important that the VAM have a high level of precision and accuracy, because these results carry important
consequences such as retention and promotion decisions (McCaffrey et al., 2004). Using a three-year average of student progress will help to lessen any instability in the value-added analysis (Glazerman et al., 2012).

Prior to OTES, school districts did not need to incorporate student progress data into teacher evaluations. The OTES was created with input from teachers, administrators, and community members. The model utilizes more than just teacher-level value-added data to evaluate overall teacher performance by incorporating formal and informal observation data and other evidence (Misco, 2008; ODE, 2012b; Sanders & Horn, 1998; Scherrer, 2011). However, value-added analysis allows for the isolation of teacher effects on student progress, which is a fairer measure for comparing teachers than achievement levels (Scherrer, 2011). The new evaluation instrument also provides feedback to teachers on their performance. OTES provides each teacher with either an improvement plan or growth plan thus encouraging the process of continuous improvement for all teachers.

In Ohio, House Bill 153 required the creation and implementation of OTES (ODE, 2012b). Figure 2.1 below illustrates the evaluation framework of the new evaluation system, which incorporates both a teachers performance and student growth measures equally to determine the overall summative rating.
The State Board of Education and Ohio Department of Education developed the new OTES to better recognize and reward teaching excellence, improve the quality of instruction for students, improve student learning, strengthen professional development, and inform retention and promotion decisions of employees (ODE, 2012i). Figure 2.2 illustrates the evaluation matrix of the Ohio Teacher Evaluation System which combines the results of both student academic growth and the teacher’s performance to create a final summative evaluation rating (ODE, 2012i). The following matrix is referenced in Ohio Revised Code 3319.111 and 3319.112 and is used to determine the final summative rating of each teacher (Ohio Revised Code, 2013).
Figure 2.2 OTES Evaluation Matrix.

There are four overall performance ratings for teachers in the OTES:

- Accomplished – a leader and model in the classroom, school, and district, exceeding expectations for performance.
- Proficient – consistently meets expectations for performance and fully demonstrates most of all competencies,
- Developing – demonstrates minimum competency in many of the teaching standards, but may struggle with others, and
- Ineffective – consistently fails to demonstrate minimum competency in one or more teaching standards (Ohio Department of Education [ODE], 2012i, p. 23).

There are three growth designations with OTES – above, expected, and below growth. For the growth metric, Ohio uses EVAAS® – which differentiates teacher effectiveness into five levels (EVAA®S, 2012). Those five levels are:
• Most Effective – teachers whose students are making substantially more progress than the state or pool average noted by a teacher index equal to or above positive two.

• Above Average – teachers whose students are making more progress than the state or pool average noted by a teacher index equal to or greater than positive one, but less than positive two.

• Average – teachers whose students are making progress similar to the state or pool average noted by a teacher index equal to or greater than negative 1, but less than positive one.

• Approaching Average – teachers whose students are making less progress than the state or pool average noted by a teacher index equal to or greater than negative two, but less than negative one.

• Least Effective – teachers whose students are making substantially less progress than the state or pool average as noted by a teacher index less than or equal to negative two (EVAAS, 2012).

The teacher growth index (TGI) is calculated by dividing a teacher’s effect score by its corresponding standard error (EVAAS, 2012). For the purpose of OTES, the middle three levels (above average, average, and approaching average) are combined to represent expected growth. This allows for a confidence band of 95 percent for each of the three levels (EVAAS, 2012). The purpose for dividing the average level into three levels is to provide a teacher with a sense of whether he or she is at the upper or lower end of average. This can help provide a context of how his or her effectiveness is trending over time. Only teachers with an above teacher growth index are eligible for the highest OTES
final summative rating of accomplished. Teachers with the highest performance ratings have more autonomy including an exception to the annual evaluation process and selection of their evaluator (ODE, 2012b). The addition of value-added into the OTES helps to ensure that during teacher layoffs the most effective teachers are retained (Glazerman et al., 2012). The final summative rating for a teacher is used to make teacher retention and promotion decisions. School districts are required to develop a process for removing poorly performing teachers that must include a teachers’ final summative rating. The passage of House Bill 555 made this an official policy as Ohio Revised Code 3319.112 (E) (Ohio Revised Code, 2013).

**Brief History of Teacher Interview Process**

The quality of the teacher is the most important factor related to student achievement (The Gallup Organization, 2011; Goe, Bell, & Little, 2008; Regan & Hayes, 2011; Wright, Horn, & Sanders, 1997). However, to ensure school districts hire effective teachers they must first identify high-quality candidates from the applicant pool. The process to identify and recruit the most highly qualified teaching candidates tends to be time-consuming (Novotny, 2009; Regan & Hayes, 2011). Many school district hiring practices are also highly subjective (Novotny, 2009). High turnover rates in some districts require the need to recruit teaching candidates continuously. The hiring process takes considerable time, effort, and financial resources. With budgetary constraints, school districts are looking for screening instruments to ensure recruiting high-quality candidates in the most efficient means possible. The problem is greater in urban school districts. (Jacob et. al., 2009; Regan & Hayes, 2011). According to Novotny (2009), the
inconsistency of the teacher-selection process in school districts across the country necessitates the use of a commercial screening instrument for standardization.

**Teacher Perceiver Interview**

In 1960, the Selective Research International/Gallup organization developed the Teacher Perceiver Interview (TPI) to help identify in a valid and reliable manner potential teacher candidates who have strengths that are specific to effective teachers (Faurer, 2004; Novotny, 2009). According to Novotny (2009), thousands of public school districts use this assessment across the nation. The assessment attempts to identify life themes in the candidate that match the habits or behaviors found in the most successful teachers (Faurer, 2004). There are three categories of questions on the TPI – situational, observational, and personal. The TPI utilizes questions that address intrapersonal, interpersonal, and extrapersonal themes (Faurer, 2004, The Gallup Organization, 2012). Questions in the intrapersonal theme attempt to measure how a candidate views the mission, investment, and focus of education. Mission relates to a deep belief that students can grow and attain self-actualization. A teacher who has mission believes it is his or her goal to make a significant contribution to others. The notion of investment represents the capacity to receive satisfaction from the academic growth of each student. A teacher shows investment through responses to each student’s learning as opposed to the teacher’s own performance. Focus refers to the teacher’s having models and goals. A teacher shows this focus through using these goals to develop intentional activities and strategies (Koerner, 2007).
The interpersonal questions can estimate the candidates’ empathy, rapport drive, listening, and objectivity (The Gallup Organization, 2012). Empathy is the ability to put yourself into another’s situation. A teacher with empathy is able to tune into a student’s thoughts and feelings and use them as feedback. Rapport drive refers to a teacher’s having an appropriate and effective relationship with each student. A teacher with positive rapport drive sees the teacher-student relationship as critical to learning. Listening is more than when a teacher actively listens to others; they must also accept and respond to these interactions. Objectivity is evident through the teacher’s response to the whole situation. A teacher demonstrates objectivity by collecting and understanding all facts prior to making a decision (Koerner, 2007).

Finally, questions in the extrapersonal theme attempt to quantify candidates’ individual perception, input drive, activation, innovation, and gestalt (Faurer, 2004; The Gallup Organization, 2012). Individual perception or individualization refers to a teacher’s considering the interests and needs of each student while developing the educational program. Input drive is the internal desire to search out ideas, resources, and experiences to help his or her students. Activation refers to a teacher’s ability to stimulate students to think, respond, and feel while learning. An innovative teacher is one who is willing to take risks by trying out new ideas and instructional techniques. Innovation is more than just being creative, as a teacher must be willing to move beyond finding these creative ideas to actual implementation. Gestalt or perfection drive is the motivation toward completeness. A teacher represents high gestalt when they strive for perfectionism and feel uncomfortable when work is not completed. Although form and structure are
vital concerning instruction, the teacher considers first the individual student (Koerner, 2007).

Resources – both financial and human – are becoming scarcer in school districts (Regan & Hayes, 2011). The Gallup Organization created the TeacherInsight (TI) as an automated online interview used by numerous school districts across the country to help identify the best possible teaching candidates (Jacob et al., 2009). The TI allows for more access of teaching candidates to human resource departments because they can complete the screener at their convenience without the assistance of a school district employee (Regan & Hayes, 2011). Using the TI also reduces the added cost of criminal background checks on less qualified candidates (Smith, 2005). It takes candidates about 30 minutes to complete the TI, and the system records each answer so that if an Internet connection is lost the candidate can begin where they left off (Novotny, 2009; Smith, 2005). The 40 question TI consists of multiple choice questions with 50 seconds to answer among the four options, forced-choice questions with 50 seconds to choose the statement that best describes the candidate between two different statements, and questions with a five-point Likert-scale for the candidate to rate his or her level of agreement to a statement in 20 seconds. If a candidate takes more time than allotted for a question, the response will not be recorded. Candidates cannot skip questions or return to them later. Candidates must have a district access code to be able to take the assessment or a direct link (The Gallup Organization, 2012) The TI scores are presented in scores ranging from a low of 0 to a high of 99 with a recommended cutoff of 67 (Regan & Hayes, 2011).

School districts should consider the TI screening results as only one of several pieces of information during the teacher selection process (The Gallup Organization,
School district human resource officials have immediate access to the results of the TI. Candidates, however, do not receive the results of the TI assessment. Without the benefit of receiving the score a student will not know how well or poorly he or she performed on the TI. Candidates may only take the TI once in a 12-month period with a school district. According to The Gallup Organization (2007), the TI results are an excellent indicator of quality teaching candidates if they answer the questions honestly and accurately. This process allows district administrators to spend their time more efficiently with the candidates identified by the TI instrument as most effective. The Gallup Organization (2007) considers the instrument a fair measure because they have researched and tested it to ensure it identifies high quality candidates. It also ensures that all applicants experience the same questions and therefore are evaluated the same way (The Gallup Organization, 2012).

The advantage of school districts moving toward using an interview instrument, such as the TI, is that it is highly structured, objective, reliable, valid, and provides an in depth understanding necessary to identify talents that result in quality teaching candidates (Faurer, 2004; The Gallup Organization, 2012; Novotny, 2009). The TI is a web-based version of the Teacher Perceiver Interview that is more convenient for both the teacher candidate and the school district (Koerner, 2007). The online structured interview process allows the candidate to provide his or her thinking without influence from the interviewer (Koerner, 2007). Another advantage to using a candidate-screening instrument such as the TI is there is research to support that candidates who score low on the TI do not possess the qualities or talent necessary to be the most effective teachers (Faurer, 2004; The Gallup Organization, 2012).
Caveats of Value-Added

The use of value-added reporting is not new; it has occurred for decades in various forms across the country. The initial uses of value-added metrics were program evaluation and district or school accountability. There are several concerns with determining ways to use value-added data including (a) attribution errors, such as accurate student linkage, removing district and student effects from teacher effects, and the impact of previous teachers’ effects on future teachers’ effect; (b) measurement errors resulting from incomplete student records, random, and system errors; (c) test construction; and (d) non-random assignment of students to classrooms. Due to these limitations, initial VAM used standard scale score point increases annually across various subjects to determine student growth. These were crude measures and did not account for concerns relating to attribution error, measurement error, and test construction (Hanushek & Hoxby, 2005; Kupermintz, 2002). Today, we have software and computers capable of handling the complex statistical modeling necessary for improving utilization of teacher-level value-added data (Hanushek & Rivkin, 2010). However, this does not eliminate the concerns relating to the use of value-added data for the purposes of accountability. Validity and reliability are critically important to any value-added measure selected to become part of an accountability system, especially one that has the potential to affect employment, pay, assignment decisions, or evaluations of staff (Hanushek & Hoxby, 2005; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). School district leaders must understand each of these issues and be able to mitigate their effect to ensure appropriate use of value-added data.
**Attribution Error**

Attribution error when used in discussions on VAM refers to ensuring that students, and their relative growth, are connecting to the appropriate teacher or teachers (Kennedy, 2010). Linking students accurately to their teachers is a common problem in creating an accurate value-added model for teachers (Amerein-Beardsley, 2008; Doran & Izumi, 2004; Hanushek & Rivkin, 2010; Martineau, 2006; McCaffrey et al., 2004). Student linkage includes linking students to all teachers who provide instruction and accurately determining the percent of instruction each teacher provides. The process of student linkage is problematic for school districts for several reasons. First, many districts do not have student information systems that can identify more than one official teacher. In Ohio, school districts must report this “teacher of record” to the Ohio Department of Education for the purposes of ensuring proper teacher certification (Ohio Department of Education [ODE], 2012e). There are also issues with changes to student schedules not being recorded in these systems throughout the year. Flexible grouping across teachers throughout the year is becoming more popular. This practice involves a team of teachers allowing students to move in and out of their classroom based on unit or chapter assessments. It allows the team of teachers to group students for specific lessons based on students strengths and weaknesses. Although this may be a promising practice it can contribute to attribution error.

