STUDENT STANDARDIZED TEST SCORES AS AN EFFECTIVE MEASURE OF TEACHER PERFORMANCE: TEACHER AND ADMINISTRATOR ATTITUDES

By Chad Lanese

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

in Educational Leadership

Northern Arizona University

May 2018

Approved:

Richard Wiggall, Ed.D., Chair

Troy Bales, Ed.D.

Walter J. Delecki, Ph.D.

Mary Dereshiwsky, Ph.D.

ABSTRACT

STUDENT STANDARDIZED TEST SCORES AS AN EFFECTIVE MEASURE OF TEACHER PERFORMANCE: TEACHER AND ADMINISTRATOR ATTITUDES

CHAD LANESE

In recent years, the focus of accountability has emphasized standardized test results and the individual teacher. Weight has increasingly been given to the use of standardized test results as an important tool to measure teacher effectiveness and has been implemented in a number of states through legislation that mandates its use on evaluation instruments. The emergence of Value-added Measures and Student Growth Percentiles as viable tools to accomplish the task of measuring teacher effectiveness using standardized test scores has become more prominent across school districts in the United States. However, much of the research on these tools has shown that they may not be stable enough to use in an evaluation. Additionally, the relative instability of these measures creates a larger concern when used in teacher evaluations because the results of evaluations can influence decision-making around teacher tenure, dismissal and even compensation. Using an unproven method to make these types of decisions is wrought with potential concerns about the issue of measuring teacher effectiveness.

This study served to describe the attitudes of high school mathematics teachers and high school administrators regarding the use of standardized test results on teacher evaluations. The researcher administered quantitative surveys and qualitative interviews for the purpose of better describing teacher and administrator attitudes. Both quantitative survey results and qualitative interview results were analyzed in order to better understand teacher and administrator attitudes toward the use of student standardized tests results as an indicator of performance on teacher evaluations. When describing teacher attitudes, three groups of mathematics teachers served as a

valuable source of data. Groups consisted of teachers of accelerated courses only, non-accelerated only, and those that instructed both accelerated and non-accelerated courses as a part of their teaching assignment. Administrator and teacher attitudes were also analyzed in order to compare the two groups. Each group responded to the same survey items and statistical analysis was applied for the purpose of comparing the two groups. The research design was mixed methods where sequential timing was applied as a subset methodology design in order to gather survey data first and then interview data.

Survey and interview questions were categorized into three themes as related to teacher evaluation. Theme one considered the concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness. Theme two questions focused on attitudes towards standardized test results and the degree of trust that participants had/did not have with regard to standardized test results. Theme three questions considered the actual process of teacher evaluations within the organization. The findings suggest that a clear attitude of disagreement exists for each group regarding theme one questions. However, the accelerated and combined (accelerated/non-accelerated) groups expressed disagreement toward theme two while the non-accelerated group was more neutral. Theme three produced similar neutral and agree responses across the three groups of teachers. Administrators also expressed an attitude of disagreement towards theme one, but were more neutral in responses to themes two and three. The idea of using standardized test results as a tool to evaluate teacher performance was met with disagreement for both teachers and administrators.

ACKNOWLEDGEMENTS

This study would not have been possible without the support and expertise of a number of outstanding individuals to whom I am forever grateful.

- To my Committee Chair, Dr. Richard Wiggall, whose regular encouragement and reflective approach always served to lead me in the right direction. Thank you for your dedication and contribution to my study.
- To Dr. Mary Dereshiwsky, a true researcher at heart and statistical wizard that served to provide crucial guidance instrumental to the successful completion of this study. Your enthusiasm and positive approach is inspirational.
- To Dr. Wally Delecki, the consummate professional and educational leader. Your willingness to problem solve, actively listen and say the right thing at the right time is second to none.
- To Dr. Edith Hartin, the editor of all editors. I am honored that you were willing to share your knowledge, wisdom, and dissertation expertise with me throughout this process. Thank you for assisting me in the completion of this academic endeavor as I could not have done it without you.
- Last but certainly not least, Dr. Troy Bales who has served as a friend and mentor from the very beginning of my career. You have always challenged me to be the very best that I can be and I am truly honored to call you my colleague and friend. Les is an angel on both of our shoulders and I know he is looking down on us with pride.

TABLE OF CONTENTS

CHAPTER		PAGE
1	Introduction	1
	Background of the Study	1
	Statement of the Problem	5
	Purpose of the Study	6
	Research Questions and Hypotheses	6
	Definitions of Terms	8
	Acronyms Used	10
	Limitations	11
	Delimitations	13
	Assumptions	14
	Significance of the Study	14
	Organization of the Study	17
	Summary	18
2	Review of the Literature	19
	Introduction	19
	Educational Accountability in the United States	19
	History/Evolution of Teacher Evaluation	25
	Current National Teacher Evaluation	53
	Arizona Teacher Evaluation	64
	Summary	69

CHAPTER PAGE

3 Research Design and Methodology 70

Introduction 70

Restatement of the Problem 70

Restatement of the Purpose of the Study 71

Restatement of Research Questions and Hypotheses 71

Research Design 73

Census and Sample 74

Instrumentations 77

Validity and Reliability 80

Data Collection Procedures 83

Data Analysis Procedures 84

Summary 91

4 Findings 92

Introduction 92

Research Question 1 94
Research Question 1a Findings 94
Quantitative RQ1a Findings: Group A 94
Quantitative RQ1a, Theme 1 Findings: Group A 96
Quantitative Summary RQ1a, Theme 1: Group A 99
Qualitative RQ1a Theme 1 Findings: Group A 100
Qualitative Summary RQ1a, Theme 1: Group A 105

Quantitative RQ1a, Theme 2 Findings: Group A 106
Quantitative Summary RQ1a, Theme 2: Group A 108
Qualitative RQ1a Theme 2 Findings: Group A 109
Qualitative Summary RQ1a, Theme 2: Group A 111

CHAPTER PAGE

Quantitative RQ1a, Theme 3 Findings: Group A 112
Quantitative Summary RQ1a, Theme 3: Group A 114
Qualitative RQ1a Theme 3 Findings: Group A 115
Qualitative Summary RQ1a, Theme 3: Group A 120

Summary for RQ1a: Group A 120

Research Question 1b Findings 121
Quantitative RQ1b Findings: Group B 121
Quantitative RQ1b, Theme 1 Findings: Group B 123
Quantitative Summary RQ1b, Theme 1: Group B 125
Qualitative RQ1b, Theme 1 Findings: Group B 127
Qualitative Summary RQ1b, Theme 1: Group B 131

Quantitative RQ1b, Theme 2 Findings: Group B 132
Quantitative Summary RQ1b, Theme 2: Group B 134
Qualitative RQ1b, Theme 2 Findings: Group B 135
Qualitative Summary RQ1b, Theme 2: Group B 138

Quantitative RQ1b, Theme 3 Findings: Group B 138
Quantitative Summary RQ1b, Theme 3: Group B 141
Qualitative RQ1b, Theme 3 Findings: Group B 143
Qualitative Summary RQ1b, Theme 3: Group B 146

Summary for RQ1b: Group B 147

Research Question 1c Findings 147
Quantitative RQ1c Findings: Group C 147
Quantitative RQ1c, Theme 1 Findings: Group C 149
Quantitative Summary 1c, Theme 1: Group C 151
Qualitative RQ1c, Theme 1 Findings: Group C 153
Qualitative Summary RQ1c, Theme 1: Group C 157

Quantitative RQ1c,, Theme 2 Findings: Group C 157
Quantitative Summary RQ1c, Theme 2: Group C 159
Qualitative RQ1c, Theme 2 Findings: Group C 160
Qualitative Summary RQ1c, Theme 2: Group C 164

Quantitative RQ1c, Theme 3 Findings: Group C 164
Quantitative Summary RQ1c, Theme 3: Group C 166
Qualitative RQ1c, Theme 3 Findings: Group C 168
Qualitative Summary RQ1c, Theme 3: Group C 171

CHAPTER PAGE

Summary for RQ1c: Group C 172

Overall Summary for RQ1 173

Research Question 2 Findings 175

Research Question 3 Findings 179
Quantitative RQ3 Findings: Administrators 180
Quantitative RQ3, Theme 1 Findings: Administrators 181
Quantitative Summary RQ3, Theme 1: Administrators 183
Qualitative RQ3, Theme 1 Findings: Administrators 184
Qualitative Summary RQ3, Theme 1: Administrators 189

Quantitative RQ3, Theme 2 Findings: Administrators 189
Quantitative Summary RQ3, Theme 2: Administrators 191
Qualitative RQ3 Theme 2 Findings: Administrators 192
Qualitative Summary RQ3, Theme 2: Administrators 195

Quantitative RQ3, Theme 3 Findings: Administrators 195
Quantitative Summary RQ3, Theme 3: Administrators 197
Qualitative RQ3 Theme 3 Findings: Administrators 199
Qualitative Summary RQ3, Theme 3: Administrators 202

Overall Summary for RQ3 202

Research Question 4 Findings 203

Summary of Findings 209

Summary 213

5 Summary, Conclusions, Implications, and Recommendations 215

Introduction 215

Summary of the Study 215

Conclusions 222

Implications for Practice 224

Recommendations for Future Studies 226

CHAPTER PAGE

Concluding Remarks....228

REFERENCES....230

APPENDICES

A Teacher and Administrator Surveys....239

B Interview Questions....245

C NAU IRB Approval....246

D NAU Informed Consent....247

E District Approval....251

F Principal Letter....253

G Teacher Cover Letter....254

BIOGRAPHICAL INFORMATION....255

LIST OF TABLES

TABLE PAGE

1 Theme Alignment of Survey and Interview Questions....80

2 Match-up of Research Questions to Corresponding Sources of Information and Data Analysis/Reporting Procedures....87

3 Multimethod Convergence Information....89

4 Group A: Gender, Degree, Total Years Teaching, and Years Teaching in District A....95

5 RQ1a, Theme 1: Group A Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)....100

6 Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Components....102

7 Student Achievement Data Supporting Intended Purpose of Teacher Evaluation....103

8 Effective Method of Teacher Evaluation using Standardized Test Results....104

9 Standardized Testing Results Serve as an Indicator of Teacher Evaluation....105

10 RQ1a, Theme 2: Group A Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)....109

11 Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation....110

12 Trust Student Standardized Tests as a Measure of Performance....111

13 RQ1a, Theme 3: Group A Accelerated Teachers (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)....115

14 Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach....117

15 Standardized Testing Results Serves as a Tool that can Influence Teacher Performance....119

16 Standardized Testing Results Serves as a Tool that can Influence Professional Growth....120

TABLE PAGE

17 Group B: Gender, Degree, Total Years Teaching, and Years Teaching in District A) 122

18 RQ1b, Theme 1: Group B Non-Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)....126

19 Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component....128

20 Student Achievement Data Supporting Intended Purpose of Teacher Evaluation 129

21 Effective Method of Teacher Evaluation using Standardized Testing Results....130

22 Standardized Testing Results Serve as an Indicator of Teacher Effectiveness 131

23 RQ1b, Theme 2: Group B Non-Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)....135

24 Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation 137

25 Trust Student Standardized Tests as a Measure of Performance....138

26 RQ1b, Theme 3: Group B Non-Accelerated Teachers (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)....142

27 Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach 144

28 Standardized Testing Results serves as a tool to Influence Teacher Performance....145

29 Standardized Testing Results serves as a tool that can Influence Professional Growth 146

30 Group C: Gender, Degree, Total Years Teaching, and Years Teaching in District A) 149

31 RQ1c, Theme 1: Group C Accelerated and Non-Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28) 152

32 Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component 154

TABLE PAGE

33 Student Achievement Data Supporting Intended Purpose of Teacher Evaluation 155

34 Effective Method of Teacher Evaluation using Standardized Testing Results.....156

35 Standardized Testing Results Serve as an Indicator of Teacher Effectiveness 157

36 RQ1c, Theme 2: Group C Accelerated and Non-Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27) 160

37 Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation 162

38 Trust Student Standardized Tests as a Measure of Performance.....164

39 RQ1c, Theme 3: Group C Accelerated and Non-Accelerated Teachers (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20) 167