Questions even arise about the feasibility of removing previous teacher effects from current teacher effects (Darling-Hammond et al., 2012; Goe et al., 2008; Martineau, 2006; Newton, Darling-Hammond, Haertel, & Thomas, 2010). Not assessing all subject areas annually poses a problem in Ohio when determining teacher effectiveness.
(McCaffrey et al., 2004). Currently assessments only exist for science in grades five, eight and ten in Ohio. There are no state assessments for writing or social studies except for grade ten (Ohio Department of Education [ODE], 2012a). Trying to calculate teacher effectiveness over several years of data instead of annually increases the attribution error of the designation and potentially increases misclassification of teacher effectiveness (Welsh, 2011). This is because it is difficult to determine which teacher is actually responsible for the student progress or lack thereof when the reports are only created at certain grade levels. Ohio’s new evaluation system addresses this concern in two ways. First, teachers are provided with annual teacher effectiveness data for the core subjects. They are also provided a running three-year average of their student progress which increases the precision of the TGI score (SAS, 2012).

Tracking students with disabilities across years and different programs can also cause problems. Students with disabilities may take the regular assessment one year and an alternate assessment in future years. It is not possible to use alternate assessment results in Ohio’s VAM so these students have gaps in their achievement history preventing their inclusion in value-added analysis for the teacher (SAS, 2012). Not including all students can have an effect on the sample size for a teacher whose classroom is mainly students with disabilities. As the sample size decreases, so can the certainty relating to the teacher effect. If the sample size becomes too low it may not be possible to calculate a value-added report (BFK, 2011).

Finally, there can be tutors, parents, or others working with students outside of the classroom that influence student growth (Darling-Hammond et al., 2012; Newton et al., 2010). It is important to recognize that student test score improvements could be due to
instruction and experiences outside of the school day (Scherrer, 2011; Welsh, 2011). This is especially true for gifted learners and students with disabilities. Gifted learners often receive enrichment through experiential learning and tutors outside of the classroom (Welsh, 2011). Students with disabilities in Ohio may be eligible for “extended school year” services, which allow funding for additional instruction over the summer beyond the typical 180-day school year (Ohio Department of Education [ODE], 2012f). Parents of students with disabilities may also obtain additional tutoring to help their child reach grade level achievement (Buzick & Laitusis, 2010). Out-of-school learning experiences are also different based on socio-economic status. Typically, advantaged students experience a great deal of out-of-school learning and their disadvantaged peers experience summer academic decay (Darling-Hammond et al., 2012; Scherrer, 2011). How can we be certain which teacher to attribute the student progress, whether it is more or less than a year’s worth of growth, if we do not address these problems?

**Measurement Error**

Measurement error typically refers to two possible issues, random error and systemic error (McCaffrey et al., 2003). Random error affects a student completely by chance. A student’s emotional or physical state could influence their score either positively or negatively on a particular day. Consider this example of a random error. A student has a broken limb, which has a negative impact on his or her achievement due to the frustration of completing the answers with their non-dominant hand, therefore providing an inaccurate picture of their achievement. Systemic error is any factor that influences the entire system (McCaffrey et al., 2003). In this instance, imagine that the
office makes an announcement over the intercom during testing. Errors of this type could have an impact on all of the students who are testing in this classroom, therefore affecting the entire system. School district leaders need to consider these caveats when determining appropriate uses of value-added data and when responding to applicable concerns.

Another measurement error concern relates to gifted learners. Gifted students are more likely to achieve high-test scores regardless of the instructional practices of their teachers (Welsh, 2011). Classrooms with high percentages of gifted students can affect the teacher effectiveness rating. Measurement error can also manifest itself with teachers only teaching low-level concepts often measured on state achievement tests (Jacob et al., 2009; Scherrer, 2011). The potentially unethical practice of teaching test-taking skills not aligned to long-lasting learning can provide an inaccurate measure of effectiveness through score inflation (Darling-Hammond et al., 2012; McCaffrey et al., 2004). As accountability increases for value-added measures, unethical practices are likely to increase (Scherrer, 2011; Welsh, 2011).

Classroom level value-added analysis poses other concerns relating to students with disabilities. There are several concerns relating to the use of test scores from individual students over two or more years due to testing accommodations and modifications (Buzick & Laitusis, 2010). It is common for students with disabilities to have either testing accommodations or modifications. An accommodation is a change in testing that does not alter the specific construct being measured by the assessment (Buzick & Laitusis, 2010). An example of a testing accommodation would be allowing extended time for completing the test, reading questions or instructions to the student, or allowing the testing to occur in a small group or one-on-one setting (Ohio Department of
A testing modification is a change that does alter the specific construct being measured by the assessment (Buzick & Laitusis, 2010). An example of a testing modification is reading the actual reading passage to a student (Ohio Department of Education [ODE], 2012d). This changes the reading assessment to a listening assessment. If school districts utilize testing modifications, the testing results and eventual value-added analysis may not accurately reflect the teacher’s effect. Although testing accommodations should not affect the testing results or value-added analysis, there is conflicting research relating to inconsistent use of accommodations over time (Buzick & Laitusis, 2010). Students with disabilities are typically well below grade level and may have additional barriers connected to previous frustrations in school. These barriers could make showing growth for these students more difficult (McCaffrey, et al., 2004).

**Non-Random Assignment of Students**

Non-random assignment of students to teachers is another common issue that can have an effect on teacher-level value-added analysis (Darling-Hammond et al., 2012; Jacob & Lefgren, 2008; Newton et al., 2010; Rothstein, 2008; Scherrrer, 2011; Stronge et al., 2011). Non-random assignment of students to teachers can have an effect on teacher-level value-added analysis at two levels. First, student populations in various schools are not random. By this, I mean there tends to be more advantaged students attending affluent school districts (Scherrrer, 2011). Teachers in schools with high percentages of advantaged students are more likely to have students who often have extended learning opportunities. The opposite is true for disadvantaged students. Both the schools, and
teachers, are likely to have students who do not have access to these extended learning opportunities (Scherrer, 2011). The problem of non-random assignment of students occurs even within schools. It is common to have classrooms with a majority of high- or low-achieving students (Scherrer, 2011). Second, non-random assignment could be due to building administrator or parent requests. For instance, it is possible that either the principal purposely assigned students who have difficulty behaving in school or students with similar abilities to a particular teacher who they felt had the best classroom management (Rothstein, 2008). The non-random assignment of students could account for the difference in the particular teachers’ effectiveness rating for the given year (Newton et al., 2010; Stronge et al., 2011). The result of non-random assignment of students could have misleading implications on teacher-level value-added results.

**Test Construction**

In Ohio, test construction issues remain a concern with value-added analysis. For instance, if a student scores proficient for two consecutive years in reading, the student would have been determined to have met a year’s worth of growth in Ohio (Nicholson, M., personal communication, July 13, 2012). The idea is that the student must have learned a year’s worth of content to be able to maintain his or her proficient status. The problem inherent in this thinking is that it assumes that the state assessments are equated at each performance level. The issue arose because over the three-year period of creating and implementing the state achievement assessments in Ohio value-added reporting did not exist. How could the test creators predict the use of value-added analysis at that time?
Therefore, the problem for district leaders is that key stakeholders will not immediately understand this effect.

The following scenario demonstrates the performance level equation issue. A student scores advanced on the third grade reading Ohio Achievement Assessment (OAA), accelerated on the fourth grade reading OAA, and then proficient on the fifth grade OAA over a three-year period. It might appear on the surface that the student is regressing over this three-year period. Although regression is possible, it is equally possible that the difference is merely a factor of differences in performance level equating. It takes more than a year’s worth of growth to maintain the advanced performance level from third to fifth grade (ODE, 2012c). School district leaders need to understand this as a potential weakness of Ohio’s current achievement system. Therefore, communication to district stakeholders is necessary to prevent unfair judgment and misconceptions about student progress and teacher effectiveness.

Another concern with the calculation of teacher effectiveness through value-added relates to students at the extremes of achievement. Many state achievement tests only accurately assess the knowledge and skills of average students (Darling-Hammond et al., 2012; Newton et al., 2010; Welsh, 2011). If the assessment does not have the adequate academic stretch for the gifted learner, it cannot appropriately measure his or her academic gains because they hit the ceiling of the assessment (Welsh, 2011). The same concern exists with students with disabilities (Scherrer, 2011). The assessment must have a low enough floor to adequately measure the low-achievers true understanding to therefore estimate the students’ academic growth. So in essence the assessment may actually be informing us that no growth has occurred on what this assessment is
measuring (Sherrer, 2011, p. 131). Standardized assessments used to estimate the effectiveness of teachers must be sensitive enough to determine the academic gain due to instructional practices otherwise the teachers could not show growth on what was measured (Buzick & Laitusis, 2010; Welsh, 2011). This could lead to a possible misclassification of a teacher effectiveness rating.

The Potential of Value-Added to Improve Instruction

There is great potential for improving instruction with the incorporation of teacher-level value-added data into the accountability system in Ohio. Many of the advancements in VAM address the concerns referenced above. The use of value-added data helps school district administrators in several ways. First, it can provide a measure of the academic effects of their curricula, instructional practices, and professional development opportunities on student achievement. It can also assist the allocation of scarce resources with data-driven decisions (Glazerman et al., 2012). School districts in Ohio already benchmark their achievement against other districts across the state; value-added data allows for similar benchmarking with regard to student progress. Finally, school districts can use value-added data to identify “bright spots” instructionally within their buildings (Misco, 2008). Administrators can then leverage these promising practices across classrooms to help improve student achievement (BFK, 2011).

Teachers are often inundated with data, but rarely does it help them in making instructional decisions (BFK, 2011). Value-added data can provide a teacher with the data necessary to inform instructional decisions that can improve student achievement in several ways. First, Value-added data provides information to identify which type of
student most benefits from his or her instruction (high, middle, or low-achievers). Value-added data allow a teacher to reflect on his or her instructional practice and reallocate, align, or focus classroom resources to help all students to achieve at high levels. The teacher can also modify his or her instruction to address the needs of all students (Misco, 2008). Differentiation of instruction allows the teacher to meet the diverse needs of the students in his or her classroom. Using individual student projections, teachers are able to monitor progress through the school year to ensure academic growth for every student. Student projections provide a measure of each student’s academic potential at the beginning of the year when there is time to make actionable, instructional changes (BFK, 2011). Teacher self-selection of professional development opportunities can also be informed by value-added data. Finally, value-added data allow teachers to identify the strengths and weaknesses of their current instructional practice (Glazerman et al., 2012).

One problem in many states is that they do not have testing every year for each subject (Hanushek & Hoxby, 2005). Ohio initially had proficiency tests only in grades four, six, and eight. Inconsistent testing increases attribution error, because testing over several years it would be impossible to know which teacher to attribute the student progress (Welsh, 2011). In 2006, Ohio implemented annual testing in grades three through eight for reading and mathematics to address this concern (ODE, 2012a). Currently, Ohio assesses every student in reading and mathematics in grades three through eight (ODE, 2012a). The Partnership for Assessment of Readiness for College and Careers (PARCC) will create new state assessments in English/reading and mathematics equating them across grade levels and performance levels alleviating the appearance of student regression (Partnership for Assessment of Readiness for College
and Careers [PARCC], 2012). These assessments will be operational in the 2014-2015 school year for grades three through eleven. The PARCC assessments will be more rigorous than the current assessments, which are based on a minimum competency rather than college and career readiness. PARCC assessments should eliminate or greatly reduce score inflation from student guessing, as the process of elimination strategy is not effective with most questions on the assessments (PARCC, 2012). Consequently, this should create value-added estimates of teacher effectiveness that are more accurate.

In Ohio, there are 120 school districts with over 830 schools participating in Project SOAR with Battelle for Kids. Project SOAR requires member school districts to assess all students in English/reading, math, science, and social studies from second grade through high school (Battelle for Kids [BFK], 2012a). The systematic testing of all students in all core areas allows for the calculation of teacher effectiveness in each of these subjects across grades three through high school. Second grade represents the baseline year of standardized achievement test scores. The standardized assessments are a mix of state criterion-referenced achievement tests and nationally norm-referenced assessments where there is no state achievement test (BFK, 2012a). In Project SOAR, the TerraNova standardized achievement test is the norm-referenced assessment used when no state achievement assessment exists (BFK, 2012a). Although VAM’s are still controversial, they do represent a promising method for estimating a teachers’ instructional effectiveness (Goe et al., 2008; Welsh, 2011).
**Student Linkage**

Linkage software that allows for more accuracy in linking students to appropriate teachers has developed over time (Goe et al., 2008). Battelle for Kids, a national, not-for-profit organization dedicated to improving instruction, has linkage software that is free to all districts in Ohio. Student linkage software accounts for detailed time of instruction, specifying time and percent of responsibility of the instruction of students by month from 20 to 100 percent during that timeframe (BFK, 2012a). It even addresses gaps due to maternity and medical leaves throughout the school year. The linkage software is beneficial to school districts and classrooms with high mobility rates because students who attend school for less than a full academic year count less in the analysis than their full-time counterparts (BFK, 2011). The linkage software enables teachers to share accurately the educational responsibility of their students. Each teacher must verify his or her student roster; then the building administrator approves each teacher roster to ensure all students are appropriately accounted. The linkage software is available to other states across the country and helps eliminate issues relating to teacher attribution from student linkage. Ohio’s future assessment system and Battelle for Kids linkage software help alleviate many of the concerns relating to attribution error.

**Mitigating Socio-economic Factors**

Another significant benefit of value-added data is that socio-economic factors do not influence value-added methodology, unlike achievement scores, making its use in accountability systems extremely advantageous (BFK, 2011; Kelly & Downey, 2010; Kupermintz, 2002; Sanders & Horn, 1998). Most school districts in major urban settings
have suffered from Ohio’s accountability system that places a substantial amount of weight on achievement measures (ODE, 2012a). It is common for districts with a high percentage of students from low socioeconomic families to have lower achievement scores than their higher wealth counterparts (ODE, 2012a). Incorporating value-added metrics into the state accountability system gives school districts an opportunity to show the academic progress occurring even though students may not yet be scoring at the proficient level (Buzick & Laitusis, 2010). Value-added metrics allow school districts with high percentages of students from low socioeconomic status to show parents and community members that academic progress is occurring (BFK, 2012b, Misco, 2008). Value-added data also ensure that the more affluent districts that have typically higher achieving students are also progressing at a yearly rate, responding to concerns that students in higher achieving districts are regressing to the mean over time (BFK, 2011). One possible reason why value-added data may not be influenced by socio-economic factors could be the multiple years of data used to create student predictions. Socio-economic status is a factor that remains constant over time and has shown to have little to no impact on student growth (BFK, 2012b). Based on the analysis of all school districts in the state of Ohio, 61.5 percent of the variance of school achievement can be explained by poverty level. This is based on the achievement and poverty data for the 2008 – 2009 school year. School district achievement is represented by the performance index score which is a weighted average of school achievement across all grades and subjects for all students. School district poverty level is represented by the free and reduced lunch percentages of each. Although there is a direct relationship between achievement and poverty levels when considering value-added gains against poverty level there is no
relationship, with only 2.3 percent of the variance of value-added gains explained by school poverty (BFK, 2012). So using value-added data is a fairer measure for school and district accountability because it is not affected by socio-economic status. Even though the use of teacher-level value-added data has some challenges, it is clear it has immense potential in improving instruction for students (Misco, 2008). The challenges include the use of teacher value-added data for promotion, employment, compensation, and assignment decisions (Hanushek & Hoxby, 2005; Kennedy, 2010; Rivkin, 2007).

**Class Size Addresses Some Measurement Error**

Although there is no way to completely account for all of the measurement error relating to standardized test scores, class size can help mitigate it in several ways. One form of measurement error is due to sampling error. This sampling error refers to the fact that a standardized test has limitations to the content it can assess (McCaffrey et al., 2004). As class size increases, McCaffrey et al. (2004) notes that the variability in an individual teacher effect due to sampling error will decrease (p. 106). Another means for addressing concerns relating to measurement error is through the calculation of standard error. The size of the standard error is influenced primarily by two factors:

- The first factor is the number of students in the classroom that included in the analysis. Typically, the greater the number of students in the classroom the smaller the standard error.
- Second is the variance of student scores. The more spread in the distribution of student gains in comparison to the classroom’s mean gain the larger the standard error (BFK, 2011, p. 8).
Teachers with larger class sizes will have more precise and accurate teacher effect designations (McCaffrey et al., 2004).

**Summary of Literature Review**

Chapter two included a review of literature related to various value-added methodologies for estimating teacher effectiveness. A brief overview of Ohio’s new evaluation system, which includes a value-added component, was shared. Teacher interview procedures typically include a multitude of teacher characteristics to identify the most qualified candidates. Over time, many school districts incorporated teacher screener instruments such as the TeacherInsight (TI) to help identify better the most qualified teachers from the candidate pool.

There are numerous caveats to consider when using value-added metrics to estimate the effectiveness of a teacher. Several of these factors were investigated including attribution error, test construction, measurement error, and non-random assignment of students to classrooms. Bearing in mind these caveats there is still considerable potential for the use of value-added measures to identify effective teachers. I investigated the various improvements to early value-added measures that lessen the initial concerns, including incorporation of student linkage software.

There are numerous screening instruments available to school districts to help identify the best teachers from the candidate pool. The TI purports to do this effectively. There is however, a cost associated with the use of the TI and not a great deal of research to support its claims. The study will use value-added measures to add to the literature
related to the TI score and teacher effectiveness. In the next chapter, I will outline and explain the methodology and research design for the study.
CHAPTER III

Research Methodology

The research methodology describes the design, procedures, and analysis needed to provide a reasonable response to the research questions and the problem statement (Creswell, 2008, p. 275). The purpose of this quantitative study focuses on the relationship between TeacherInsight (TI), a teacher-screener instrument, scores and eventual teacher growth index (TGI) scores. The research study explores how well the TI teacher-screener instrument accurately identifies teaching candidates who are able to facilitate a year’s worth of student growth. The chapter explains the research design of the study and includes a thorough description of the statistical techniques used to address each research question and the variables included. The chapter also shares demographic and performance data of the study school district and the teachers participating in the study to provide context and the rationale for choosing the particular school district. In addition, the section reviews the procedures and processes for data collection and obtaining approval from the study school district. Finally, a summary of the chapter is provided.

Purpose

The purpose of this study is to determine the accuracy of TI scores in identifying effective English/reading and mathematics teachers as reported by Teachers Connecting Achievement and Progress (TCAP) reports in a large, affluent, suburban school district in Ohio. This study may assist school districts through determining the predictive value of
this teacher-screening instrument. Teacher-screening instruments such as the TI have significant financial costs and school districts want assurances that the results warrant the expenditure. The investigation focuses on the relationship between the TI scores used in the teacher selection process in a Midwestern, suburban school district and teacher effectiveness in English/reading and mathematics in grades three through twelve as measured by TGI scores. In Ohio, the TGI score is reported to school districts via the TCAP reports provided by SAS®.

**Instrumentation and Variables**

During the district’s online application process, every teaching candidate must take the TI. Figure 3.1 depicts the procedure for a candidate moving through the application process.

![Figure 3.1. Study District Application Process](image-url)

Figure 3.1. Study District Application Process
The TI instrument did report scores as percentile ranks, but that changed in 2003. The TI instrument since the 2003 – 2004 school year bases scores on a percentage of points possible from 1 to 99 (Koerner, 2007). The research study only used TI scores from 2003 – 2004 to 2011 – 2012 school years. The study school district categorizes all TI scores into one of three groups based on these scores. Table 3.1 summarizes each level.

Table 3.1

Interview status based on TI level

<table>
<thead>
<tr>
<th>TI Level</th>
<th>TI Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Recommend for an interview</td>
<td>81 to 99</td>
</tr>
<tr>
<td>Recommend for an interview</td>
<td>72 to 80</td>
</tr>
<tr>
<td>Not Recommend for an interview</td>
<td>1 to 71</td>
</tr>
</tbody>
</table>

The TI scores are used to determine if applicants are good candidates for a specific teaching position. Applicants who are a highly recommend are placed on the “A” list of candidates. Applicants who score in the recommend level are placed on the “B” list. All other applicants are placed on the “C” list. Building administrators are able to interview applicants from any list, although they are encouraged by the human resource director to interview “A” and “B” applicants first. Due to the school districts’ student population growth it attracts an enormous volume of applicants. It would not be possible for building administrators to interview all of the applicants. The TI provides a means for screening applicants to move into the interview process. Some teaching positions, third through fifth grade, attract so many applicants that only category “A” candidates are interviewed. Some positions are so difficult to fill, high school physics or chemistry, that all applicants are interviewed regardless of the TI category.
**Teacher Growth Index**

The TCAP report contains the number of students included in the value-added analysis, standard error, teacher effect, subject, grade level of analysis, TGI, and an effectiveness rating. The teacher effect is the difference between the mean predicted score and mean observed score of the students in the analysis, minus an adjustment via a shrinkage estimator. The shrinkage estimation is included to help adjust for any measurement errors that may not have otherwise been taken into account. The shrinkage estimator pulls each teacher effect closer to zero and therefore makes it more difficult for a teacher to be inaccurately rated as most or least effective. Standard error provides a confidence band for the teacher effect. The TGI score is calculated by dividing the teacher effect by its corresponding standard error. The TGI score determines the teacher effectiveness rating. Table 3.2 identifies the five effectiveness ratings based on TGI scores (EVAAS, 2012). The middle three TCAP effectiveness ratings only differ by one standard error of measure. This provides only a 67 percent confidence band for each. The purpose of this difference is to help teachers identify whether they are trending toward most effective or least effective over time. For the purpose of the study, a slightly different effectiveness rating was used. Each effectiveness rating will include at least two standard errors of measure to ensure a 95 percent confidence band for each rating. The level of uncertainty is reduced by using two standard errors of measurement. The three middle TCAP effectiveness ratings are combined to create the expected rating for the study.
Table 3.2  
*Teacher effectiveness ratings*

<table>
<thead>
<tr>
<th>TGI</th>
<th>TCAP Effectiveness Rating</th>
<th>OTES Effectiveness Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A TGI score greater than positive 2</td>
<td>Most Effective</td>
<td>Above</td>
</tr>
<tr>
<td>A TGI score greater than positive 1 but less than positive 2.</td>
<td>Above Average</td>
<td>Expected</td>
</tr>
<tr>
<td>A TGI score between negative 1 and positive 1</td>
<td>Average</td>
<td>Expected</td>
</tr>
<tr>
<td>A TGI score less than negative 1, but greater than negative 2.</td>
<td>Approaching Average</td>
<td>Expected</td>
</tr>
<tr>
<td>A TGI score less than negative 2.</td>
<td>Least Effective</td>
<td>Below</td>
</tr>
</tbody>
</table>

The variables for the study are TI scores, TGI scores, gender, and years of experience. Table 3.3 provides the full name, description, abbreviation, and the type of measure for each variable. The TGI score is the dependent variable. TGI scores are a continuous variable created by dividing the teacher effect by its standard error. The TI score represents the independent variable. The TI score is a continuous variable on a scale from 1 to 99. Gender and years of experience are covariates of the study. Literature suggests that the more experience a teacher has the higher the quality of the teacher (Jacob et al., 2009; Koerner, 2007; Regan & Hayes, 2011). By controlling for years of experience and gender in the analysis, the researcher can determine how much of the variance is explained by TI scores alone.
Table 3.3
*Variables of study*

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>TeacherInsight™ Score</td>
<td>TI</td>
<td>A teacher-screening instrument</td>
<td>Continuous (1 - 99)</td>
</tr>
<tr>
<td>Teacher Growth Index</td>
<td>TGI</td>
<td>An estimate of teacher effectiveness</td>
<td>Continuous</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender</td>
<td>Sex of the participant</td>
<td>Categorical (F or M)</td>
</tr>
<tr>
<td>Years of Experience</td>
<td>YOE</td>
<td>Number of years a participant has taught</td>
<td>Continuous</td>
</tr>
</tbody>
</table>

The value-added data come from several standardized testing sources in Ohio. The study school district is a member of the SOAR project and Table 3.4 lists the various standardized assessments, grade levels, and subjects used in the analysis of the VAM. These assessments are both norm and criterion referenced assessments. All assessments are highly correlated with curricular objectives being taught in the classroom, which is critical for effective value-added analysis (BFK, 2011). The inclusion of assessments beyond the Ohio Achievement Assessments (OAA) allows for more teachers to be included in the value-added analysis.