40 Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach 169

41 Standardized Testing Results Serve as a Tool to Influence Teacher Performance.....170

42 Standardized Testing Results Serve as a Tool to Influence Professional Growth 171

43 Kruskal Wallis Comparisons of Accelerated, Non-Accelerated and Both Groups 179

44 Administrators: Gender, Degree, Total Years Teaching, and Years Teaching in District A) 181

45 RQ3, Theme 1: Administrators (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28).....184

46 Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component 186

47 Student Achievement Data Supporting Intended Purpose of Teacher Evaluation 187

48 Effective Method of Teacher Evaluation using Standardized Testing Results.....188

49 Standardized Testing Results Serve as an Indicator of Teacher Effectiveness 189

TABLE PAGE

50 RQ3, Theme 2: Administrators (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ26, SQ27) 192

51 Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation 193

52 Trust Student Standardized Tests as a Measure of Performance 194

53 RQ3, Theme 3: Administrators (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20) 198

54 Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach 200

55 Standardized Testing Results Serves as a Tool that can Influence Teacher Performance 200

56 Standardized Testing Results Serves as a Tool that can Influence Professional Growth 201

57 Mann-Whitney U Test of Administrator and Teacher Survey Responses 209

58 Summary Findings of Survey Questions and Interviews for each Research Question 211

DEDICATION

This work is dedicated to the most important people in my life, my wonderful family. Your support, encouragement and words of wisdom were essential to the completion of this work. To my amazing wife Songhui, I simply could not have done this without you. You have been my best friend and cheerleader since day one. To my children, Katelyn and Ethan, the pride and joy of my life. Although I have been a bit distracted, I look forward to making that up to you and truly hope that I serve as a positive role model that inspires you to pursue whatever it is in life that you choose. Lastly, to my parents Connie and Jeff. Your unwavering support has been instrumental as this was certainly a team effort. I am thankful for your guidance and hope that you are as proud to be my parents as I am proud to be your son.

CHAPTER 1

Introduction

Background of the Study

In the educational world accountability is a key concept that has emerged over the last two decades as an important priority, as noted by Guthrie (2003), "The United States is on a sustained and intense path seeking means for rendering the education system more effective" (p. ix). Educational reform has taken many approaches and invariably places emphasis on student achievement as related to standardized tests. For example, Guthrie (2003) stated that "High performance schools, high-stakes testing, academic accountability, teacher productivity…are illustrative of the slogans, issues, and topics that dominate American education policy and practice at the onset of the twenty-first century" (p. x). A specific focus in recent years has been on using the student achievement data from standardized tests to measure the effectiveness of schools and the classroom performance of teachers.

Efforts to legally mandate the use of these data on teacher evaluations have become more prevalent and, in many cases, an important tool used in order to make decisions related to compensation, tenure, and dismissal. According to the Center on Great Teachers and Leaders at the American Institute for Research (2013), the databases on state teacher and principal evaluation policies note that 21 states use teacher evaluation systems to determine varying levels of compensation and 27 use evaluation instruments for purposes of dismissal.

While no logical argument exists that refutes the importance of a focus on continuous improvement as necessary in all educational settings, it could prove important to consider the process one uses to make the determination of whether or not a teacher is considered effective. As reforms continue to evolve in the educational world, standardized testing is increasingly

utilized as a tool that is relied upon more heavily in order to measure teacher performance and ultimately decide if their performance is considered effective.

In education, teacher evaluation systems are the primary instrument that is used to measure a teacher's classroom performance. There has been an increase in accountability and an increased integration of teacher evaluation instruments with student achievement data in order to determine teacher efficacy. For example, Bergin (2015) noted that

> The U.S. has had a short history of using achievement data to evaluate a school or district; the new movement is to use it to evaluate individual teachers' effectiveness. Indeed, the US Department of Education has used the Race to the Top (RTTT) funds and the ESEA waiver process to obligate states to use student achievement data as a 'significant' part of teacher evaluation. (p. 1)

Bergin (2015) described that educational reform efforts have placed an emphasis on accountability intended to identify both effective and ineffective teachers in that "The hope is that student achievement data differentiates teachers better than old, inadequate evaluation systems that simply labeled teachers as 'satisfactory' or not" (p. 1).

Efforts at emphasizing standardized testing results with regard to teacher performance are evident in federal and state legislation throughout recent history in the United States. For example, legislation such as No Child Left Behind (NCLB) and RTTT serve as examples where increased accountability, that has a strong connection to the use of standardized testing results, is seen. According to Dee and Jacob (2009), "The No Child Left Behind (NCLB) Act is arguably the most far-reaching education-policy initiative in the United States over the last four decades" (p. 2). NCLB served to label public schools based upon their performance on mandated, standardized testing results. The RTTT initiative shifted the focus of NCLB from schools to a

much more limited emphasis on teachers and leaders. Evidence of this focus is seen with regard to how points are awarded on the application for successful completion of the RTTT competitive grant. According to McGuinn (2014), the percentage of points is allocated in a very specific manner as it relates to the six broad categories described in the application. For example, the percentages are structured as follows: State Success Factors (25%), Standards and Assessments (14%), Data Systems (9%), Teachers and Leaders (28%), School Turnaround (10%), and General (14%). What is most telling about the percentage distribution is the emphasis on the category of Teachers and Leaders. This portion of the application has the highest percentage of points, which results in drawing the conclusion that this section is an emphasis of federal reform. The specific focus in the Teachers and Leaders section on the RTTT application is even more interesting when one considers McGuinn's (2014) description of this section. He noted that "The section on Teachers and Leaders (28%) pushed states to develop better teacher and principal training, evaluation, and distribution all with a focus on student performance" (p. 65).

In looking at both the development of NCLB and RTTT, a clear picture emerges in how accountability for educators has been shaped over the past decade. A local emphasis on accountability, particularly related to student achievement and teacher performance, also exists in the educational setting. In Arizona, ARS 15-203 (2012) provides one such example of how accountability manifests itself in the system. This legal requirement, as found on the Arizona Department of Education (ADE) website, states:

> The Arizona Board of Education shall adopt and maintain a model framework for a teacher and principal evaluation instrument that includes quantitative data on student academic progress that accounts for between thirty-three percent and fifty percent of the evaluation outcomes. (ARS§15-203(A)(38), 2012, section 38)

The narrative described illustrates how accountability has changed in education to further emphasize the use of standardized testing results as a tool that measures teacher performance.

While the intention to develop methods of measuring teacher performance is necessary and can certainly impact the educational world in a positive manner, it is also important to note that both federal (RTTT) and state (A.R.S. 15-203 in Arizona) entities are legally mandating these initiatives as an approach in the public education system. RTTT illustrates a federal reform with a strong connection between evaluation and student achievement. McGuinn stated (2014), "among the fourteen criteria for RTTT eligibility were requirements that states not have a firewall preventing the linking of student achievement data with individual teacher information" (p. 69). A clear priority from the federal level is set with regard to the connection between using student achievement data and measuring individual teacher performance. In Arizona, A.R.S 15-203 provides us with an example at the state level mandating that student achievement data be included as a part of all principal and teacher evaluations (ARS§15-203(A)(38), 2012, section 38). At the state level, Arizona is certainly not alone in its efforts to mandate that student achievement be a part of teacher performance evaluations. According to the Center on Great Teachers and Leaders (2013) at the American Institute for Research, the databases of state teacher and principal evaluation policies, student data components on teacher evaluations are required by legislation in more than 30 states. Additionally, according to Lash, Makkonen, Tran, and Huang (2016), "as of early 2014, 40 states and the District of Columbia were using or piloting methods to evaluate teachers in part according to the amount students learn" (p. 3).

It could prove meaningful to pose a few questions when considering the use of student achievement to measure teacher effectiveness including, "what quantitative data on student academic performance effectively measures teacher performance?" and "how can that be used

successfully on an evaluation instrument?" In an effort to answer these questions and quantify teacher effectiveness, many studies have been conducted looking at a variety of accountability models that attempt to use student achievement data to make high-stakes decisions about teacher performance. Further, these initiatives are moving forward before a proven system has been identified that can actually accomplish this task. For example, in Berliner's (2014) analytic essay, he noted that "There has been rapid growth in value-added assessment of teachers to meet the widely supported policy goal of identifying the most effective and the most ineffective teachers in a school system" (p. 1). The use of student growth percentiles as well as value-added models, have been examined by a number of researchers in the field in order to inform policymakers. When considering student growth percentiles as a tool to measure teacher efficacy, Lash et al. (2016) contended that "This is perhaps the first published study on the stability of the teacher-level growth scores derived under the student growth percentile model, a common model used by states in teacher evaluation systems" (p. 7).

Statement of the Problem

The problem examined in the study involved the attitudes of teachers and principals. More specifically, what are teacher and principal attitudes towards standardized testing results when they served as an indicator that measured an individual teacher's performance? As educational organizations attempt to tie teacher effectiveness to student achievement, a great deal of research indicates that the measures currently utilized to accomplish this task are not considered stable enough to do so. More specifically, value-added measures (VAM) and student growth percentiles (SGP) are two instruments that have been widely utilized for this purpose. Each has been studied and identified as insufficient for the purpose of measuring teacher performance in connection with student achievement data. For example, Corcoran's (2010)

work provided examples with regard to VAMs in that "given the extent of uncertainty in teacher value-added scores, it would not be surprising if these estimates fluctuated a great deal from year to year" (p. 6). Corcoran specifically identified evaluation, promotion, compensation, and dismissal of teachers as outcomes for using value-added systems. Often times, this research provides an indication that these accountability systems are unstable or require further examination. Additionally, Lash et al. (2016) noted that "the findings indicate that even when computed as an average of annual teacher-level growth scores over three years, estimates of teacher effectiveness do not meet the level of stability that some argue is needed for high-stakes decisions about individuals" (p. 7). While these measures are under debate about their relative effectiveness, the problem arises in that they have been used as a piece of teacher performance evaluations.

Purpose of the Study

The purpose of this study was to examine attitudes of both high school math teachers and administrators regarding whether or not student standardized test results were seen as effective measures of instructional performance.

Research Questions and Hypotheses

The research questions and hypotheses that guided this study included:

1. What are the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?
    a. What are the attitudes of teachers instructing accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

b. What are the attitudes of teachers instructing non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

c. What are the attitudes of teachers that instruct both accelerated and non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

2. Is there a statistically significant difference between the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

   $H_0$2. There is no statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

   H2. There is a statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

3. What are the attitudes of high school administrators regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?

4. Is there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?

   $H_0$4. There is no statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of math teacher instructional performance in the classroom.

   H4. There is a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of math teacher instructional performance in the classroom.

Definitions of Terms

There are a number of key terms and operational definitions that must be defined for the purpose of this research.

Academic Achievement: According to Steinmayr, MeiBner, Weidinger, and Wirthwein (2015), academic achievement can be defined as:

> The performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university. School systems mostly define cognitive goals that either apply across multiple subject areas (e.g., critical thinking) or include the acquisition of knowledge and understanding in a specific intellectual domain (e.g., numeracy, literacy, science, history). Therefore, academic achievement should be

considered to be a multifaceted construct that comprises different domains of learning. (p. 1)

In applying this broad definition of academic achievement one can operationally define it to include the performance level of students who took the AZMerit examination, grades earned in classes, and performance on academic tasks within the classroom, typically developed by the teacher of that class.

Administrator: For the purposes of this study, the term administrator referenced three specific groups. The three groups were high school principals, high school assistant principals, and district level administrators. The definition also included the notion that each administrator needed to either evaluate or train high school math teachers.

AZ Merit: According to the Arizona Department of Education's AzMerit parent brochure (n.d.),

> AzMerit stands for Arizona's Measurement of Educational Readiness to Inform Teaching. AzMerit is a computer-based test which provides engaging questions and measures critical thinking skills for college and career readiness. AzMERIT is aligned to Arizona's state learning standards which detail the concepts covered in select courses. The test is designed to measure student mastery of course-specific skills and readiness for college or career. (p. 2)

Standardized test: According to The Glossary of Education Reform (2015),

> A standardized test is any form of test that (1) requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that (2) is scored in a 'standard' or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. (p. 1)

In this case, the one test that serves as an example of this operational definition is AzMerit, which is considered to be a standardized test that is administered to all students in public schools in the state of Arizona and according to the AzMerit Testing Conditions, Tools and Accommodations Guidance document (2017), "AzMerit is a standardized test" (p. 1).

Teacher: The definition of teacher included specific criteria. For example, teachers included individuals that instruct math courses for students in grades 9-12 that were consistent across the five comprehensive high schools in the district and instruction of courses that were either accelerated (AP/Honors), not accelerated or both.

Teacher Evaluation: Operationally defined as the use of standardized test results/academic performance of students as a component that feeds an overall performance rating of the teacher on a teacher evaluation instrument used in one specific school district. According to Danielson (2016),

> The Framework for Teaching is a research-based set of components of instruction, aligned to the INTASC standards, and grounded in a constructivist view of learning and teaching. The complex activity of teaching is divided into 22 components (and 76 smaller elements) clustered into four domains of teaching responsibility: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. (p. 1)

Acronyms Used

ADE: Arizona Department of Education

AzMerit: Arizona Merit Standardized Test

NCLB: No Child Left Behind

RTTT: Race to the Top

SGP: Student Growth Percentile

VAMs: Value-added Measures

Limitations

While limitations are outside the power of the researcher's influence, delimitations are controlled by the researcher (Roberts, 2010, p. 13). The limitations of this study include:

1. One potential limitation of this study is the issue of inter-rater reliability. While the same evaluation instrument is used by all administrators to evaluate performance in this K-12 district, each individual administrator brings with them varied experiences with regard to their own understanding of quality instruction. The end result could be the possibility that implementation of rubrics utilized in the evaluation instrument can be applied too rigidly or loosely across administrators, thus compromising data. The human resources department maintains what can be described as an online filing cabinet where all evaluation results are maintained for employees. These data have been accessed by the district's human resources department for the purpose of analyzing results in order to identify the distribution of teacher performance ratings across the district in order to continuously train administrators on inter-rater reliability as it relates to evaluating classroom instruction while using the evaluation instrument's rubrics.
2. An additional limitation can be attributed to the fact that the researcher studied a K-12 unified school district, which means data is not consistent across every teaching assignment. Classroom teachers use different measures to demonstrate student performance which is dependent upon the classroom and content being taught. For example, a math or English teacher can utilize a common assessment developed by

the district that resembles a standardized test. However, a fine arts teacher does not have this same measure to demonstrate evidence of student performance. Different assessments are used to assess student progress, which creates a comparative challenge. To mitigate this threat to validity, the researcher only included high school teachers in grades nine through twelve who teach mathematics and also administer the AzMerit standardized test. This allows for more consistency with regard to data collected on teacher perceptions towards measuring performance. Doing so provided a consistent data point for determining student achievement among teachers who have regular experience analyzing this type of test data.

3. Educational programming within the high schools in this district vary significantly. A possible limitation is that specific high schools may offer programs focused on STEM, Career and Technical Education, Advanced Placement, and Dual Enrollment. Each program utilizes varied measures to determine what student success looks like and each may be quite different from one another. This phenomenon also impacts comparisons between teachers with regard to student success measures. Therefore, in order to address this limitation, the researcher has maintained a sample to include high school mathematics teachers. The district has developed guaranteed and viable curriculum maps for this content area as well as accompanying assessments. This means that the curricular expectations are far more common when considering mathematics. Additionally, the development of curriculum maps helps to establish more consistency among mathematics teachers with regard to indicators of performance.

4. The possibility of non-responses of survey participants is considered to be a limitation of this study. Weekly reminders will be sent to all survey participants as a mechanism to increase participation of survey participants.
5. The possibility of a limited number of survey responses from specific sub-groups (accelerated, non-accelerated, and accelerated/non-accelerated) of mathematics teachers is a limitation of this study as teachers self-reported teaching assignments through the survey and were categorized into sub-groups based upon responses. This led to a smaller number of responses on the part of the accelerated group of mathematics teachers.

Delimitations

When considering the scope of this study, it is important to identify a number of parameters. According to Roberts (2010), "a delimitation differs, principally, in that it is controlled by the researcher" (p. 139).

1. Groups included in the sample will consist of high school mathematics teachers and high school administrators.
2. The primary criteria for selection to each group will be that teachers currently serve in the same school district as defined by those who work in a setting that serves students in grades 9 through 12.
3. The primary criteria for selection to each group when considering administrators will be that they are employed in the same K-12 school district and supervise/train mathematics teachers.
4. Data gathered from various teachers will be treated in aggregate and not separated by individual schools.

5. Data gathered from various administrators will be treated in aggregate and not separated by individual schools.

Assumptions

There are a number of assumptions throughout the course of this study:

1. Both teachers and administrators understand that standardized tests include state mandated tests that students take each year.
2. The standardized tests have students answer the same questions and all are graded in the same way.
3. References will be made to standardized tests that have been utilized to gather student performance data in Arizona such as AIMS and AzMerit tests.

Significance of Study

An important goal of this study was to examine the attitudes that high school mathematics teachers hold regarding the effectiveness of standardized testing results as effective measures of performance. Further, administrator attitudes on whether the use of student achievement data as an accurate measure of teacher performance was also examined in this study. A study that focuses on teacher and administrator attitudes provides the opportunity to better define how professionals in the field see standardized tests as a tool to measure their performance. Teachers spend their lives dedicated to the pursuit of making sure that student learning is taking place in the classroom. Furthermore, administrators hold a comprehensive view of instruction and student learning as much of their time is dedicated to evaluation of teachers' instructional performance. This background certainly provides an insight into the fact that both teachers and administrators spend an extensive amount of time looking for ways to identify effective instructional performance and helps one understand that a level of credibility

exists among these professionals. By examining their attitudes on how standardized tests factor into that conversation, it may be possible to add information to the field of education with regard to the most appropriate role of standardized tests as related to instructional performance. Furthermore, a study of this nature served to better inform policy makers in how they influence educational reform as related to accountability.

Much of today's research has revealed that the use of standardized testing to determine teacher effectiveness is difficult to do in an ethical, consistent manner. This dilemma warrants a need to solicit input and gather data directly from professionals in the field of education on this topic. Teachers and administrators work in schools on a daily basis with students and serve to provide a necessary perspective on the relationship that exists between the measurement of their performance and standardized testing results. There are a number of models currently being utilized across the United States in order to evaluate teacher performance, specifically VAMs and SGP. Darling Hammond (2011) noted that "Value-added models of teacher effectiveness are highly unstable: Teachers' ratings differ substantially from class to class and from year to year, as well as from one test to the next" (p. 1). The problem arises through the fact that the research has described these models as unstable and inconsistent. Lash, et al. (2016) stated that "The current finding that teacher-level growth scores are so unstable as to raise questions about their use in teacher evaluation systems is similar to conclusions that other researchers have drawn about value-added measures of teacher effectiveness" (p. 7). Often times, the recommendations made by researchers in the field indicate that the use of these models should not be intended to influence decisions about teacher performance, compensation, and tenure. Corcoran (2010) noted that "In Washington, D.C., and Houston, teachers can be granted or denied tenure partially based on value-added, and Houston awards bonuses to its high value-

added teachers" (p. 1). Although this information is developed in order to assist policymakers in making appropriate decisions about legislation focused on accountability, one still sees legal mandates that require districts to implement a system that utilizes standardized tests as a tool to measure teacher performance through formal channels such as teacher evaluations systems. For example, many states attribute anywhere from 20% to 50% of their teacher evaluations to student growth and/or achievement data (Center on Great Teachers and Leaders, 2013).

A study of this type can provide practitioners in the field of education with more insight into how teachers and administrators view the use of standardized testing results as a measure of performance. A description of both teacher and administrator perceptions adds more information to the field of education in that it helps to add voice to a group that is impacted by accountability and reform measures that rely heavily on using standardized testing results as a way to categorize the performance of both schools and individual teachers. Both NCLB and RTTT have been cited as accountability measures in education reform that serve to emphasize the use of standardized testing results as an important tool in measuring educational performance. For example, Pedulla, Abrams, Madaus, Russell, Ramos and Miao (2003) argued that

> Education reform efforts since 1983 have generally had three main components: (1) educational goals or standards, (2) a test designed to measure the degree to which these goals have been achieved, and (3) high stakes attached to the results, which are intended to influence the behavior of teachers and students. (p. 10)

It becomes clear that an examination of both teacher and administrator attitudes around the use of standardized testing and performance could provide more insight as we further our efforts in education reform. Lastly, there is significant value in the notion that policy makers can have

access to information that provides us with insight into teacher and administrator attitudes on the role of standardized testing as a measure of performance.

Organization of Study

When considering the organization of this study, along with this chapter, are four additional chapters. Chapter 2 includes a comprehensive literature review of the history of accountability and teacher evaluation in American education that highlights a movement toward increased use of standardized tests as a measurement of both school and teacher effectiveness. Both federal and state initiatives are identified in order to clearly show how the public system has been measured in terms of effectiveness based upon standardized test results. Chapter 2 also identifies and discusses significant studies that have analyzed the use of standardized test results for the purpose of making determinations about teacher performance. The studies included serve to clearly show both perspectives with regard to using standardized test results on evaluation instruments. Studies also illustrate that an effective method for measuring teacher performance based upon standardized testing data has not yet been identified and that current models are frequently deemed ineffective in attempting to draw conclusions about teacher performance.

Chapter 3 describes the methodological design of the study and provides detailed information about the process and instruments used to gather data from both teachers and administrators. The sample descriptions and the procedures utilized to select members of the sample population are also included in Chapter 3. Chapter 4 includes thorough analyses of the data generated through a mixed-method design as well as findings. Chapter 5 provides a summary of the study, conclusions drawn as a result of this research, implications, as well as recommendations for further research. The final section of this research includes both a bibliography and appendices that support all aspects of this dissertation.