Table 3.4
*Assessments for TCAP analysis*

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Type</th>
<th>Grades</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>TerraNova</td>
<td>TN</td>
<td>Norm</td>
<td>2 - 8</td>
<td>Reading &amp; mathematics</td>
</tr>
<tr>
<td>Ohio Achievement Assessment</td>
<td>OAA</td>
<td>Criterion</td>
<td>3 - 8</td>
<td>Reading &amp; mathematics</td>
</tr>
<tr>
<td>ACT Quality Core</td>
<td>ACT QC</td>
<td>Criterion</td>
<td>High school</td>
<td>Algebra I, Algebra II, Geometry, Pre-Calculus, &amp; English 9 – 12</td>
</tr>
</tbody>
</table>
Research Design

In this correlational study, a quantitative methodology was selected for the analysis of the data. The researcher used Hierarchical Multiple Regression (HMR) to explore the predictive value of TI scores on TGI scores of English/reading and mathematics teachers. HMR was selected because there are multiple independent variables (TI score, years of experience, and gender) and one dependent variable (TGI scores for English/reading and mathematics). HMR is based on correlation, but allows for a more sophisticated exploration of the interrelationship between TI scores, years of experience, gender, and TGI scores for English/reading and mathematics (Pallant, 2010, p. 148). HMR permits the investigation of how well the independent variables predict the TGI scores of English/reading and mathematics teachers. HMR makes the testing of each independent variable individually to determine how it contributes to the predictive ability of the model and control for each possible. The research model explores how well TI scores predict TGI scores in English/reading and mathematics (Block 2) after controlling first for years of experience and gender in Block 1. Gender and YOE are entered into block 1 together because they are both considered demographic variables. After all independent variables are entered, the overall model is assessed on its ability to predict the dependent variable, TGI scores for English/reading and mathematics. The relative contribution of each independent variable is also assessed. These statistical techniques were selected because of their appropriateness to the data gathered in the study. Listed below are the statistical techniques for each research question. Regression analysis is the best fit for the quantitative correlational study because the objective is to (a) identify any relationship existing between the TI scores, years of experience, and gender with TGI
scores for English/reading and mathematics and (b) determine the predictive value of TI scores to TGI scores of English/reading and mathematics.

**Procedures**

The data for the study were obtained from several databases within the school district and the Ohio Department of Education. There were several meetings with the data, research, and accountability (DRA) and human resource departments to identify the data necessary for analysis and determine the most efficient way to combine it. The DRA department prepared the data file with the necessary data. This required merging data from the DRA, human resource, and Ohio Department of Education databases. The study school district human resource department database contained the TI scores and gender for each teacher in the study. The DRA department database contained TGI scores, standard error, student count for the effectiveness report, year of the effectiveness report, teacher effectiveness rating, and subjects and grade level taught at the time of the TCAP report. The Ohio Department of Education database contains the years of experience for each teacher in the study. The teacher data were merged using teacher credential identification numbers, which are unique for every teacher in the entire state and included in each of the databases. The VLOOKUP function in Microsoft Excel 2010 was used to perform the merging of data. Once these data were merged on the teacher credential identification number, school district officials removed the identifying teacher information before providing me with the data set. Removing all personally identifiable information protects the identity of all of the teachers whose data are included in the study. Data collected for teachers in the study includes TI score levels and scores, years
of experience, gender, grade level and subject taught at the time of the teacher
effectiveness report, TGI, standard error, student count for effectiveness report, teacher
effectiveness rating, and year of report. Teachers without any of the necessary data points
were excluded from the final data set shared with the researcher by the school district.

Research Questions and Related Hypotheses

Three research questions were created in this study to review the relationship
between TI scores and teacher effectiveness. The research questions and hypotheses are
as follows:

1.1: Is there a relationship between the TI scores and teacher effectiveness – as a
proxy of student progress – as measured by the teacher growth index (TGI) score for
English/reading, controlling for teacher demographics (years of experience and gender)?

1.1Hₐ: Teachers with a higher level of TI scores will have a higher TGI score in
English/reading.

1.2. Is there a relationship between the TI score and teacher effectiveness as
measured by the TGI score for mathematics controlling for teacher demographics (years
of experience and gender)?

1.2Hₐ: Teachers with a higher level of TI scores will have a higher TGI score in
mathematics.

2. Does the TI score predict effective English/reading teachers better than
effective math teachers as measured by the TGI scores?

2Hₐ: The TI score will predict effective mathematics teachers better than it will
predict effective English/reading teachers.
Participants

The study school district was affluent with the highest median income in the state (ODE, 2012g). The school district has experienced high student population growth over the past ten years. During the years of this study, enrollment moved from 13,000 students to over 17,000 students (ODE, 2012a). During the time of the study this included roughly an additional 1,000 students per year from the 2007 – 2008 to 2011 -2012 school years (ODE, 2012a). The result of the student population growth required the school district to hire approximately 100 new teachers per year during the analysis. Each new teacher was required to take the TI online screener to move from the candidate pool into the interview process. If the teacher taught English/reading or mathematics in grades three through twelve then he or she would also have a TCAP report and be included in the analysis.

In order to reserve interviews for only the highest quality applicants, the school district began using the online TI screening instrument to select teaching applicants in 2004 (personal communication, Marsh G., January 6, 2013). The school district had teacher-level value-added reports that estimated teacher effectiveness in the 2007 – 2008 school year for teachers in reading and mathematics in grades three through eight. In the 2008 – 2009 school year, teacher-level value-added reports also became available for English and mathematics teachers in grades nine through twelve. The high school value-added reports include English 9 through 12, Algebra I, Geometry, Algebra II, and Pre-Calculus.

The study school district received the highest state rating of Excellent or Excellent with Distinction for each year of the data collection from 2007 through 2012. This rating was based on student achievement indicators for grades three through eight and ten,
student attendance rate, and high school graduation rate. The state of Ohio included a value-added composite metric to the accountability system that combined student growth in reading and math across grades three through eight in 2008 -2009 school year. The study school district maintained a year’s worth of growth or better for each of those school years (ODE, 2012g). The study school district’s high achievement and high academic growth led to numerous applicants for each teaching vacancy that arose.

**Sample Size**

The sample for the study includes English/reading and mathematics teachers who had teacher-level value-added data and a valid TI score. The student population growth the study school district experienced over the five-year period allowed for a robust sample of student data to create each unique TCAP report. There were 452 total teachers included in the sample of the study for English/reading and mathematics. This included 234 English/reading and 218 mathematics teachers in grades three through twelve. The 452 TCAP reports included the academic progress of over 14,000 different students. It is important to note that while there were 452 total different data points within the sample these do not reflect the total number of unique teachers in the sample. The difference is because there are many elementary teachers in grades three through five that teach both reading and mathematics. For the purpose of the study they were included in both the English/reading and mathematics analysis to better determine any relationship between the TI and TGI scores. The sample size is sufficient for a reliable equation of this social science research study based on Tabachnick and Fidell’s (2007) formula (p. 123). The formula takes into account the number of independent variables used in the study to
determine an appropriate sample size; \( N > 50 + 8m \); where \( N \) equals the recommended minimum number of the sample and \( m \) equals the number of independent variables. The study employs three independent variables (TI scores, years of experience, and gender). Using the sample size formula \( (N > 50 + 8(3)) \), the necessary minimum number of participants is 74. Each sample English/reading (234) and mathematics (218) are well above the minimum number recommended by Tabachnick and Fidell (2007). The participants for the study were selected because they began employment with the school district from the 2003-2004, the first year the TI online screener was used, through the 2011-2012 school year, had a TI score, and taught in a position that received an eventual TCAP report. All teachers who participated in this study were certified and/or licensed in the state of Ohio and provided direct instruction to the students included in their analysis.

**Research Approval**

The Office of the Superintendent of the study school district was contacted and provided with the prospectus of the research study. After the researcher met with the Office of the Superintendent to explain the purpose of the study and the data necessary, a meeting was set up to meet with the DRA department and human resource department. The director of the DRA department removed all teacher-identifiable information and replaced them with a unique teacher identification number that allows for analysis of data, but prevents the researcher from linking any data with a teacher. The director of curriculum and instruction provided final approval for use of the data for the study. A written letter of consent was obtained to use the district data for the research (see
Appendix B). The Ashland University Human Subjects Review Board (HSRB) received the school district approval with the HSRB consent form (see Appendix A).

**Analysis of Data**

The quantitative data were then presented using descriptive statistics including frequency, range, percentage, and mean (Creswell, 2008). Then preliminary analysis was performed using Pearson-Product correlation and independent samples t-tests to explore relationships and differences among the variables of the study. Hierarchical multiple regression (HMR) was selected as the main statistical method of the correlational research study. The goal was to answer the research questions which in summary inquired: the predictive value of TI scores on the TGI scores of English/reading and mathematics teachers; whether YOE and gender have a predictive value to TGI scores; and if TI scores predict TGI score of mathematics teachers better than English/reading teachers. The study school district merged data from all databases with Microsoft Excel 2010. The data was then imported into the IBM® SPSS® 19.0 statistical and data management package and coded for analysis.

To answer the research questions the following regression equations were used. The first step is written as

\[ TGI_R = \beta_0 + \beta_1 yoe + \beta_2 gender + e_i \]

and the second step,

\[ TGI_R = \beta_0 + \beta_1 yoe + \beta_2 gender + \beta_3 TI + e_i \]

where \( TGI_R \) = dependent variable prediction (in this case English/reading)

\( \beta_0 \) = the intercept, the value of TGI score when all the other variables are zero.
$\beta_1 \text{yoe} = \text{independent variable years of experience regression coefficient.}$

$\beta_2 \text{gender} = \text{independent variable gender regression coefficient.}$

$B_3 \text{TI} = \text{independent variable TI score regression coefficient.}$

$e_i = \text{represents the random error.}$

The following are the regression equations for mathematics. It is identical to the above equation with the exception it involves mathematics data. The first step is written as

$$TGI_M = \beta_0 + \beta_1 \text{yoe} + \beta_2 \text{gender} + e_i$$

and the second step,

$$TGI_M = \beta_0 + \beta_1 \text{yoe} + \beta_2 \text{gender} + \beta_3 \text{TI} + e_i$$

where $TGI_M = \text{dependent variable prediction (in this case mathematics)}$

$\beta_0 = \text{the intercept, the value of TGI score when all the other variables are zero.}$

$\beta_1 \text{yoe} = \text{independent variable years of experience regression coefficient.}$

$\beta_2 \text{gender} = \text{independent variable gender regression coefficient.}$

$B_3 \text{TI} = \text{independent variable TI score regression coefficient.}$

$e_i = \text{represents the random error.}$

This equation allows the effect of each independent variable to be assessed in blocks.

YOE and gender are considered to be of lesser importance and therefore will be entered in the first block. TI is expected to have the greatest predictive value of TGI scores so it will purposely be added last. Finally, summary statistics will be provided after the final block is analyzed. The F test describes the statistical significance of adding the various independent variables to the prediction equation sequentially. The test of the hypothesis is presented in the output as analysis of variance. The F ratio for mean square regression over mean square residual tests the significance of multiple R.
Question 1 explored the relationship between TI scores and TGI scores for English/reading and mathematics teachers. The dependent variable is the TGI score and the independent variable is the TI score. A Pearson bivariate analysis was performed to determine if there was a relationship between TI and TGI scores. There was a Pearson correlation coefficient calculated for English/reading and mathematics TGI scores separately.

Question 2 explored the how TI scores, YOE, and gender impacted the TGI scores of English/reading and mathematics. HMR was used to see how much of the variance of TGI scores could be explained by the various independent variables. TI scores, YOE, and gender were the independent variables and TGI scores for English/reading and mathematics teachers were the dependent variables. The independent variables were entered in blocks. The first block of independent variables included YOE and gender. The second block of data included TI scores as the independent variable.

**Summary**

Selecting and maintaining a highly effective teaching staff is the first priority of all school districts across the country. An effective teacher is able to facilitate maximum learning for every student. The charge of school district leaders is to provide an effective teacher for every classroom. The research study illuminated the importance of the recruitment process and value-added measures of effective teachers. Both of these topics deserve attention due to their extreme importance to the continuous improvement of student achievement.
The study used data to explore the relationship between TI scores, years of experience, gender, and TGI scores for teachers of English/reading and mathematics. If the TI scores accurately predict effective teachers as Gallup purports, there should be a direct correlation between a teacher’s TI score and his or her eventual effectiveness rating as measured by TCAP reports. If this is true, it could have a considerable impact on school district’s hiring practices. District staff not currently using the TI instrument could eliminate numerous hours of time and effort to identify the most qualified candidates in a more efficient manner. Chapter IV provides an examination of the results of the data analyzed in the research study.
CHAPTER IV

Results of Data Analysis

The purpose of the research study was to determine if the TeacherInsight (TI) score accurately predicts teacher effectiveness in English/reading and mathematics. The quantitative correlational study used teacher value-added data from a purposive sample of teachers from a large, Midwestern, suburban school district. The data included English/reading and mathematics teachers for grades three through twelve. The research questions explored the relationship between TI scores and teacher effectiveness in English/reading and mathematics as measured by the teacher growth index (TGI) score. This chapter provides a brief introduction; review of the research questions; descriptive statistics; and reports the findings of the statistical analyses of the research study.