Summary

It is important to note that a study of this nature is intended to describe the attitudes of professionals in the field of education that may assist in guiding policy and legislation as related to how standardized testing results can be utilized to measure the effectiveness of teachers in the classroom. According to Bergin (2015),

> quantifying student learning and then connecting the quantity of learning to specific teachers is far from simple. While there is consensus that student learning, in addition to quality of teacher practice, should be part of teacher evaluation, there is no consensus about how to include measures of student learning. (p. 1)

The trend in education over the last two decades has emphasized increased accountability from both the federal and state level. For example, Bergin (2015), stated that "States and districts vary enormously in the extent to which student achievement data is weighted in the evaluation process, from 10% to 50% of the evaluation formula" (p. 5). This historic trend is aimed at improving the overall system of education in America; however, there has not yet been success in identifying proven methods that actually use standardized test results to effectively measure teacher performance. It is essential to create multiple opportunities that examine the attitudes of educators as it relates to effective measures of their own performance. This serves to create more opportunities for studies of this nature to identify possible trends that will help guide in establishing effective measures of performance that will serve to benefit educators and ultimately all students.

CHAPTER 2

Review of the Literature

Introduction

The chapter will provide a review of the literature as related to the history and evolution of both accountability and teacher evaluation in the United States. A historical perspective on the development of teacher evaluation will serve to highlight how accountability has changed over time and how the educational legislation in America has shaped the current views of measuring teacher performance through evaluation. This historical perspective of both accountability and changes in teacher evaluation leading up to RTTT legislation illustrates how the system has transformed throughout history. The chapter will also include literature that describes the current trends in the United States regarding teacher evaluation. Lastly, research on what is taking place specifically in Arizona concerning teacher evaluation will provide context for this study and insight into the approach taken in the state.

Educational Accountability in the United States

When considering the historical perspective of educational accountability, one must actually look at early efforts in America about educational reform. According to Ellis (2007), "the launch of Sputnik I October 4, 1957 changed the world and education forever" (p. 222). Ellis noted that Sputnik spurred competition between the American and Soviet powers with regard to the space race. The perceived Soviet superiority led to the realization by American society that they were falling behind, especially in science and math education. As with most

government initiatives and the existence of the cold war, this was a reaction to a global event that impacted federal involvement in education. In his analysis, Ellis (2007) pointed out:

> Even though the federal government has no constitutional authority in the area of education, the impact of Sputnik placed education front and center in the mind of the public and created a mindset for the federal government's involvement in public education. (p. 222)

Sputnik was a historical event that provided a pathway for future educational reform. This can be evidenced through subsequent legislation that emerged in America.

The *Elementary and Secondary Education Act* (ESEA) of 1965 is described by Marsh and Willis (2003) as "the beginning of a period of unprecedented federal activism in education" (p. 237). The ESEA served to expand the federal government's role in education and since its inception in 1965 has expanded and evolved over the years. This evolution is described by Ellis (2007) in the following manner, "over the last 4 decades, the ESEA 1965 has been amended 42 times including the last and most extensive amendment, NCLB" (p. 224).

As NCLB is considered, a definitive trend in how the federal government increased its authority over education in America when considering standardized testing is seen. The composition of NCLB furthers the narrative with regard to how the federal government's role has served to influence accountability in education. For example, NCLB made it mandatory for states to administer tests that would serve to directly measure student performance in relation to state academic standards. The focus of NCLB was to label and identify schools that, according to Dee and Jacob (2009), were "failing to make 'adequate yearly progress' (AYP) towards the stated goal of having all students achieve proficiency in reading and math by 2013-14 and to institute sanctions and rewards based on each school's AYP status" (p. 3). The quote highlights

the influence of this legislation in how it incorporates mandatory testing as a primary tool to measure school effectiveness and student learning. Neill (2016), further described the impact in relation to standardized testing and NCLB in that,

> It required statewide tests in grades three through eight, as well as reading and math tests in one year of high school, and science tests in three. It mandated rigid, draconian rates of improvement (all students were to score 'proficient' by 2014) and a series of punitive, escalating sanctions for failure to improve quickly enough. (p. 12)

Provisions in NCLB of this nature provide insight into how student achievement data, specifically standardized testing results, are viewed as the central tool for making a determination about the degree of effectiveness as related to schools and districts.

While testing is a large component of NCLB accountability measures, this legislation also was directly connected to funding sources for schools as well as curriculum. If funding was to be awarded, each state was required to submit a plan that, according to Ellis (2007), described how it would "ensure that high-quality academic assessments, accountability systems, teacher preparation and training, curriculum, and instructional materials are aligned with challenging State academic standards so that students, teachers, parents, and administrators can measure progress against common expectations for student academic achievement" (p. 224).

Additional curricular stipulations also emerged as a result of NCLB. These provisions represented oversight in relation to the curriculum that states utilized for instructional purposes. Ellis (2007) explained that any funds associated with NCLB could not be used for curricular materials unless they satisfied the requirement that it be connected to scientifically based instructional strategies (p. 224). While schools and districts still had choice with regard to the implementation of curriculum, this detail highlights the manner in which choices were narrowed

as a result of NCLB. Ellis (2007) further described this provision, "NCLB represents the first time in 40 years of federal involvement with local education that the federal government has attempted to dictate curriculum" (p. 225).

There is further evidence of federal involvement through the existence of RTTT, which was established in 2009 and emerged as a result of political disagreement with regard to educational reform and the much debated success of NCLB. According to McGuinn (2014), "faced with divided control and partisan gridlock in Congress—which has been unable to reauthorize the ESEA, the largest federal education program—the Obama administration has opted to make education policy from the executive branch" (p. 61). RTTT emerged as the next federal legislation that would serve to govern the United States educational system. As McGuinn (2014) described, RTTT is a competitive grant program which allows states to complete a waiver of NCLB and referred to this program as one that awards states grant funding for "for developing effective school reforms that are in line with federal goals and approaches" (p. 64).

A review of the RTTT grant components and expectations serves to illustrate how the federal government is able to connect funding to the initiatives and political ideals that exist for the educational system. McGuinn (2014) provided an overview of the grant application for RTTT which revealed the federal government's focus and priorities for education. There are six components scored within the application:

1. state success factors,
2. standards and assessments,
3. data systems to support instruction,
4. great teachers and leaders,

5. turning around the lowest-achieving schools, and
6. general (p. 65).

It is also important to note that the broad categories are broken down into 19 more detailed categories that expand on how federal reforms are to be carried out. The application sections for RTTT are also assigned a point value in order to provide a way to score applications with the end result being approval or denial.

The analysis of the grant application creates a perspective on how RTTT furthers the federal approach to accountability. It also highlights the evolution of teacher evaluation and how federal legislation begins to influence the structure of performance evaluations. Additionally, the accountability focus in NCLB was quite different from RTTT; for example, the focus for NCLB was on mandated testing in order to measure school and district success. In reviewing RTTT, there is a very different focus that leads to a clear understanding of the federal emphasis. According to McGuinn (2014), the percentage of points is allocated in a very specific manner related to the six broad categories: State Success Factors (25%), Standards and Assessments (14%), Data Systems (9%), Teachers and Leaders (28%), School Turnaround (10%), and General (14%) (p. 66). What is most telling about the percentage distribution is the emphasis on the category of Teachers and Leaders, representing the highest percentage of points, which results in drawing the conclusion that this section is an emphasis of federal reform. The focus on the Teachers and Leaders section in the RTTT application is interesting when considering McGuinn's (2014) description, "The section on Teachers and Leaders (28%) pushed states to develop better teacher and principal training, evaluation, and distribution all with a focus on student performance" (p. 65). A clear priority is set with regard to the connection between using student achievement data and measuring individual teacher performance.

Teacher evaluation becomes a polarizing subject as it's considered one of the key focus points of RTTT. The impact that RTTT makes within this context of education is astounding and marks a major shift in relation to accountability. Neill (2016) described the impact as the:

> primary result of the RTTT deal was a further rapid increase in testing students, because the waivers required every teacher to be judged by state or local test scores, but states had few tests in subjects other than those mandated by NCLB, and no state had them in every subject or grade. (p. 14)

The focus of RTTT directs state education agencies to emphasize the performance of individual teachers. McGuinn (2014) explained:

> Perhaps no issue better represents RTTT's potential to drive changes in discourse, politics, and policy—as well as its limitations—than teacher accountability. Numerous studies have demonstrated how existing state teacher evaluation, tenure, and dismissal policies are broken and have impeded efforts to improve teacher quality and student achievement. Prior to RTTT, the norm across the country was to give teachers tenure automatically after 3 years in the classroom, with no meaningful evaluation of their teaching effectiveness and little risk of their being fired during their career no matter how ineffective they are. Despite the abundant evidence that major evaluation and tenure reform was necessary, virtually no state had taken serious, sustained action before RTTT. (p. 71)

McGuinn (2014) furthered this stance, noting that the financial support along with the cited rhetoric has created astonishing rates of change when considering teacher evaluation systems reforms (p. 73).

In looking at both the development of NCLB and RTTT, a clear picture emerges in how accountability for educators has been shaped over the past decade. The emphasis of NCLB served to elevate the importance of using student achievement data from mandated standardized tests to measure the effectiveness of schools and districts. This higher level of accountability has evolved through the implementation of RTTT by shifting its focus to individual teachers. The discussion of these two major reform initiatives represents the increased involvement of the federal government in education. Additionally, the implementation of ESSA or the Every Student Succeeds Act took place after this study and, as a result, was not included as of the literature review.

History/Evolution of Teacher Evaluation

The history of teacher evaluation runs parallel to the development of the United States education system. As schools developed, the need to evaluate those that were responsible for providing instruction also grew; however, the early purpose of schools influenced who actually evaluated educators. According to Jewell (2017),

> Public schools were established in 1647 in Massachusetts in an effort to equip children with the ability to read and understand the principles of religion and capital laws of this country. Massachusetts was a pioneer in early education…outside of New England, schooling opportunities varied widely. (p. 371)

Jewell (2017) describes early schools as those with a single classroom structure having students from varying grade ranges with a curricular emphasis on discipline and memorization. Local community leaders were charged with developing the academic focus, where "the family, society, and religious institutions were of primary significance" (p. 371). The overall prominence of many schools, according to Jewell (2017), was often times directly related to

religion and the emphasis of teacher effectiveness was minimal as moral standing and the appropriateness of curriculum took center stage (p. 372). During this time teaching was not determined to be a professional field or occupation, and according to Marzano, Frontier, and Livingston (2011), local government and clergy were responsible for hiring teachers and evaluating instructional performance (p. 12). Marzano et al. (2011) elaborated on the notion that the responsibility for evaluation of teacher performance was the responsibility of religious officials in that "clergy were considered to be logical choices for this role because of their extensive education and presumed ability to guide religious instruction in schools" (p. 12). Therefore, the priorities of the community governed the focus of those charged with assessing teacher performance.