Research Questions

The research questions in this study focused on determining the relationship between a teacher-screening tool and teacher effectiveness as measured by student academic growth. The specific teacher-screening tool used in this study was the TI created by The Gallup Organization and used throughout the United States. Teacher effectiveness was estimated by using a TGI score which was calculated by dividing a teacher’s effect score by his or her specific standard error from the TCAP reports. These reports are provided to teachers through EVAAS®.
1.1: Is there a relationship between the TI scores and teacher effectiveness – as a proxy of student progress – as measured by the teacher growth index (TGI) score for English/reading, controlling for teacher demographics (Years of experience and gender)?

1.1H₀: Teachers with a higher level of TI will have a higher TGI score in English/reading.

1.2. Is there a relationship between the TI score and teacher effectiveness as measured by the TGI score for mathematics controlling for teacher demographics (Years of experience and gender)?

1.2H₀: Teachers with a higher level of TI will have a higher TGI score in mathematics.

2. Does the TI score predict effective English/reading teachers better than effective math teachers as measured by the TGI score?

2H₀: The TI score will predict effective mathematics teachers better than it will predict effective English/reading teachers.

**Descriptive Statistics**

This section reports the descriptive statistics of the teacher data used in the research data. First, the descriptive statistics of the combined teacher dataset were examined. The study school district was a large, affluent, suburban school district located in the Midwestern United States. The combined dataset included teacher value-added data of English/reading and mathematics teachers in grades three through twelve from the 2007 – 2008 to 2011 – 2012 school years. The data were merged from databases from the DRA and human resource departments and Ohio Department of Education. Table 4.1
displays the frequency of the number of sampled teachers by gender for each subject. The total number of unique teachers for each subject area was higher than the number of teachers in the study, because most elementary teachers (grades three through five) taught both reading and mathematics. They were included in the analysis of each subject to determine the relationship of TI scores for each subject to the TGI score. No teacher data were counted more than once within each subject. Although many teachers had multiple TGI scores from various years, only the first TGI score they received was used in the analysis. This was to remain consistent with those staff that only had one year of TGI data. A more detailed description of the study school district can be found in Chapter III.

Independent Variables

There were three independent variables for the purpose of this research study (gender, years of experience, and TI score). Years of experience (YOE) and TI score were measured as a continuous variable and gender as a categorical variable. Teachers’ gender was coded as 0 for Female; and 1 for Male. Table 4.1 reports the frequency of the gender percentages and numbers for each subject. There were many more female than male teachers in the study for both English/reading and mathematics. Overall there were roughly the same number of English/reading (234) and mathematics (218) teachers. The percentage of female teachers for English/reading (84.2%) was very consistent with the percentage of female teachers for mathematics (83.9%).
Table 4.1  
*Frequency: Number of teachers by gender*

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>183</td>
<td>83.9%</td>
</tr>
<tr>
<td>Male</td>
<td>35</td>
<td>16.1%</td>
</tr>
<tr>
<td>Total</td>
<td>218</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English/Reading</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>197</td>
<td>84.2%</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>15.8%</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>100%</td>
</tr>
</tbody>
</table>

The descriptive statistics revealed that the mean YOE of the 234 English/reading teachers sampled was 8.38 years (SD = 5.743). The 218 mathematics teachers had a similar mean YOE of 8.34 (SD = 5.832). In addition, the skewness of 1.145 (SE = 0.159) and kurtosis of 1.137 (SE = 0.317) suggested that YOE was not normally distributed as seen in Table 4.2. Skewness refers to the symmetry of the distribution. A normal distribution has a mean that is located in the center of the distribution (Tabachnick & Fidell, 2007). The sample YOE mean for English/reading teachers is positively skewed; indicating the mean score is located to the far left of the distribution. Kurtosis refers to the peakedness of a distribution. The sample mean YOE for English/reading teachers has a positive kurtosis of 1.137 (SE = 0.317), meaning there is a narrow peak to the distribution. Dividing the skewness or kurtosis by their corresponding standard error is a method for determining if they are normally distributed. A value of zero suggests a perfectly normal distribution (Tabachnick & Fidell, 2007). If the values are close to zero, a distribution has not violated the assumptions of normality.

Tabachnick and Fidell, suggested using transformations to improve the normality of the variable (2007, p. 87). To address the skewness and kurtosis issues a square root transformation was performed. This is accomplished by taking the square root of the
YOE variable for every participant. After the transformation was performed, the skewness of .392 ($SE = .159$) and kurtosis of – .326 ($SE = .317$) gave evidence it was normally distributed. With large samples over 200, there is a diminished impact of skewness and kurtosis values slightly greater or less than zero (Tabachnick & Fidell, 2007). For the remainder of the study YOE will refer to the new square root transformation of the YOE statistic. Using the square root of a variable does not affect the results of the regression analysis (Tabachnick & Fidell, 2007).

The mean TI scores were consistent for English/reading at 71.58 (SD = 8.688) and mathematics teachers at 71.12 (SD = 8.611), as noted in Table 4.2 and 4.3 respectively. The skewness of .080 ($SE = .159$) for English/reading teachers and – .037 ($SE = .328$) of mathematics teachers were both normally distributed. The kurtosis of .239 ($SE = .317$) for English/reading teachers and .484 ($SE = .328$) for mathematics teachers is also considered normally distributed. The variable required no transformation as a result.

Table 4.2
Descriptive Statistics for English/reading Teachers

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>Skewness ($SE$)</th>
<th>Kurtosis ($SE$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI Score</td>
<td>234</td>
<td>71.58</td>
<td>8.688</td>
<td>.080 (.159)</td>
<td>.239 (.317)</td>
</tr>
<tr>
<td>YOE</td>
<td>234</td>
<td>8.38</td>
<td>5.743</td>
<td>1.145 (.159)</td>
<td>1.137 (.317)</td>
</tr>
<tr>
<td>YOE*</td>
<td>234</td>
<td>2.73</td>
<td>.965</td>
<td>.392 (.159)</td>
<td>-.326 (.317)</td>
</tr>
</tbody>
</table>

*After square root transformation
Table 4.3
Descriptive Statistics for Mathematics Teachers

<table>
<thead>
<tr>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Skewness (SE)</th>
<th>Kurtosis (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI Score</td>
<td>218</td>
<td>71.12</td>
<td>8.611</td>
<td>-.037 (.165)</td>
</tr>
<tr>
<td>YOE</td>
<td>218</td>
<td>8.34</td>
<td>5.832</td>
<td>1.125 (.165)</td>
</tr>
<tr>
<td>YOE*</td>
<td>218</td>
<td>2.71</td>
<td>.991</td>
<td>.334 (.165)</td>
</tr>
</tbody>
</table>

*After square root transformation

Dependent Variable

The dataset used in the research study contained 452 unique TGI scores for English/reading and mathematics teachers from the 2007 – 2008 to 2011 – 2012 school years who also had a valid TI score. This included 218 unique mathematics teachers and 234 English/reading teachers. Table 4.4 displays the descriptive statistics of the TGI score for English/reading and mathematics teachers. The positive mean TGI score for both English/reading and mathematics teachers indicating that on average teachers in the sample were effective. The mean TGI score for English/reading teachers was .202 ($SD = 1.820$), and for mathematics teachers was .537 ($SD = 2.395$). An independent samples $t$-test suggested that the mean TGI scores for English/reading teachers and mathematics teachers were not significantly different ($t = 1.618, p = .106$).

Table 4.4
Descriptive Statistics for TGI Score of English/reading and Math Teachers

<table>
<thead>
<tr>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Skewness (SE)</th>
<th>Kurtosis (SE)</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>234</td>
<td>.212</td>
<td>1.820</td>
<td>-.046 (.159)</td>
<td>2.580 (.317)</td>
<td>1.618</td>
</tr>
<tr>
<td>Math</td>
<td>218</td>
<td>.537</td>
<td>2.395</td>
<td>1.170 (.165)</td>
<td>2.707 (.328)</td>
<td></td>
</tr>
</tbody>
</table>
Data Analysis

This section reports the statistical correlations from the analysis of the collected teacher data. The first section is for preliminary analyses to explore relationships between the variables of the study. The second section reports the analysis of the research questions 1.1 and 1.2. Finally, the last section reviews the data for the second research question. The statistical analyses were described centered on the three research questions presented in the study at the 90 percent confidence level ($\alpha = .10$). This level was selected because the research is exploratory in nature. Using this level only allowed for a 10 percent probability that the results were due solely to chance.

Preliminary Analysis

Preliminary analyses were performed to better understand the relationships between the three continuous variables (YOE, TI score, and TGI score) and one categorical variable (gender) of the study. These analyses do not attempt to answer the questions, but rather provide some perspective to better understand the big picture of these variables. Table 4.5 and 4.6 show the results of the Pearson correlation of the three continuous variables for English/reading and mathematics teachers respectively.

The relationship between TI scores and teacher effectiveness in English/reading (as measured by the TGI score) was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. There was a weak positive correlation between the two variables, ($r = .113, p = .085$), with high scores on the TI
screening instrument associated with high English/reading teacher effectiveness scores, though it was so close to zero that it has not practical significance.

The relationship between TI scores and teacher effectiveness in mathematics (as measured by the TGI score) was investigated using Pearson product-moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. There was no relationship discovered between the TI and TGI scores.

There are no significant relationships between any of the continuous variables for mathematics teachers. For English/reading teachers there is only a significant relationship between TI score and TGI score (R = .113, p = .085). In the correlational model, TI scores were hypothesized as an accurate predictor of teacher effectiveness (TGI) for English/reading and mathematics teachers. The results do not support the hypothesis that TI scores accurately predict the TGI score of English/reading or mathematics teachers.

Table 4.5  
*significant at α = 0.1

Table 4.6  

To explore the effect of gender on the continuous variables an independent samples t-test was performed. Table 4.7 and 4.8 report the results of this analysis. On
average, female teachers in the study had more experience than male teachers. It is important to note that many more teachers in the study were female than male for both English/reading and mathematics. The results in Table 4.7 show no difference in the average TI score between males (71.65, SD = 9.047) and females (71.56, SD = 8.643) for English/reading teachers. Male mathematics teachers (71.89, SD = 7.15) scored almost an entire point higher than their female counterparts (70.98, SD = 8.87) on the TI. Overall in the study, there were slightly more English/reading teachers (234) than mathematics teachers (218). Table 4.7 and 4.8 show there was a significant difference in the YOE for male and female teachers in the study for both English/reading (t = 4.10, p < .001) and mathematics teachers (t = 3.54, p < .001). There were no significant differences by gender on mean TI or TGI scores for English/reading teachers. There was also not a significant difference between the mean TI scores of male and female mathematics teachers. There was a significant difference between the TGI score of male and female teachers (t = – 0.239, p = .004). Male mathematics teacher’s TGI score is significantly higher than female teachers suggesting that male teachers were significantly more effective at facilitating student academic progress than their female counterparts. However, both male and female mathematics teachers’ average TGI scores were above zero, indicating that both groups were effective.
Table 4.7  
**Independent Samples t-test for English/reading Teachers**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>37</td>
<td>71.65</td>
<td>9.047</td>
<td>-0.053</td>
<td>.958</td>
</tr>
<tr>
<td>F</td>
<td>197</td>
<td>71.56</td>
<td>8.643</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TGI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>37</td>
<td>.033</td>
<td>2.13</td>
<td>.571</td>
<td>.571</td>
</tr>
<tr>
<td>F</td>
<td>197</td>
<td>.245</td>
<td>1.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>YOE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>37</td>
<td>2.19</td>
<td>.856</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>197</td>
<td>2.83</td>
<td>.952</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant at α = 0.01

Table 4.8  
**Independent Samples t-test for Mathematics Teachers**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>35</td>
<td>71.89</td>
<td>7.15</td>
<td>-5.70</td>
<td>.569</td>
</tr>
<tr>
<td>F</td>
<td>183</td>
<td>70.98</td>
<td>8.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TGI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>35</td>
<td>1.41</td>
<td>3.14</td>
<td>-2.39</td>
<td>.004*</td>
</tr>
<tr>
<td>F</td>
<td>183</td>
<td>.370</td>
<td>2.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>YOE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>35</td>
<td>2.18</td>
<td>.878</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>183</td>
<td>2.81</td>
<td>.982</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant at α = 0.1

**Question 1.1**

Question 1.1 asked: Is there a relationship between the TI scores and teacher effectiveness – as a proxy of student progress – as measured by the teacher growth index (TGI) score for English/reading, controlling for teacher demographic factors (YOE and gender)? This question was analyzed by the use of a hierarchical multiple regression in SPSS. Gender, YOE, TI and TGI variables were used to address this question. As shown in Table 4.9, the following hypotheses were answered based on the results:
Reject $1.1H_a$: Teachers with a higher level of TI will have a higher TGI score in English/reading after controlling for the demographic variables of YOE and gender. The hypothesis is rejected because while there is statistical significance the percent of variance that can be explained by the TI score is relatively zero and therefore has no practical significance.