The education system for measuring teacher quality took on a very different approach from what takes place in schools today. Rather than emphasizing student achievement, "the teacher evaluation process in early colonies was primarily a system of inspection. Many teachers in these early years of public education did not have a great deal more education than their students" (Jewell, 2017, p. 372). Therefore, teacher efficacy was a construct that connected to the essential characteristics determined by communities and religion; "A teacher's effectiveness was evaluated through community and religious mores rather than through quality of instruction and student achievement" (Jewell, 2017, p. 372). As a result, the primary outcome for teachers who failed to perform in accordance with the ideals of the community in which they resided and worked was dismissal. Opportunities for feedback and training in order to facilitate improvement were not provided or extensively considered in the early education system, "Teachers were expected to be immediately responsive to the direction of community leaders who had the power to terminate the teacher for any infraction" (Jewell, 2017, p. 372). The

practice described is illustrative of a rudimentary system of supervision that emphasized inspection. Marzano et al. (2011) described the approach to supervision in that,

> Supervisors had nearly unlimited power to establish criteria for effective instruction and to hire and fire teachers. Because there was no necessary agreement as to the importance or nature of pedagogical expertise, the quality and type of feedback to teachers was highly varied. (p. 12)

This early time in education can be best summarized by Tracy (1995) when considering teacher evaluation:

> First, it was assumed that supervisors had a right to intervene directly in the classroom; local and state legislation reinforced this assumption. Second, it was assumed that the teacher was the servant of the community and, as such, should be expected to respond to the community's directives. Third, the criteria for effective instruction were established by the community. Effectiveness was defined in terms of the desired out-comes--students who could read the Scriptures and who adopted the mores of the community. The power vested in the committee to immediately dismiss the teacher meant that the observers' suggestions were meant to be taken seriously. (p. 320)

While the system of evaluation served a very basic purpose in early schools, the 1800s marked a shift in how the American education system viewed teachers and evaluation of their performance. Marzano et al. (2011) described the nature of this change:

> A rising industrial base and the common schooling movement that extended through the 1800s spawned large urban areas with more complex school systems. In these larger schools and districts, a demand grew for teachers who held expertise in specific disciplines and for administrators who could assume more complex roles. One teacher

> within a building was often selected to assume administrative duties…and ultimately grew into the role of building principal. (p. 13)

The Marzano et al. (2011) anecdote highlighted how the Industrial Revolution impacted education in a variety of ways, which influenced the need for teachers as professionals rather than servants of the community. Jewell (2017), described this phenomenon, "between 1820 and 1860, the Industrial Revolution increased urban development…This resulted in the need for better-educated teachers with better training by an expert, or 'principal', teacher" (p. 373). The need for more, educated teachers spurred reform efforts in education and in the 1840s, the call for education reform targeted teachers in a way that focused on limiting their autonomy and creating more administrative control (Jewell, 2017, p. 373). Some experts refer to this time in education as the professionalism phase. For example, according to Tracy (1995),

> The professionalization phase of assisting and assessing began with the end of the community accountability phase and lasted through most of the 1800s, when the responsibility for the overall operation of schools shifted from community leadership to a cadre of professional educators. (p. 320)

This approach led to two important shifts in education. The first was that supervisors began to focus on the improvement of instruction. Blumberg (1985) described this shift, "Much as the development of good schools was seen as central to the development of the local community and the nation itself, so was the position of teacher held in high esteem by the supervisors" (p. 58). A clear evolution exists and many of the statements made by researchers indicate that the role of educator, at this time, is seen as more professional in nature.

> While superintendents initially inspected schools to see that teachers were following the prescribed curriculum and that students were able to recite their lessons, the

> multiplication of schools soon made this an impossible task for superintendents and the job was delegated to the school principal. (Starratt, n.d., p. 1)

Thus, the second shift is that local observation of teacher performance, by a specific expert, is introduced as a mechanism to support the development of educators rather than for dismissal purposes.

The 1800s continued to mark a period of time where both school systems and the notion of evaluating teacher performance evolved, "As the numbers of schools grew, so too did the interest in improving teacher pedagogy" (Jewell, 2017, p. 374). While the idea of developing instructional practice became important, Marzano et al. (2011) also explained that:

> The period from the beginning of formal education in the United States up to the mid-1800's saw the dawning of the awareness that pedagogical skills are a necessary component of effective teaching. Although there was little or no formal discussion about these skills. (p. 13)

The evolution of educational ideals also played a role in shaping the assessment of teacher performance. Significant historical figures like Horace Mann developed the idea of formalizing school and teacher training. For example, Mann lobbied to make school attendance a requirement and was more concerned that students learn both socially and develop moral character (Jewell, 2017, p. 374). The mid 1800s also marked a time where Mann's passion for education furthered the need to develop the skill set of teachers, thus the emergence of the Normal School. According to Jeynes (2007),

> Mann was concerned about the American perception that city schools were of far greater quality than rural schools, and he claimed that if schools had a common curriculum,

> educational leaders could found teacher institutes that could train teachers to be effective no matter which common school they taught in. (p. 149)

Jewell (2017) elaborates on the function of normal schools as well, "In these schools, students honed their teaching skills through observation and feedback" (p. 374). While training and feedback were emerging practices in the field of education, it can also be seen that, as a whole, these tools were not well-established when considering a societal perspective, "While administrative oversight was the accepted model for teacher evaluation, this oversight was limited. For example, in 1890, the city of Baltimore had two superintendents to oversee the entire district, which employed 1,200 teachers" (Jewell, 2017, p. 375). Jewell's historical perspective illustrates the limited and varied implementation of teacher evaluation practices in America during this time.

The late 19th and early 20th centuries continued to mark an era of greater development when considering the form and function of teacher evaluation in American education. Marzano et al. (2011) referred to this time in history as the Period of Scientific Management (p. 14). Other researchers in the field have also offered descriptions of this period; Tracy (1995) elaborated by stating,

> Scientific supervision was the concept of measuring the methods of teaching to determine the most productive ones in relation to student outcomes. The emphasis on measurement led to increased attention to direct classroom observation and data gathering, particularly through use of an observation checklist, a tool commonly used today. (p. 320)

This period was a time that included two views of education which, in many ways ran counter to one another (Marzano et al., 2011). The first of those views belonged to John Dewey. While John Dewey is known as a prominent educational thinker in American history, his philosophical

approach aligned directly to the development of democracy through schools so that students could practice citizenship.

> Dewey thought that educational leaders—teachers, principals, superintendents, and so on—should facilitate and manage the resources, energy, focus, and engagement of the school organization toward the freeing of intelligence for the well-being of students to promote their becoming individuals who live democratically with one another. (Dewey & Simpson 2010, p. 119)

The second view is described by Jewell (2017), "between 1900 and 1920, it was proposed that teaching could be measured and made more efficient using successful business productivity models" (p. 378). Ideas of this nature belonged to Frederick Taylor who, according to Marzano et al. (2011), "believed that measurement of specific behaviors of factory workers was perhaps the most powerful means to improve production" (p. 13). Marzano et al. (2011) furthered Taylor's scientific approach by explaining that this way of thinking "resonated with engineers and business owners, and colleges of engineering and were well positioned to infuse is principles into their course" (p. 13).

The K-12 system began to experience the scientific approach through Ellwood Cubberley who described how Taylor's model "could be applied when visiting teachers' classrooms. He described specific feedback that a supervisor might provide to a teacher" (as cited in Marzano et al., 2011, p. 14). For example, Cubberley provided a rating form that included a scale of A to F, where a 6th grade teacher was given a D for her arithmetic lesson. Feedback provided by Marzano et al. (2011):

> Weak Points: Entirely wrong procedure for type of problems used. No attempt at problem-solving instruction.…

> Suggestions Made: Explained to her that, being a new teacher to our schools, she evidently did not know how we taught Arithmetic. Explained faults of the lesson, but commended her managerial ability. Told her how she should handle such work, and gave her Newcomb's *Modern Methods of Teaching Arithmetic* to take home and read designated chapters. (p. 14)

This description provides insight into the scientific approach with regard to how feedback was to be provided to teachers regarding instructional performance pointing out the significance of the system and how it may have impacted more recent approaches to evaluation of teacher performance. Marzano et al. (2011), identified that "through the 1930s, there was continued tension between the scientific approach to schooling, including a greater reliance on standardized testing and the approach that focused on social development and democratic values" (p. 15). They also pointed out that "Cubberley and Wetzel's recommendations might be considered precursors to some of our recommendations regarding the use of data for feedback" (p. 15). It becomes evident that these two approaches have an important influence on prevailing philosophies of education which impacts the approach of how teachers were to receive feedback about their classroom performance. Marzano et al. (2011) summed it up,

> One can use data for feedback but still maintain the goal of an education system that fosters democratic ideals. Nonetheless, the two perspectives were not described or perceived in a fashion that allowed for integration, and the tension between them continued through the Great Depression. (p. 15)

During this time period it became clear that business productivity models had made an impact on the approach to supervision of teachers and measuring performance. Jewell (2017), highlights this impact and how evaluation changed,

> In contrast to the early colonial model in which teachers were expected to perform well or suffer the consequences, the objective evaluation model required teachers and administrators to work together to improve the overall quality of the teachers' skills; the goal was retention and improvement rather than dismissal. (p. 379)

As the scientific management period in education eventually drew to a close, there were significant historical events that shifted the focus of education, which ultimately impacted the approach to supervision. The end of World War II marked an important time, "the G.I. Bill was sending greater numbers of students than ever to college…in the post-war world there was great demand for math and science skills to meet the industry and governmental needs" (Jewell, 2017, p. 379). Jewell further describes *Brown v. The Board of Education* as a landmark decision that created a focus on school inequality across America and "opportunities for African-American children became imperative" (p. 380). As the complexity of schools and the teaching profession grew, the mid-20th century brought with it a more individually focused approach to supervision of teachers. Tracy (1995), referred to this period as the Human Relations Phase, noting:

> The 1930s and early 1940s saw the pendulum swing from a scientific perspective focusing on achievement of organizational goals to a human relations perspective that focused on the individuals within the organization. Oversight of instruction became conceived of as a form of guidance rather than direction of instruction. (p. 320)

Robert Marzano et al. (2011) further confirmed this shift in education by stating,

> The period immediately after World War II began with a swing away from the scientific approach to schooling. Emphasis was placed on not only assisting the teacher to develop his or her unique skills, but also tending to his or her emotional needs. (p. 16)

The individual teacher certainly became an emphasis and has been characterized in a variety of ways. Jewell (2017) describes this characterization by noting that "if teachers were treated as valued partners in the educational process, improved teaching quality would automatically result" (p. 380). This new approach is also referred to as the cooperation model and was emphasized as "teacher participation and autonomy" (Jewell, 2017, p. 381).