Table 4.9 reports the model summary of the hierarchical multiple regression including the R-square, standard error of the estimate, and F change for Block 1 and Block 2 of English/reading teachers. The models are evaluated to determine how much of the variance of TGI score is explained by TI scores after the effects of years of experience and gender are removed. Then, each of the three independent variables (TI score, YOE, and gender) is evaluated to determine how well each variable contributes to the final equation. Table 4.2 and 4.3 display the average score, standard deviation, and sample size of each data set by subject.

Table 4.9 shows that Block 1, containing YOE and gender did not predict a teacher’s TGI score. The R-square value is .002, indicating that only 0.2 percent of the variance of TGI scores for English/reading teachers were explained by teacher demographic factors. The inclusion of the TI score in Block 2 significantly improved the model versus using the demographics of YOE and gender alone ($F = 2.972; df = 1, 230; p = .086$). The R-square for Block 2 of .015 indicates that 1.5 percent of the variance of TGI scores can be explained by the TI score. The part or semi-partial correlation explains the proportion of the remaining variance a variable can explain after controlling for other variables in the model (Tabachnick & Fidell, 2007, p. 145). After controlling for gender and YOE, 1.3 percent of the variance can be explained by the TI score uniquely.
Table 4.9
Regression results for English/reading teachers

<table>
<thead>
<tr>
<th>Block</th>
<th>Predictors</th>
<th>R</th>
<th>( R^2 )</th>
<th>( R^2 )-change</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender, YOE</td>
<td>.043</td>
<td>.002</td>
<td>.002</td>
<td>.217</td>
<td>2</td>
<td>231</td>
<td>.805</td>
</tr>
<tr>
<td>2</td>
<td>Gender, YOE, TI</td>
<td>.121</td>
<td>.015</td>
<td>.013</td>
<td>2.97</td>
<td>1</td>
<td>230</td>
<td>.086*</td>
</tr>
</tbody>
</table>

Coefficients from Block 2

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficients</th>
<th>t</th>
<th>p value</th>
<th>Part correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOE</td>
<td>.010</td>
<td>.075</td>
<td>.941</td>
<td>.005</td>
</tr>
<tr>
<td>Gender</td>
<td>-.209</td>
<td>-.621</td>
<td>.535</td>
<td>-.041</td>
</tr>
<tr>
<td>TI</td>
<td>.024</td>
<td>1.724</td>
<td>.086*</td>
<td>.113</td>
</tr>
</tbody>
</table>

*significant at \( \alpha = 0.1 \)

**Question 1.2**

Question 1.2 asked: Is there a relationship between the TI score and teacher effectiveness as measured by the TGI score for mathematics after controlling for the teacher demographics of YOE and gender? This question was analyzed by the use of HMR and the results are reported in Table 4.10. HMR was used to assess the ability of the TI score to accurately predict teacher effectiveness, after controlling for the influence of YOE and gender. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity. YOE and gender were entered in Block 1 first as a means of controlling for the demographic variables. The R-square of Block 1 was .029; meaning 2.9 percent of the variance of the TGI scores of mathematics teachers was explained by the demographic factors of YOE and gender. Block 2 allowed for the inclusion of the TI score, but it did not significantly improve the model from Block 1. The R-square of Block 2 was .030 with only a change
of .001. So TI scores are not a significant predictor of TGI scores for mathematics teachers. Gender however is a significant predictor of TGI scores for mathematics teachers (t = 2.51, p = .013). For mathematics teachers gender uniquely explains 2.9 percent of the variance in TGI scores, as noted by the part correlation in Table 4.10. The part correlation (also known as semi-partial correlation) between gender and the TGI score is .169. The square of .169 is .029, which means gender uniquely explains 2.9 percent of the variance in the TGI score, after controlling for YOE and the TI score.

The following hypotheses were answered based on the previous results:

Reject 1.2H₀: Teachers with a higher TI score will have a higher TGI score in mathematics.

Table 4.10

<table>
<thead>
<tr>
<th>Block</th>
<th>Predictors</th>
<th>R</th>
<th>R²</th>
<th>R²-change</th>
<th>F change</th>
<th>df1</th>
<th>df2</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender, YOE</td>
<td>.171</td>
<td>.029</td>
<td>.029</td>
<td>3.24</td>
<td>2</td>
<td>215</td>
<td>.041*</td>
</tr>
<tr>
<td>2</td>
<td>Gender, YOE, TI</td>
<td>.175</td>
<td>.030</td>
<td>.001</td>
<td>.265</td>
<td>1</td>
<td>214</td>
<td>.607</td>
</tr>
</tbody>
</table>

Coefficients from Block 2

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficients</th>
<th>t</th>
<th>p value</th>
<th>Part correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOE</td>
<td>.152</td>
<td>.911</td>
<td>.363</td>
<td>.061</td>
</tr>
<tr>
<td>Gender</td>
<td>1.130</td>
<td>2.51</td>
<td>.013*</td>
<td>.169</td>
</tr>
<tr>
<td>TI</td>
<td>.010</td>
<td>.515</td>
<td>.607</td>
<td>.035</td>
</tr>
</tbody>
</table>

*significant at α = 0.1

Question 2

Question 2 asked: Does the TI score predict effective English/reading teachers better than effective math teachers as measured by the TGI score?
The following hypotheses were answered based on the previous results:

Reject $2H_a$: The TI score will predict effective mathematics teachers better than it will predict effective English/reading teachers.

This question was analyzed by the use of HMR. The TI score did not predict effective mathematics teachers as there was no significant relationship. There was a statistically significant relationship between the TI and TGI score of English/reading teachers, but it has no practical significance. The R-square of the model for English/reading teachers was .015 meaning that 1.5 percent of the variance of TGI scores can be explained by the three variables (TI score, YOE, and gender). However, the TI score was responsible for explaining 1.3 percent of the variance of the TGI score. The mathematics model R-square is actually .030 which is twice that of the model for English/reading teachers. However, this is with the inclusion of all three variables, not just the TI score. Gender is the only variable in the model that is a significant predictor of TGI score for mathematics ($t = 2.51, p = .013$).

Table 4.11
Block 2 Summary Results for English/reading and Mathematics Teachers

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$R^2$</th>
<th>$F$-change</th>
<th>$Df$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English/reading</td>
<td>.121</td>
<td>.015</td>
<td>2.92</td>
<td>1, 230</td>
<td>.086*</td>
</tr>
<tr>
<td>Mathematics</td>
<td>.175</td>
<td>.030</td>
<td>.265</td>
<td>1, 214</td>
<td>.607</td>
</tr>
</tbody>
</table>

*significant at $\alpha = 0.1$

**Summary of Findings**

The data provided results for the research questions of the study. Question 1.1 examined the relationship between TI scores and TGI scores for English/reading teachers after controlling for the demographic variables of YOE and gender. There was only a
weak relationship discovered between TI scores and TGI scores of English/reading teachers, explaining only 1.3 percent of the variance. There was no significant relationship between demographics and TGI scores.

Question 1.2 examined the relationship between TI scores and TGI scores for mathematics teachers after controlling for the demographic variables of YOE and gender. There was no significant relationship discovered between TI scores or the demographic variable of YOE and TGI scores of mathematics teachers. There was a weak relationship between gender and TGI scores. Gender uniquely explains 2.9 percent of the variance in the TGI score of mathematics teachers, as noted in Table 4.10.

Question 2 explored whether TI scores explained the TGI score better for English/reading or mathematics teachers. Although only English/reading teachers TI scores were significantly related to TGI scores, it was a weak relationship. The TI score for English/reading teachers only explained 1.3 percent of the variance in the TGI score uniquely.

The next chapter will discuss, based on the results from the analyses, the critical findings of this study, limitations and delimitation, practical implications for the use of TI scores as a teacher-screening instrument, and recommendations for future research.
CHAPTER V

Discussion

This chapter reports the results of the statistical analyses of the study. The content of this chapter includes an introduction, review of the purpose of the study, a discussion of findings for each research question, limitations and delimitations of the study, implications for school districts using the TI screening instrument, recommendations for future research, and conclusions.

School accountability has increased dramatically over the past ten years with the creation of proficiency testing, local report cards, and value-added reporting. All of these tools can be used as a means for improving instruction for students in a school district, building, or specific classroom. Unfortunately, in some cases they are used to name, blame, and shame school districts rather than a means for continuous improvement. The inclusion of the performance index rankings on the local report card, a system used to rank school districts from highest to lowest in regard to their testing performance is the most recent example. Now there are efforts across the country to quantify the overall effectiveness of individual teachers on improving student achievement. This pressure from legislators, state departments of education, parents, and community members for immediate improvement has forced school districts to review their hiring practices.

There is little question that teachers matter and high-quality teachers matter most. Teacher screening instruments were advertised as a method to ensure school districts hired the most capable teachers considering specific characteristics that would translate into superior student academic performance. Many districts currently use the
TeacherInsight (TI) and other teacher screening instruments to ensure they hire teachers who can immediately improve instruction, increase student learning, and therefore produce positive results on these new accountability measures.

At the time of this research, the study school district used the TI score as a means to identify high-quality teachers for the interview stage of their application process as noted in Figure 5.1. The study used teacher data from the 2007 – 2008 to 2011 – 2012 school years. The study included data for 234 English/reading and 218 mathematics teachers. During the time of the study, the school district was experiencing an increase in student population that led to the need to hire approximately 100 additional teachers each year during that five-year period.

The focus of this study was to investigate the predictive value of the TI score for identifying the most effective teachers as noted by TGI scores in English/reading and mathematics. Specifically, the study inspected: (a) the relationship between TI scores and mathematics teacher effectiveness, (b) the relationship between TI scores and English/reading teacher effectiveness, and (c) if the TI score predicts effective mathematics teachers better than English/reading teachers.

The study also examined if the demographic factors of years of experience (YOE) or gender had any influence on the effectiveness for English/reading or mathematics teachers. If the TI screening instrument successfully predicts effective teachers in English/reading and mathematics, then school districts can use this information to narrow the number of candidates for interviews in the hiring process. Candidate interviews are the most time-consuming step in the hiring process and as such school districts are constantly looking for mechanisms to make the process more efficient.
In order to answer these research questions, data were collected from several databases: (a) TI score from the school district’s human resources database, (b) YOE and gender from the state department of education database, and (c) TGI score from the school district’s data, research, and accountability database. These data were combined from the various databases using the VLOOKUP function in Microsoft Excel 2010. There was preliminary data analysis using descriptive statistics, correlation, and independent sample t-tests. Finally, research questions were explored using a hierarchical multiple regression model.

Figure 5.1. Study District Application Process
**TI Scores as a Predictor of Teacher Effectiveness**

Literature connecting teacher effectiveness and the TI score is limited. Most of the available literature comes directly from The Gallup Organization, creator of the TI screening instrument. With very little third-party research available, it is difficult to determine the accuracy of the results (The Gallup Organization, 2007; The Gallup Organization, 2011). These studies report a positive relationship between the TI score and indicators of teacher effectiveness. There is much more literature in relation to value-added data, but most of it is at the school or grade-level and not at the classroom or teacher level. Teacher-level value-added data are very new, and there should be additional literature available in the future. The 2012 – 2013 school year will be the first year that all grade four through eight English/reading and mathematics public school teachers in the state of Ohio will receive classroom level value-added data. This study used independent research as well as studies developed by The Gallup Organization to examine the relationships between TI scores and teacher effectiveness, as measured by TGI scores. The results of this study address the gap in literature regarding the TI screening instrument and teacher-level value-added data, as measured by the TGI score for English/reading and mathematics.

**Summary of Results for Research Question 1.1**

Is there a relationship between the TI scores and English/reading teacher effectiveness, as measured by TGI scores, after controlling for years of experience and gender?
There were 234 unique teacher data points available for the analysis of this research question. Each English/reading teacher included in the analysis of this question included a valid TI score, TGI score, YOE, and gender. Any teacher in the data set without one or more of these variables was excluded from the analysis. Preliminary data analysis included correlations between the three continuous variables (YOE, TI score and TGI score). The preliminary analysis found a weak positive correlation between TI and TGI scores for English/reading teachers. Although the relationship between the TI score and TGI score was significant ($r = .113, p = .085$), it does not have practical significance. The TI score can only explain 1.3 percent of the variance between the two variables, which is so close to zero it may not warrant the expense of the screening tool. There was no significant relationship between YOE with either the TI or TGI scores for English/reading teachers.