A number of educational publications during this timeframe served as evidence for the emergence of a clear trend and focus on the individual teacher. For example, a 1946 edition of *Educational Leadership* provided insight into this education approach. In the article by Hankamp (1946) entitled, *Are Teachers People?* He stated,

> We believe that a careful consideration of the basic human needs and rights of teachers is an important educational issue at the present time. In this number of Educational Leadership-through the statements of educators in many and varied positions—the case is presented. (p. 250)

His article certainly highlights the emphasis during this time period in education along with advocacy towards the idea that strong relationships must exist within the context of leadership responsibilities. In the same issue of *Educational Leadership* there is an article written by Willard E. Goslin (1946) entitled *Know Your Teacher* where he stated that "one of the most important aspects of a leader's work is the continuous endeavor to know and to understand those

with whom he works" (p. 260). The article also provided insight into the attitude that teachers need to be understood like any other human. Goslin (1946) explained:

> The school administrator needs to realize that teachers *are* human beings and that they bring with them into their schoolrooms all their human frailties, all the ups and downs of their physical and mental health, all their varied interests and enthusiasms, as well as their lesson plans and their teaching techniques. (p. 260)

Although this literature provides insight into the theme that understanding the teacher as a human being during this era is crucial, it is also important to examine what was being discussed regarding supervision and teacher evaluation within this context. Educational literature related to teacher evaluation in mid-20th century provided more development of this concept; for example, according to Melchior (1950), "two major functions of professional leadership are commonly accepted—administration and instructional supervision" (p. 5). When digging more deeply into Melchior's work, much of the text is focused on the role of administration which included such functions as "getting citizens to provide buildings and grounds; it continues with through maintenance, securing teachers and supplies, and providing general oversight of the situation in which teacher and pupil are together in a classroom" (1950, p. 4). One area initially absent was the description of the importance of supervision as related to instructional supervision and teacher evaluation; however, Melchior (1950) made reference to this function in a subsequent chapter and described it as, "good supervision is good teaching" (p. 6). There is also the cooperative approach described in Jewell's (2017) work with regard to teacher evaluation that is evident in Melchior's work. Melchior (1950) described instructional supervision:

> In a supervisory program teachers and supervisors work together and play together as do classroom pupils and their teachers. A supervisor's activities, such as group and

> individual conferences, exchange of ideas, preparation of written and oral work, visitation, and the use of instructional materials, may be compared to the classroom and outside-classroom activities of pupils and teachers. (pp. 4 - 5)

It becomes clear that the relationship between teacher and evaluator is compared to that of the student and teacher relationship. Furthermore, this relationship can be characterized as being cooperative in nature. In early chapters of his text, Melchior (1950) described supervision as a philosophical approach and as a tool to develop the teacher and students to be good workers and ultimately good citizens (p. 8). In later chapters, he spent time describing the idea of classroom visitations and more specifically classroom observations. The description provides great understanding of what classroom observation looked like at the time as well as the purpose it served, "Classroom visitation as practiced in the old days by supervisory officer is disappearing. In fact the modern supervisor frowns upon the use of the world, especially when it is accompanied by the term 'demonstration'" (Melchior, 1950, p. 364). Melchior explained that a teacher's feeling towards being observed, whether positive or negative, is greatly dependent upon the supervisor's approach and relationship with said teacher. According to Melchior (1950), "the fault lies with the supervisor if teachers do not want to be observed or do not want to observe the supervisor at work with children" (p. 364). This again highlights a teacher centered approach focused on addressing the individual needs of the educator within the context of classroom observation. When considering the process of visitation, Melchior (1950) described a number of observational scenarios where a supervisor is invited by the teacher to visit the

classroom and work together in order to influence instruction (p. 366). He noted that an effective program of teacher supervision included:

> Observation of a closely similar situation—age level or subject-matter area—for a particular phase of the work or for more inclusive purposes; observation of a grade level below or above the one being taught by the visiting teacher, for purposes of continuity of the curriculum and teaching procedures; and observation of a class in a subject-matter area other than the visitor's, to see how subject matter is related. In other words, the program includes both vertical and horizontal correlation. (p. 365)

The cooperative approach to evaluation is evident in other works during this time period. For example, in Coleman's 1945 *Educational Leadership* article, *The "Supervisory Visit"*, the approach was characterized:

> The supervisor today recognizes the complexity of a visit to the teacher, and utilizes every possibility to make it a mutually satisfying and worthwhile experience. Such a concept of the 'supervisory visit' involves understanding the individual teacher and building readiness for supervision. (pp. 164 - 165)

It is clear that a positive relationship between supervisor and teacher is a primary focus when looking to improve performance in the classroom. This idea was quite evident in one of Coleman's (1945) concluding statements, "The supervisor merely supplements in whatever ways may seem best and never supersedes the teacher" (p. 167). During this time in education it was clear that the relationship with teacher was essential as classroom performance was measured. As a result, there were also outcomes that could be considered negative with regard to the

manner in which supervisors conducted classroom observations. Tracy (1995) expounded on these outcomes when he stated,

> Supervisors concentrated on building positive relationships with teachers. Unfortunately, an outcome of this relational emphasis was that supervisors sometimes feared upsetting the relationship by conducting direct classroom observation. Thus, in practice, human relations supervision all too often equated with hands-off supervision, where little actual assistance was provided. (p. 320)

As a result, the need for change began to emerge from the human relations or cooperation phase of performance evaluation with regard to how teacher assessment took place in the education field. For example, Whitehead's (1952) study chronicled the need for change, "teachers in the state of North Carolina have looked frankly and sincerely at this business of supervision, and are of the opinion that administrators should pay more attention to the chief aim of education—effective teaching" (p. 106).

As the transition away from the human relations or cooperative period of supervision took place, Tracy (1995) contended, "Over the next several years, each dominant supervisory practice represented a reaction to the previous phase. The post-human relations phase was no exception" (p. 320). Tracy claimed that a second scientific phase emerged which was similar to that of the previous phase, which was characterized by "techniques for observing and recording what occurred in the classroom would provide data that could stimulate instructional improvement" (p. 320). Additionally, one may recall a time where the identification of instructional methods was connected to teacher behaviors and viewed as a tool to enhance and improve performance. According to Gillis (2015), "With the 1950s came an era of research on teaching methods and an effort to link student outcomes with particular teaching behaviors.

Classroom observations and checklists identifying favored teaching behaviors became common components of local teacher evaluation approaches" (p. 26).

As teacher evaluation continued to evolve, one of most prominent eras began in the late 1950s resulting in a significant impact in the educational community. Marzano et al. (2011) characterized this era by stating, "few innovations in the field of education spread as quickly as clinical supervision" (p. 17). Clinical supervision was initially developed by "Morris Cogan, a professor at Harvard's Masters of Arts in Teaching program in the 1950s. He and his colleagues developed a systematic approach to working with student teachers" (Marzano et al., 2011, p. 17). According to multiple researchers, Robert Goldhammer who was a student of Cogan's, was also responsible for the development of this approach to supervision.

> The emergence of clinical supervision in the late 1960s attempted to combine the tools and the techniques of the scientific phases with the supervisor/teacher team approach of the human relations phase. Borrowing from the relational and motivational concerns of the human relations phase, clinical supervision required sustained teacher and supervisor interactions in order to mutually solve classroom problems. In the work of both Cogan and Goldhammer, this interaction was to occur during pre-and post-observation conferences. (Tracy, 1995, p. 320)

Goldhammer also developed the concept of clinical supervision and maintained that the model was fashioned after practices implemented in teaching hospitals. Goldhammer believed that "the process involved a purposeful, symbiotic relationship between practitioner and resident, where observation and discussion drove both parties to higher levels of growth and effectiveness" (as cited in Marzano et al., 2011, p. 18). In order to better describe the approach involved in Clinical Supervision, it is important to consider the direct work of Goldhammer. In the text by

Goldhammer (1969), *Clinical Supervision: Special Methods for the Supervision of Teachers*, a process was outlined that clearly highlighted the approach taken on the model. The model included a five-phase procedure inclusive of: a pre-observation process, classroom observation, analysis, supervision conference, and analysis of the data collected (Goldhammer, 1969, p. 60). These phases involved actions that take with the teacher and supervisor. Reavis (1976) provided a detailed description of the different phases of the clinical supervision model. As noted, the pre-observation process occurs first, and according to Reavis (1976),

> The purposes of this conference are to establish rapport, get an orientation to the group the supervisor will be observing, receive information on the lesson to be taught, suggest minor changes that might improve the lesson, and develop a contract, that is, an agreement between teacher and supervisor about the purpose of the lesson. (p. 361)

The observation serves as an opportunity for the evaluator to write down any notes associated with the lesson and ultimately becomes the foundation for later discussion. The analysis phase is described as the time where the supervisor analyzes the data collected, looking for patterns in teacher exchanges. Reavis (1976) further elaborated that "the supervisor must clear his or her mind of all pet theories and biases and deal directly with the day" (p. 361). The next step is the conference between supervisor and teacher where the focus is on what the teacher may have indicated were concerns in the pre-observation meeting. This is also where plan remediation is developed between both parties. Reavis (1976) concluded that "the final step in the sequence is the post-conference analysis…supervisor reviews actions taken in each of the preceding step with regard to whether they facilitated improved instruction and teacher growth" (p. 361).

During the time in educational history where this model for evaluation was prominent, researchers touted Clinical Supervision as a tool that would serve to greatly impact instruction in

a positive manner. For example, Reavis (1976) concluded that "Clinical supervision, providing clarity and specificity in in-class supervision, has the potential to accomplish what all evaluation attempts—to improve the quality of instruction provided to children" (p. 363). While other approaches emerged, clinical supervision also demonstrated the ability to stand the test of time. In Tracy's (1995) article, *How Historical Concepts of Supervision Relate to Supervisory Practices Today*, it was noted that "The assumptions were that a sustained cycle of assistance is necessary for teaching to improve and that the analysis of teaching behavior patterns can lead to useful insights" (p. 320).

At the turn of the 20th century, evidence of clinical supervision still surfaced as a viable format for teacher evaluation models. According to Pajak (2001),

> Classroom observation and feedback have been mainstays in the clinical supervision of both preservice and in-service teaching for many years and are likely to continue to play an important part in the ongoing quest to further the professional growth of beginning and experienced teachers. (p. 233)

While clinical supervision has maintained an enduring impact, the need for an evolution of systems to measure teacher performance also continued to rise. It became apparent that flaws existed with regard to its purpose and implementation; according to Marzano et al. (2011), "few models in the entire field of education—let alone in the specific domain of educational supervision—have been as widely deployed, as widely disparaged, or as widely misunderstood" (p. 18). Goldhammer's five phases ultimately became the primary focus and system for how

evaluations were organized, but did not include the rich dialogue intended by Goldhammer Marzano et al. (2011).

> In some cases, the rich, trusting dialogue envisioned by Goldhammer was reduced to a ritualistic set of steps to be followed. Regardless of the reasons for its demise, Goldhammer's vision of supervision as a…quest for more effective instructional practices quickly disappeared. (p. 19)

It is clear that clinical supervision would not be the last evaluative tactic to influence the educational method to teacher evaluation.

A significant body of work that emerged in the educational world following clinical supervision was during the 1980s from Madeline Hunter. According to Marzano et al. (2011), "the next major influence on supervision was the work of Madeline Hunter" (p. 20). While her work had an enormous impact on the supervision of teachers, the actual focus of Hunter was:

> designed for the explicit purpose of having students get it right the first time through. Erroneously some school administrators have used the model to analyze teaching performances…during her lifetime, Dr. Hunter was emphatic that it was never the intention that her model should be used as a teacher evaluation tool. (Wilson, 2017, p. 1)

What Hunter did design was a model for effective instruction that served to impact the instructional practices of classroom teachers in relation to planning and delivery. Hunter (1976) described the intent of this model, "It is important to state that our work was not basic research but the identification and labeling of teaching decisions and actions which incorporated principles that *research already had established as having the potential to affect learning*" (p. 163). Hunter's work was inclusive of a lesson design that included seven different components including: anticipatory set, objective and purpose, input, modeling, checking for understanding,

guided practice, and independent practice (Marzano et al., 2011, p. 20). These elements were to be incorporated into lessons by teachers as a way to deliver instruction to students. There were significant benefits to Hunter's work in that specific teaching strategies and instructional design were identified and could be applied by the classroom teachers. Orange (2002) provided a description of these benefits, "She had labeled the techniques and explained the underlying psychological theory of why these techniques work. Hunter's model is prescriptive in that it outlines a way to teach in a conscious and deliberate fashion to increase student learning" (p. 103).

In addition to Hunter's seven steps, there were other contributions she made as a result of her work within the context of teacher supervision. While the labeling of specific strategies had a significant influence on instruction and supervision, Hunter also advocated for observation and scripting of lessons. Marzano et al. (2011) stated that "observation and script taping were critical components of Hunter's process of supervision. During script taping, a supervisor recorded teaching behaviors and then later categorized them into those that promoted learning" (p. 20). The intent was for supervisors to conference with teachers in order to discuss teacher behaviors using objective, concrete data as a tool to influence practice. Hunter (1983) also provided a publication that emphasized the importance of script taping as a process that could assist the evaluator. She described how scripting a lesson could serve as a benefit to supervisors in that "It (script taping) is the process of capturing with pen and pad what happened in an observed segment of teaching so that cause-effect relationships can later be examined" (p. 43). Hunter went on to say, "The fundamental purpose of all supervision is to accelerate the growth of those who are supervised" (p. 43). While Hunter's 7-step lesson design process was not necessarily

meant to be a tool for evaluation of teachers, she certainly developed work that served to aid in support of the effective supervision of teachers.