An independent samples t-test was performed to identify any significant differences between gender, a categorical variable, and the continuous variables of YOE, TI score, and TGI score of English/reading teachers. There were no significant differences between the TI and TGI scores of English/reading teachers based on gender. Both groups of English/reading teachers on average are effective teachers based on the average positive TGI scores. It may be possible that other factors are at work with the average teacher having a positive TGI score. Some of these factors could be the induction program of the school district, professional development opportunities offered, among other factors. There was a significant difference between the YOE of male and female English/reading teachers, with females having a significantly higher mean YOE than males.
A hierarchical multiple regression was performed to determine if there was a significant relationship between the TI score and English/reading teacher effectiveness, as measured by the TGI score, while controlling for the demographic factors of YOE and gender. YOE and gender were entered into the model first in Block 1 as control variables. Neither was a good predictor of the TGI score for English/reading teachers. The addition of the TI score in Block 2 did significantly improve the overall model ($F = 2.972; df = 1,230; p = .086$). However, the addition of the TI score only explains 1.5 percent of the variance of the TGI scores for English/reading teachers. So although this result is statistically significant, it does not have much practical value for school districts. Looking at the part or semi-partial correlation, the TI score only explains 1.3 percent of the TGI scores for English/reading teachers uniquely after removing the effects of YOE and gender. Taking these analyses into consideration it does not appear that the TI score alone is extremely effective at identifying English/reading teachers who will have high TGI scores.

**Summary of Results for Research Question 1.2**

Is there a relationship between the TI scores and mathematics teacher effectiveness, as measured by TGI scores, after controlling for years of experience and gender?

There were 218 unique teacher data points available for the analyses of the research question. Each mathematics teacher included in the analyses of this question included a valid TI score, TGI score, YOE, and gender. Any teacher in the data set without one or more of these variables was excluded from the analyses. Preliminary data
analysis included correlations between the three continuous variables (YOE, TI score and TGI score). The analysis found no significant relationship between TI and TGI scores for mathematics teachers. There was also no significant relationship between YOE with either the TI or TGI scores for mathematics teachers.

An independent samples t-test was performed to identify any significant differences between gender, a categorical variable, and the continuous variables of YOE, TI score and TGI score of mathematics teachers. Male mathematics teachers scored almost one full point higher than their female counterparts on the TI screening instrument. While this is not a significant difference, this could have practical significance for the school district when it comes to the hiring process. In the hiring process, candidates who score higher on the TI screening instrument have a better chance of obtaining an interview and therefore being hired. So TI scores that are on the fringe of one of the three categories the study school district uses, as noted in Table 3.1, could have an impact on obtaining an interview. For instance, if the candidates’ TI score places them into category A versus B of the applicant pool or category B versus C the candidate would have a better chance of obtaining a position because principals interview candidate A applicants prior to category B and rarely interview candidate C applicants. However, from the five years of data there were far more female mathematics teachers (183) than males (35). So this does not appear to have had an effect on the balance of gender in favor of males for mathematics positions in the school district. The question is whether the school district receives equal amounts of male and female applicants. If not, it could be possible that male mathematics teachers are hired at a higher percentage than females simply because they scored higher on the TI screening instrument.
Table 3.1

<table>
<thead>
<tr>
<th>Interview Status Based on TI Level</th>
<th>TI Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Recommend for an interview</td>
<td>81 to 99</td>
</tr>
<tr>
<td>Recommend for an interview</td>
<td>72 to 80</td>
</tr>
<tr>
<td>Not Recommend for an interview</td>
<td>1 to 71</td>
</tr>
</tbody>
</table>

There is also a significant difference between the mean TGI scores of male and female mathematics teachers. Male mathematics teachers have a significantly higher mean TGI score than females, suggesting that they are more effective teachers. However, it is important to note that both groups have average scores that are positive, meaning that both groups are effective. Males are more effective with a TGI over one point higher than females. This could be very practical for school districts because having a higher TGI score could allow a school district to have a higher overall rating on the Local Report Card. School districts that are able to facilitate above expected growth in the aggregate across grade levels and subjects are eligible for the highest rating of Excellent with Distinction, by the State of Ohio (ODE, 2012a). Just as with English/reading teachers, there was a significant difference between the YOE of male and female mathematics teachers, with females having a significantly higher mean YOE than males. There appear to be no practical implications in relation to YOE based on gender as males are obtaining significantly better TGI scores than females.

A hierarchical multiple regression was performed to determine if there was a significant relationship between the TI score and mathematics teacher effectiveness, as measured by the TGI score, while controlling for the demographic factors of YOE and gender. YOE and gender were entered into the model first in Block 1 as control variables. The model for Block 1 was significant \( (F = 3.24; \ df = 2,215; \ p = .041) \) suggesting it was
a good predictor of the TGI score for mathematics teachers versus random selection. The
addition of the TI score in Block 2 did not significantly improve the overall model. This
tells us that the TI score does not accurately predict the TGI score of mathematics
teachers. Reviewing Block 1 variables, it does appear that gender is a significant
predictor of the TGI score for mathematics teachers ($t = 2.51, p = .013$). Although gender
is a significant predictor, it is a weak relationship and therefore not practical.

Summary of Results for Research Question 2

Does the TI score predict effective English/reading teachers better than
mathematics teachers as measured by TGI scores?

The analysis of data found that the TI score of English/reading teachers had a
weak positive relationship with the TGI score. There was however, no significant
relationship between the TI score of mathematics teachers and their corresponding TGI
score. The model for English/reading teachers could explain only 1.5 percent of the
variance among TGI scores, with the TI score being solely responsible for explain 1.3
percent. The regression model for mathematics was actually twice as effective at
predicting the TGI score as that of English/reading teachers. The mathematics model
could explain 3.0 percent of the variance of the TGI scores versus only 1.5 percent by the
English/reading model. The mathematics model includes all three variables (YOE,
gender, and TI score) with the TI score not significantly helping the model. The
mathematics model depends solely on the significance of gender.

Although both models are significant, only in the English/reading model is the TI
scores a significant predictor of the TGI scores. Both models have very weak
relationships explaining only a small percentage of the variance between TGI scores. With such a small percentage of the variance between TGI scores being explained by either model, there is little practical significance to school districts. It would be wise to continue reviewing other factors that might better predict the TGI scores of teaching candidates in the applicant pool. Therefore the TI score (for English/reading teachers) and gender (for mathematics teachers) do not predict a large enough percentage of the TGI score to be recommended for use in the application process as a means of selecting highly-effective teaching candidates to move to the interview stage.

**Recommendations and Implications**

This study examined the relationship between the TI score and student academic growth as measured by the TGI scores for English/reading and mathematics teachers. The demographic factors of YOE and gender were also examined to determine if they had any relationship to TGI scores. There are numerous studies that could be conducted using these and other data from additional school districts. Teacher screening instruments have the potential to save human resource department’s precious time and effort sifting through hundreds of applications and resumes of candidates in a short period of time. Based on the results of this study, there are also several implications for school districts that currently use the TI screening instrument. Limitations and delimitations exist for every research study. It is important to make note of each to help guide future research studies.
Limitations

There were several limitations of the study identified. First, the analyses only include data from one large affluent, suburban, public school district in central Ohio. The study only utilizes data from teachers who taught English/reading and mathematics courses in grades three through eight and certain high school courses and therefore cannot be generalized to other subjects. It is possible that the TI is identifying talent that cannot be measured accurately by the TGI score. The effectiveness of a teacher can be defined in numerous ways beyond just a TGI score. The OAA are based on a minimum competency of the academic content standards. It is also conceivable that the TI and TGI scores are measuring two completely different things.

Additional limitations based on the design of the study may also exist. The researcher did not expect gender to be a significant factor for the TGI score of either subject. Why gender became a significant factor for math (male teachers have a higher TGI score than female teachers) is troubling. It might be possible that “gender” was confounded with the subject or grade span they taught. It might also be possible that that were more male math teachers at middle or high school where the TGI score was calculated from end of course exams. The high school exams are created by the ACT organization as opposed to the Ohio Achievement Assessment (OAA) which is created by Pearson (ODE, 2012c). The end of course exams could have a different pattern of TGI score than the OAA’s at the lower grade levels. The study did not collect the grade level taught for each mathematics teacher. If there were more female teachers at the lower grade levels and more male teachers at the higher grade levels that might be confounding the significance of gender.
Delimitations

There are also several delimitations of the study to consider. First, the teacher sample is not random because the study school district uses the TI scores to identify high-quality teaching candidates for the interview stage in the hiring process as noted in Figure 5.1. It is highly unlikely that many teachers who scored low on the TI screening instrument were able to obtain an interview and eventually be hired by the school district. Also, teachers who had been hired in the district prior to the 2003 – 2004 school year did not take the TI screening instrument and therefore are not included in the analyses, even if they have a valid TGI score. Finally, the study school district only utilizes the TI screening instrument for their application process, which is one of several teacher-screening instruments available for school districts to purchase. Any relationship that does or does not exist between the TI scores and TGI scores does not address any relationships that might exist with other teacher screening instruments. These delimitations provide a boundary for the analyses of the data.

Implications for the use of the TI screening instrument

The results of the study reveal no practical relationship between the TI score and TGI scores for English/reading and mathematics teachers. This does not mean that nothing of importance can be determined from these results. Districts will want to be cautious using the TI solely to screen teaching candidates. If the TI does not accurately predict the TGI score in any practical manner for English/reading or mathematics teachers, then it may not be an effective tool for screening teaching candidates. It is possible that school districts might inadvertently identify the wrong candidates for the
interview stage of their hiring process. Before a school district begins or continues using
the TI screening instrument for the purpose of identifying high quality teaching
candidates from the applicant pool financial, personal, and ethical implications must be
considered.

**Ethical implications of study.** There are several ethical implications identified
by the study that districts should consider with regard to the TI and TGI scores. First, the
definition of an effective teacher used in the study may be too narrow. Only the TGI
score of English/reading and mathematics teachers were used to determine teacher
effectiveness. A TGI score is derived from dividing the teacher effect by its standard
error. The TGI score is based on standardized achievement tests that only sample the
taught curriculum. Most of the test items on the OAA are multiple choice questions
which are of lower complexity than performance-based assessments. There are also only
between 46 to 51 points possible on each assessment which may not be enough to
accurately measure the taught curriculum (ODE, 2012c). It is possible that the
assessment did not match what content was taught by the teacher or that the assessment
items could not accurately measure the complexity level of the teacher’s instruction.

Next, it is important to note that the TI score is not provided to a candidate after
they complete the screener. Why would the score not be shared with the candidate? The
Gallup Organization (2007) explains that the TI is considered a pre-employment
interview and as such is not shared with the teaching candidate. Their rationale is that a
candidate would not receive feedback from any other employment interview and is
therefore congruent with typical school district practices. If a candidate does not score
well enough to enter the interview stage of the hiring process, he or she needs some type of feedback to improve. Otherwise, how will this candidate ever obtain a position in a school district? Is there any other assessment that a student would take where the score or some other feedback is not provided? By not sharing the TI score and how they may improve the score, it gives the impression that teaching is a profession that requires innate skills that cannot be taught. This seems contradictory for a profession that values continuous improvement and life-long learning for both its staff and students. So districts need to consider the message they are sending by using a screening instrument that appears to be counter to its mantra.

If a school district uses only the TI score to eliminate candidates from the applicant pool it can serve as a gate keeper (Metzger & Wu, 2007). School district administrators need to be aware of unknowingly discriminating against candidates based on race, color, sex, age, religion, or national origin with their selection process. Koerner (2007), found an inverse relationship between TI score and YOE for mathematics teachers. If a district only used the TI for teacher selection, then they may be inadvertently discriminating based on age in this example.

Finally, The Gallup Organization (2007) recommends that districts use the TI results as one of several means for identifying quality teachers for the interview stage of the hiring process. If a district is not utilizing multiple metrics when determining whether or not the candidate is worthy of an interview this is a problem. It is unethical to use the assessment results solely when the organization who created the assessment warns against it. Candidates should be assessed on multiple means beyond the TI score.
**Financial implications of study.** The TI teacher screening instrument costs school districts thousands of dollars a year. Based on the findings of this study and continued reductions in state funding, school districts must decide if the information provided by the TI score is worth the expense. One of the significant advantages of using the TI score in the hiring process is the amount of time and effort of human resources department saves. If additional criteria are recommended by The Gallup Organization (2007), then the time-efficiency may be lost.