As with any major initiative, criticisms and implementation issues also emerged with regard to Hunter's work. Often times, when considering implementation and significantly impactful movements, models like Hunter's can ultimately be used in ways that they not originally intended. Wilson (2017) emphasizes how implementation of the Hunter model was not well-suited for evaluative purposes.

> Erroneously some school administrators have used the model to analyze teaching performances. Please note that during her lifetime, Dr. Hunter was emphatic that it was never the intention that her model should be used as a teacher evaluation tool. Indeed, as a seasoned educator I am sure Hunter was aware that there are many great models of teaching other than her own, and that teaching is both an art and a science and therefore cannot be relegated to a simple formulaic 7-step checklist. (p. 1)

Other criticisms emerged regarding Hunter's 7-step lesson design in that the overall model was cumbersome and it may not have a place in the planning process for science teachers. Hunter published a number of articles in response to these criticisms, which characterized the nature of the concerns with regard to her model. In Hunter's 1990-1991 *Educational Leadership* article entitled, *Lesson Design Helps Achieve the Goals of Science Instruction,* she explained that "Even though Berg and Clough quote me in describing my model as 'deceptively simple in conceptualization, incredibly complex in application,' they make the most unsophisticated and incorrect interpretations" (p. 79). Hunter went on to explain that Berg and Clough characterized the model as overly cumbersome with regard to practical implementation.

Additionally, Wilson's (2017) analysis highlights possible drawbacks with regard to the Hunter model. For example, "the model's repetitive structure it is not appropriate for open-ended learning experiences, discovery learning sessions, or exploratory educational experiences, especially ones requiring divergent thinking skills, creative problem solving, or higher level thinking skills" (p. 1). Hunter published other articles responding to criticisms. One such article was in response to David Gibboney's critiques of the Hunter model. Hunter's (1987) introductory statement in the article provided clarity with regard to the concern and seemed to echo Wilson's previous statement as well.

> I am somewhat bewildered, however, by the position taken by Richard Gibboney. As I understand it he is concerned that there is no research base for the model, the model is nonintellectual, that, unlike him, I discriminate between curriculum and instruction. (p. 51)

This response points out, along with Wilson's description, that at times the Hunter model may have been described as a framework that did not necessarily promote higher levels of thinking for students when implemented by classroom teachers.

While Hunter's work was certainly widespread during the 1980s, a variety of researchers advocated for varied approaches that impacted teacher evaluation.

> Alternate evaluation models were also proposed in the 1980s; some emphasized individualized career development for teachers, and others proposed different types of evaluation and oversight depending on the teacher's experience, age, and developmental level, but Hunter's model was foremost. (Jewell, 2017, p. 386)

Other researchers described these alternate evaluation models differently; Marzano et al. (2011) characterized this period as the "era of Developmental/Reflective Models" and they elaborated

that in "the mid-1980s, researchers and theorists in supervision began to articulate alternate perspectives, primarily in relation to the prescription applications of clinical supervision and mastery teaching" (p. 22). Glatthorn and Holler (1987), in their study of a Maryland school district that developed an evaluation system stated, "the Calvert County, Maryland, school system has developed and implemented a differentiated teacher evaluation system that enable supervisors and administrators to collaborate closely in both rating teachers and helping them improve performance" (p. 56). The model that Glatthorn and Horn (1987) described emphasized systematic visitations of classrooms, clear definition of administrator responsibilities, and professional learning as a key component of the system. The researchers contended that this is a collaboratively developed system and is differentiated to meet the goal of improving instruction. Further, Glatthorn and Horn's (1987) description provided more insight into the aforementioned models discussed in the Marzano et al. (2011) work regarding the Development/Reflective era of teacher evaluation. The perspective on teacher evaluation during this timeframe is highlighted in the work of Darling-Hammond, Wise and Pease (1983):

> Over the last decade teacher evaluation has assumed increasing importance. The demand for accountability in education has shifted from broad issues of finance and program management to specific concerns about the quality of classroom teaching and teachers. These concerns have led to a resurgence of interest in evaluating teachers and to the development of new systems for teacher evaluation. (p. 285)

Darling-Hammond et al. (1983) conducted a review of the literature on teacher evaluation in their work and emphasized how the need to measure quality teaching was at odds with the focus of educational organizations on positive relationships with educators,

> We examine how external demands for accountability are at odds with internal organizational needs for stability and trust; how loosely coupled organizations like school systems handle these competing demands; and how teacher evaluation may affect organizational operations and teaching work. (p. 286)

A conflict existed between accountability and the cooperative/relational models of evaluation that were developed in previous eras; thus, the emergence of varying evaluation models aimed at meeting the divergent views that Darling-Hammond et al. (1983) cited were common in the 1980s.

As a result of varying ideas surrounding the concern over how best to evaluate teacher performance, the RAND group sponsored a study in order to identify what was actually taking place in the American education system. According to Marzano et al. (2011), this study sought to "determine what types of supervisory and evaluation practices were actually occurring in school districts across the United States" (p. 22). The RAND study served to evaluate teacher evaluation practices using case studies as a result of emerging trends that impacted teacher accountability. Wise, Darling-Hammond, Tyson-Bernstein, & McLaughlin (1984) noted that "the new concern for the quality of education of teachers is being translated into merit-pay, career-ladder, and master-teacher policies that presuppose the existence of effective teacher evaluation systems" (p. v). In the RAND study, Wise et al. (1984) explained that as a result of these practices, many school systems in America would be seeking to adjust and "reassess teacher evaluation practices" (p. v). The study included a survey of 32 school districts in order

to better understand evaluation practices. As a result of the survey data, the researchers then selected four case study districts; "We selected four case study districts representing diverse teacher evaluation processes and organizational environments" (Wise et al., 1984, p. vii). The case study districts were then investigated through interviews of employees within each district including superintendents, administrators, teachers, school board members, and various other stakeholders.

When considering the results of the RAND study, it was noted that when analyzing survey data, "evaluation practices differed substantially in the 32 school districts. Although the practices seemed similar in broad outline, they diverged as local implementation choices were made" (p. vi). Wise et al. (1984) elaborated, exclaiming that in spite of the many differences between the districts surveyed, there were significant problems associated with the instruments utilized to conduct performance appraisals. For example, Wise et al. (1984) stated that "respondents felt that principals lacked sufficient resolve and competence to evaluate effectively" (p. vi). There were additional problems with evaluation practices described in the study which ranged from teacher resistance, lack of consistency with evaluation, and ineffective evaluator education; however, the study also uncovered successful aspects of evaluation practices as a result of the analysis and information gathered from the case study districts. According to Wise et al. (1984), "Attention to these four factors—organizational commitment, evaluator competence, teacher-administrator collaboration, and strategic compatibility—has elevated evaluation in these districts from what is often a superficial exercise to a meaningful process that produces useful results" (p. vii).

The RAND study also served to highlight how the historical perspective during this time period elevated the importance of effective teacher evaluation. Wise et al. (1984) noted that in

"*A Nation at Risk: The Imperative for Educational Reform*, several of the commission's recommendations concerned with teaching would require teacher evaluation" (p. 1). This statement certainly provides insight into the notion that within the political arena at the time, the measurement of teacher quality was essential in improving the overall education system. The report further described how crucial teacher evaluation had become in the educational world, "proper teacher evaluation can determine whether new teachers can teach, help all teacher to improve, and indicate when a teacher can or will no longer teach effectively" (Wise et al., 1984, p. 1). Ultimately, the purpose of the RAND study was to "effectively assess teacher evaluation practices with a view to analyzing how teacher evaluation can be used to improve personnel decisions and staff development" (p. 2).

The RAND study provided a frame of reference for educational organizations to utilize as efforts at developing teacher evaluation systems continued throughout this time period in history. This study also served to provide recommendations and drew specific conclusions that were aimed at improving teacher evaluation.

> Our conclusions and recommendation constitute a set of necessary, but not sufficient, conditions for successful teacher evaluation. In practice, educational policies and procedures must be tailored to local circumstances. Consequently, these conclusions and recommendations may be thought of as heuristics, or starting strategies to be modified on the basis of local experience. (Wise et al., 1984, p. xi)

In reviewing the perspective of teacher evaluation during the late 1980s and early 1990s, it became clear that varied approaches to teacher evaluation were the norm. Not one specific tool was used to accomplish this task and many school systems utilized different approaches to determine how teacher effectiveness and performance could be measured. According to Barrett

(1986), “Literature exists to support all evaluation methods. Coker (1985) observes that the lack of consensus about evaluation issues represents the lack of knowledge about effective teaching and measurement technology” (p. 1). The nature of these statements in addition to the recommendations from the RAND study served to identify that using a varied approach to teacher evaluation could be implemented in order to meet the needs of local school districts.

The mid-1990s brought forth an important development when considering the evaluation of teacher performance.

> In 1996, a seminal work on supervision and evaluation was published by Charlotte Danielson. *Enhancing Professional Practice: A Framework for Teaching*, which was updated in 2007, was based on her work with the Educational Testing Service that focused on measuring the competence of preservice teachers. (Marzano et al., 2011, p. 23)

During this time, the Danielson model moved to the forefront of educational supervision as the primary tool that would influence teacher evaluation in many school systems. For example, Jewell (2017), states that, “*Enhancing Professional Practice. A Framework for Testing* was published by Charlotte Danielson and, because of its popularity, became the professional standard for teacher evaluation” (p. 387). The Danielson Framework,

> is a research-based set of components of instruction, aligned to the INTASC standards, and grounded in a constructivist view of learning and teaching. The complex activity of teaching is divided into 22 components (and 76 smaller elements) clustered into four domains of teaching responsibility. (The Danielson Group. n.d., p. 1)

Additionally, there are four domains that the framework is composed of which includes planning and preparation, classroom environment, instruction, and professional responsibilities. Within

each domain there are specific components that serve to describe a distinct piece of the domain, "Levels of teaching performance (rubrics) describe each component and provide a roadmap for improvement of teaching" (The Danielson Group, n.d., p. 1).

Danielson's work is considered significant within the context for instructional supervision, "The level of specificity supplied in the Danielson model provided the foundation for the most detailed and comprehensive approach to evaluation to that time" (Marzano et al., 2011, p. 23). The Danielson Group (n. d.) discussed the impact of Danielson, "Her *Framework for Teaching* has become the most widely used definition of teaching in the United States, and has been adopted as the single model, or one of several approved models, in over 20 states" (p. 1). Although Danielson's model has been touted as an effective tool significantly impacting the realm of teacher evaluation, the model is not free from criticisms. Johnson (2016), described in detail that Danielson's

> framework is now used in school districts across the nation—more frequently than her rival Robert Marzano's model—and her name is often known to invoke a sense of dread when spoken aloud to an overworked and underpaid public school teacher. (p. 44)

Additionally, in a survey of educators conducted by Barrett et al. (2016), Danielson's framework is described, "In this system, snapshots of instruction take on oversized importance as measurements of ability, devoid of context" (p. 5). As with any evaluation instrument, both support and concerns emerge for a system as well-known as Danielson's model.