**Personal implications of study.** Teaching is a service profession because it involves constant human interactions. Finding the best teaching candidate for a position can involve more than the TI or TGI can identify. Smith (2012) surveyed over 300 rural superintendents about what they valued when hiring teaching candidates. The superintendents favored personal skills like enthusiasm and creativity over professional skills. They believed the professional skills would be taught throughout professional development opportunities over time.

All Ohio school districts are required to provide high-quality professional development each year. More training can have a positive impact on teacher quality (Faurer, 2004). Finding the right fit for the specific position and school context is important (Smith, 2005). For instance if your school has many student behavior issues, you may need a candidate who can effectively de-escalate situations. These are just a few of the personal implications that must be considered with using the TI score in the teacher selection process.
Recommendations for future research studies

A recommendation to include other districts’ data is important because the study school district was an affluent, high-achieving, high-growth district (ODE, 2012a). It offered a competitive salary, had high parental involvement, and a low-incidence of discipline problems (ODE, 2012h). These factors could have an impact on the quality of all of the teaching candidates who applied. School districts with a fewer number of applicants for teaching vacancies might produce different results. Incorporating a wide range of school districts could address this question. So the recommendation is to incorporate data from all three typologies of school districts (rural, urban, and suburban) in future research studies of the TI screening instrument. Also including districts with varying achievement and growth data would be beneficial to the research literature. Many private and parochial school districts do not give the state assessments necessary for calculating a valid TGI score. However, if there were enough private and parochial school districts that have valid TGI scores and give the TI screening instrument it would allow for an additional factor to investigate.

Districts that currently use the TI screening instrument should also consider giving the TI to all staff with a valid TGI score in English/reading or mathematics. The study school district only began giving the TI screening instrument to prospective staff in the 2003 – 2004 school year and TGI scores were first available in the 2007 – 2008 school year. This would provide a larger sample with a greater range of scores. Adding more staff should increase the number of teachers who have very low scores on the TI. Examining the TGI scores of teachers who score low on the TI screening instrument can provide more understanding on its ability to screen effective teachers. The study school
district has unlimited access to the TI screening instrument so the only cost would be the
time for current teachers to take the assessment. The assessment only takes about 30
minutes to complete so it should be possible (Novotny, 2009; Smith, 2005). It would also
allow for the investigation of additional subject areas like science and social studies. The
initial data set had too few staff to include in the analyses. Although this study did not
find a relationship between the TI score and TGI score of English/reading and
mathematics one may exist with science or social studies teachers.

Another recommendation is to conduct a similar study with all districts that use
the TI screening instrument within all states that use the same value-added calculation.
The data set would contain a large enough sample to include an appropriate range of TI
and TGI scores. The results would also be more generalizable to those districts that did
not use the TI screening instrument because it included the TGI metric from their specific
state. Other teacher screening instruments could also be included in the analysis to
provide comparisons across a wide-range of tools. Through interviews with human
resource departments additional selection criteria could be identified to be incorporated
into the analysis.

The teacher-level value-added metric, measured by TGI scores, is relatively new
for many states (Kober & Rentner, 2012). An additional recommendation would be to
review the TGI scores of teachers three or five years after their initial teacher
effectiveness report. Prior to the reporting of TGI scores, teachers had little to no means
of knowing if their instruction was providing positive student academic growth. The
diagnostic information available from the TCAP report can provide an effective teacher
with information to drive continual improvement in his or her instruction. Would the
teachers identified as having “talent” by the TI screening instrument show higher growth than those with lower TI scores over a three or five year span? Reviewing TGI scores over several years could provide an additional factor to examine the effectiveness of the TI screening instrument.

With the results of regression analysis and the independent t-test for mathematics teacher’s males score a full point higher on the TI screening instrument than their female counterparts. Male teachers also obtain significantly higher TGI scores than their female counterparts indicating they are more effective as measured by this metric in the study school district. This might suggest that by simply being male in this school district on average you are a more effective mathematics teacher. This however poses two problems. First, no other evidence was located that gender alone predicts the effectiveness of mathematics teachers. Second, suggesting that school districts interview more males than females is not a practical approach as it would surely open up potential litigation on gender bias. Another question that arises is whether men leave the teaching profession early accounting for the significant difference in YOE? Based on these results, it is suggested that further research on additional factors is necessary to identify the variables that can assist a school district in ensuring the highest-quality candidates.

Finally, it would be interesting to see how The Gallup Organization might make changes to the TI screening instrument based on these and additional results. Additional questions might be created to better identify teachers who would be able to obtain effective TGI scores for English/reading and mathematics. Currently, school districts only receive a candidates overall TI score. Possibly providing sub-scores on each of the three themes could provide more insight. Regression toward each of the sub-scores might
provide a relationship to TGI scores. Pressure from state departments of education, parents, and community members will force district administrators to find candidates that can facilitate student academic growth. If the TI screening instrument can be shown to accurately identify candidates that will be able to facilitate this academic growth, many more districts would allocate the necessary resources to purchase such an instrument.

**Conclusions**

Districts are now more than ever hyper-vigilant in hiring the best teaching candidates possible from the applicant pool with the pressure from accountability measures. There is also more public awareness of teacher effectiveness through public media, as noted by the recent Cleveland Plain Dealer article and teacher effectiveness ranking database. The Cleveland Plain Dealer felt it was “more important to provide information, even if flawed, to help parents understand their children’s education” (O’Donnell, 2013, para. 11). Public pressures from parent and community members are only likely to increase with additional media outlets publicizing the TGI scores and ratings of teachers. With Ohio school districts routinely asking for tax levies to support the cost of educating students, parents and community members will demand effective classroom teachers.

Many districts already use teacher screeners to help separate the highest-quality candidates from the rest of the pool to enter the interview stage of the hiring process. With accountability and public pressures increasing, it is likely that more school districts will consider purchasing a teacher screening instrument to incorporate into their hiring process. School districts will not be able to risk hiring teaching candidates that cannot
attain high TGI scores quickly. The purpose of the study was to determine how well teacher-screening instruments identify the highest quality teachers from the candidate pool. This study examined the relationship between one specific teacher screening instrument, the TeacherInsight, and an estimate of teacher effectiveness, the teacher growth index score for English/reading and mathematics teachers. The expectation is that school districts using the TI screening instrument as a means to ensure that high-quality candidates in their hiring process can be assured to be interviewed. This could mean that effective teachers never make it to the interview stage of the hiring process.

This study has shown that the use of the TI screening instrument does not significantly predict the TGI scores of mathematics teachers. The TI score does significantly predict the TGI score of English/reading teachers, however this relationship is weak and is therefore not of practical use. The results of the study may not change the hiring process of school districts, although many need some method to sort through the hundreds of applications they receive for each teaching position that is available. These results should at least make school districts consider the benefit of using this particular teacher screening instrument in its hiring process. With the cost of the TI screening instrument when determining its benefit or at least including it as one of several criteria for moving from the candidate pool to the interview stage of the application process and not the sole determination should be considered.

The TeacherInsight and other teacher screeners are used across the nation in the hiring process. These results suggest that additional independent research into the ability of these instruments to identify the most effective teachers is warranted. Possibly the TI screening instrument is identifying talent that is critical to high-quality teaching, but
cannot be measured by the TGI score alone. Add to this the fact that the accountability systems of many states now includes a value-added metric will push school districts to ensure their teaching candidates can produce high student academic growth (Kober & Rentner, 2012).
References


Ohio Department of Education. (2012a). *District achievement testing results*. Retrieved from
http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&
TopicRelationID=222&ContentID=15606&Content=128217

http://esb.ode.state.oh.us/

http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&
TopicRelationID=285&ContentID=9479&Content=131427

http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&
TopicRelationID=1327&ContentID=43712

http://www.education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&
TopicRelationID=1292&ContentID=100905&Content=135642

Ohio Department of Education. (2012f). *Whose IDEA is this: A parent’s guide to the individuals with disabilities education improvement act of 2004 (IDEA)*. Retrieved from
http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&
TopicRelationID=1540&ContentID=11128&Content=124547


Ohio Department of Education. (2012i). *Ohio teacher evaluation system: Training workbook* (pp. 1-165, Rep.) Columbus, Ohio: Ohio Department of Education.


Regan, N., & Hayes, K., (2011). Accuracy of the TeacherInsight online perceiver tool in determining the effectiveness of high rated and low rated math and science teachers following one year and three years of single school district employment. *International Journal of Educational Leadership Preparation, 6*(4). Retrieved from http://cnx.org/content/m41321/latest/


APPENDIX A

HSRB APPROVAL
TO: Tom Fry
FROM: Brent Mattingly, HSRB Acting Chair
DATE: April 23, 2013
SUBJECT: Human Subjects Review Board Approval
PROJECT TITLE: Relationship between Teacher Insight and Teacher Growth Index
HSRB APPROVAL CODE: 04-1913-8112

The Human Subjects Review Board has approved the research proposal you submitted. You may proceed with the project.

The primary function of the HSRB is to ensure protection of human research subjects. As a result of this mandate, we ask that you pay close attention to the fundamental ethical principles of autonomy, justice, and beneficence when establishing your research proposal. These ethical principles pertain specifically to the issues of informed consent, fair selection of subjects, and risk/benefit considerations.

If you have any questions, please contact me.

Sincerely,

Brent Mattingly, Ph.D.
Phone: 419-289-5342
E-mail: bmattingly@ashland.edu
APPENDIX B

DISTRICT APPROVAL
Data Approval Form

Relationship between TeacherInsight and Teacher Effectiveness

A. PURPOSE AND BACKGROUND

The objective of my research is to determine the relationship between TeacherInsight (TI) scores and teacher effectiveness as measured by TCAP reports. You are being asked to participate because you utilize both TI and TCAP reports.

I will use the information to analyze the relationship between TI and TCAP reports. The results will also be reported in a dissertation for my doctoral studies at Ashland University and may be reported in a journal article. No identifiable teacher information will be shared.

B. PROCEDURES

If you agree to be in the study, the following will occur:

1. I will meet with district personnel to identify necessary data including TI scores, TCAP reports, years of experience, and gender.

2. OLSD will combine data and scrub any teacher identifiable information.

C. RISKS/DISCOMFORTS

There are no immediate or long-term risks/discomforts.

D. BENEFITS

The information that you provide will inform OLSD of the relationship between TI and teacher effectiveness. This research will add to the sparse literature linking TI and teacher effectiveness.

E. COSTS

There will be no costs to you as a result of taking part in this study.

F. PAYMENT

There will be no payment made to you as a result of taking part in this study.

G. QUESTIONS

You have talked to Tom Fry about this study and have had your questions answered. If you have further questions, you may e-mail him at tfry1@ashland.edu or call him at 614-218-1569.

If you have any comments or concerns about participation in this study, you should first talk with the researcher. If for some reason you do not wish to do this, you may contact Mr. Fry’s advisor, Dr. Carla Edlefsen at 614-794-0803, x1112, or the Ashland University Human Subjects Review Board, which is
concerned with the protection of volunteers in research projects. You may reach the board office between 8:00 and 5:00, Monday through Friday, by calling (419) 207-6198 or writing the Dean of the Graduate School, 100 Founders Hall, Ashland University, Ashland, Ohio 44805.

H. CONSENT

You will be given a copy of this consent form to keep. PARTICIPATION IN RESEARCH IS VOLUNTARY. You are free to decline to be in this study, or to withdraw from it at any point. Your decision as to whether or not to participate in this study will have no adverse consequences.

If you agree to participate, you should sign below:

[Signature]

Date: 12-21-12

[Signature]

Director of Curriculum and Instruction
APPENDIX

ODE APPROVAL
Tom,

Any documents that ODE created are considered part of the public domain. You can use them without permission --- just cite the source as ODE.

Good luck!!

Marianne E. Mottley  
Assistant Director  
Office of Accountability  

25 South Front Street | Columbus, Ohio 43215-4183  
(614) 995-9944 | (877) 644-6338  
Marianne.mottley@education.ohio.gov  
education.ohio.gov

From: Thomas Fry [mailto:tfry@granvilleschools.org]  
Sent: Thursday, September 26, 2013 7:06 PM  
To: Mottley, Marianne  
Subject: ODE website

I would like to use the two figures from the OTES documents on the ODE website for my dissertation. Do you know who I can ask for permission or if it is needed? My oral defense is Oct 10th and I need to have this figured out by then. I appreciate any help you can provide.

Thank you,

Tom

Thomas Fry Jr.  
Assistant Superintendent  
Granville Exempted Village Schools  
740.587.8186 work  
614.218.1569 cell

PLEASE NOTE: This message and any response to it may constitute a public record, and therefore may be available upon request in accordance with Ohio public records law. (ORC 149.43)