The arrival of the 21st century brought with it a different perspective regarding teacher evaluation. In 2009, a report entitled *The Widget Effect* significantly changed the narrative with regard to measuring performance of educators. Weisberg, Sexton Mulhern, and Keeling (2009) stated, "Our report examines our pervasive and long-standing failure to recognize and respond to

variations in the effectiveness of teachers" (p. 32). As indicated in the report, the intent was to examine teacher evaluation systems in order to determine how they distinguished between varying levels of performance among teachers. Furthermore, Weisberg et al. (2009) described The Widget Effect stating it "describes the tendency of school districts to assume effectiveness is the same from teacher to teacher. This fallacy fosters an environment in which teachers cease to be understood as individual professionals" (p. 32). The basis for the report was on research that took place in a variety of states, across multiple school districts.

> Our report is the product of extensive research spanning 12 districts and four states. It reflects survey responses from approximately 15,000 teachers and 1,300 administrators, and it has benefited from the insight of more than 80 local and state education officials, teacher union leaders, policy makers and advocates who participated in advisory panels in each state. (Weisberg et al., 2009, p. 32)

It became clear that the results from this study were arrived at through a comprehensive approach which involved a wide range of perspectives in the educational world.

When reviewing the key aspects of this report, two areas surfaced that shaped the way teacher evaluation systems and measuring educator performance were characterized. The report stated that The Widget Effect is "characterized by institutional indifference to variations in teacher performance" (Weisberg et al., 2009, p. 33). This indifference was characterized in a number of ways through the evaluation systems that were studied, "99% of teacher the satisfactory rating and districts that use a broader range or ratings do little better—here, 94% of teacher receive one of the top two ratings" (p. 33). The researchers also noted that as a result of so many positive teacher ratings, it became far more difficult to actually identify truly outstanding teachers. Other issues were addressed in this study such as inadequate professional

development, no special attention to novice teachers, and poor performance going unaddressed. Weisberg et al. (2009) contended that these factors contributed to the reinforcement of The Widget Effect, deeming the current education system as failing to properly rate teacher with varying levels of performance (p. 33). The impact of this report served to drastically influence how teacher evaluation systems would need to be changed in order to effectively measure teacher performance, distinguishing between poor performers and high performers. Marzano et al. (2011) described the impact of The Widget Effect, "Final conclusions from the report suggested a complete overhaul of the teacher evaluation process…Clearly, by the end of the first decade of the 21st century, teacher evaluation practices were under siege" (p. 27).

Current National Teacher Evaluation

In today's current education system, teacher evaluation is a key component within the context of educational reform. The question of how to go about measuring teacher quality is accompanied by significant debate that comes from a variety of sources. According to Cohen and Goldhaber (2016),

> Most agree with the high level assertion that teacher evaluation ought to be meaningful, which entails reforming the content and structure of evaluations. Despite this general consensus, there has been a great deal of controversy surrounding the substance of proposed reforms. (p. 378)

A wide range of research exists on the topic and perspectives vary greatly across both political and educational lines. The national perspective on teacher evaluation can be summarized in a federal report by Anderson, Butler, Palmiter, and Arcaira (2016),

> Today, efforts are underway across the country to transform teacher evaluation into a useful tool for improving teaching and learning. These efforts are supported by state

statutes as well as by improvements in state and local data systems and the development of new student assessments. (p. ix)

In recent years, the movement to use student achievement data as a tool to evaluate teachers has grown stronger in the United States. Bergin (2015) explained that "the new movement is to use it (student achievement data) to evaluate individual teachers' effectiveness" (p. 1). As a result of the increased focus on using student achievement data in teacher evaluation, it is important to consider what is happening around the United States on current evaluation systems. According to the Center on Great Teachers and Leaders at the American Institute for Research, the databases on state teacher and principal evaluation policies, student data components on teacher evaluations are required by legislation in more than 30 states (Center on Great Teachers and Leaders, 2013). According to Lash, Makkonen, Tran, and Huang (2016), "as of early 2014, 40 states and the District of Columbia were using or piloting methods to evaluate teachers in part according to the amount students learn" (p. 3). Further investigation of the data yields a similar pattern with regard to legislation in the United States that mandates the use of student achievement data on teacher evaluations. Legislation mandating the use of student achievement data on evaluations has a greater impact when considering the high-stakes decisions that are made based upon the results of teacher evaluation. For example, many states attribute anywhere from 20% to 50% of their teacher evaluations to student growth and/or achievement data (Center on Great Teachers and Leaders, 2013). According to the Center on Great Teachers and Leaders at the American Institute for Research (2013), the databases on state teacher and principal evaluation policies note that 21 states use teacher evaluation systems to determine varying levels of compensation and 27 use evaluation instruments for purposes of dismissal. Hull (2013) explained that "States are now expected to evaluate teachers at least partially on the

impact they have on their students' achievement" (p. 9). Hull's report provided greater detail about the fact that "twenty-three states require or recommend that student achievement indicators comprise half of a teachers' evaluation" (p. 12). The question then becomes, what specific models are used in order to tie student achievement to teacher performance in the United States? Two specific models are widely recognized to accomplish this task.

> Statistical methods for linking scores to teacher performance can vary considerably but can be generally described in two ways, both of which attempt to capture student growth: Value-added models (VAM): Attempt to isolate the impact a teacher has on students' academic growth from other factors that impact student learning such as a student's socioeconomic status or their achievement on prior tests. Student growth percentiles (SGP): Measure how much progress a student has made relative to other students. (Hull, 2013, p. 14)

Much of today's research has revealed that the use of value-added measures and student growth percentiles in order to determine teacher efficacy is a difficult task. When considering value-added models, research has indicated that the system is capable of serving as an effective measure of teacher effectiveness, but also a number of issues exist with the system.

> Prominent among these new approaches are value-added models (VAM) for examining changes in student test scores over time. These models control for prior scores and some student characteristics known to be related to achievement when looking at score gains. When linked to individual teachers, they are sometimes promoted as measuring teacher effectiveness. (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2011, p. 2)

Corcoran (2010) stated,

> The simple fact that teachers and principals are receiving regular and timely feedback on their students' achievement is an accomplishment in and of itself, and it is hard to argue that stimulating conversation around improving student achievement is not a positive thing. (p. 8)

Lastly, Hull (2013) noted additional benefits of valued-added measures:

> …realizes not all students are likely to make the same growth from year to year; Only model that attempts to isolate the impact a teacher has on student growth More accurately identifies effective teachers compared to to other measures including Student Growth Percentiles. (p. 15)

Although notable benefits around the implementation of VAMs are evident, significant problems have also been identified through research. According to Darling-Hammond et al. (2011), the results yielded from the study of value-added models and related data used in five separate school districts indicated that teachers who performed in the bottom 20% using value-added rankings in one year did not consistently perform at the same level the following year; the study indicated that only 20%-30% of teachers had similar results the following year (p. 4). The researchers also described that the inconsistency of rankings also occurred for those who performed in the tops levels of achievement. The researchers used this data to highlight the relative instability associated with VAMs; they concluded that other considerations noted were the lack of random assignment of students to teachers as well as other factors that influenced student progress (poverty, family support, etc.) (Darling-Hammond et al., 2011). The researchers also determined alternative ways to evaluate performance including professional standards, artifacts (lesson plans, assignments, etc.), teacher collaboration, and training (p. 5).

When considering value-added scores, there is further evidence that provides insight into why this measurement system may not be the most effective tool when examining teacher performance. Sean P. Corcoran (2010) took a closer look at VAMs of teacher effectiveness as related to the national movement of accountability in education in his executive summary entitled, *Can Teachers be Evaluated by their Students' Test Score? Should They Be?* A great deal of his research emphasized human capital in the teaching profession, education finance, and school choice. He specifically identified evaluation, promotion, compensation, and dismissal of teachers as outcomes for using value-added systems. In his study he examined two prominent systems in New York and Houston. Corcoran (2010) maintained that there were underlying issues with any testing system as it relates to what is measured, not measured, overrepresented, and underrepresented. Further, he stated he did find more consistency in value-added results that utilized a reliable measure over a constant time period; the more years of data accumulated, translated into more accurate results within the context of value-added measures. One conclusion drawn in his study was that great power exists in the idea of using statistical formulas to calculate teacher impact and effectiveness. Corcoran's (2010) work clarified that when considering value-added teacher scores:

> Teachers, policymakers, and school leaders should not be seduced by the elegant simplicity of 'value-added'. Before adopting these measures wholesale, policy-makers should be fully aware of their limitations and consider whether the minimal benefits of their adoption outweigh the cost. (p. 8)

He asserted that his findings indicated that there are numerous challenges and the two programs examined in his study were "crude indicators" of the contribution that teachers make towards student learning (p. 7). As one looks at a tool such as value-added that lacks consistency in

scores, it proves more difficult to use them in a way that provides viable insight about teacher performance. David Berliner (2014) stated,

> I conclude that because of the effects of countless exogenous variables on student classroom achievement, value-added assessments do not now and may never be stable enough from class to class or year to year to use in evaluating teachers. (p. 1)

The use of student growth percentiles has been another method that has been under examination when attempting to quantify teacher performance. According to Monroe and Cai (2015), "An SGP locates a student's current achievement score in a conditional distribution dependent on the student's prior achievement scores. In this way, an SGP provides context for the current achievement" (p. 21). When considering SGPs as a tool to measure teacher efficacy, Lash et al. (2016) contended that:

> This is perhaps the first published study on the stability of the teacher-level growth score derived under the student growth percentile model, a common model used by states in teacher evaluation systems. States or districts may have conducted studies to explore the model; if so, those studies have not appeared in the published literature or been posted on the Internet (see appendix A for related research). The findings indicate that even when computed as an average of annual teacher-level growth scores over three years, estimates of teacher effectiveness do not meet the level of stability that some argue is needed for high-stakes decisions about individuals. (p. 7)

This statement provides an indication that the model does not demonstrate the levels of reliability needed to effectively evaluate teachers; however, that conclusion is further reinforced when considering the Monroe and Cai (2015) examination of student growth percentiles.

> Given that SGPs may be used for high-stakes decisions, such as teacher evaluation, it is important that the statistical properties of the estimates are well understood. The present research focuses primarily on the reliability of SGP estimates. Generally, research has shown that SGP estimates have low levels of reliability at the student level. (p. 21)

Although these sources discuss the relative instability of student growth percentiles scores which discourages their use in teacher evaluations, there is also work that provides a contrary perspective. As stated in Xu, Grant, and Ward (2016),

> Our findings suggest that the teachers in this study with SGPs data were not held to a higher standard than teachers with student achievement goal setting. One concern among teachers is that teachers with state assessment data are being rated more harshly than teachers without state assessment data. The findings suggest the contrary. (p. 216)

Further benefits of student growth percentiles have also been identified with regard to measurement of teacher performance. Hull (2013) described the benefits of student growth percentiles as a

> good measure of individual student growth from one year to the next; cheaper and easier to calculate plus easier to understand than VAMs; More accurate at evaluating teachers than student test scores, which capture performance at one point in time; Tend to be more popular with stakeholders than VAMs since their limitations are not as well known. (p. 16)

When considering the national trends with regard to teacher evaluation in the United States, it is important to note that although accountability is an emphasis, specifically regarding the connection of using student achievement to evaluate teacher performance, there are a variety of approaches that are also considered for measuring teacher effectiveness.

> Forty-one states now require or recommend that teachers be evaluated using multiple measures of teacher performance. These include: Student achievement data, Classroom observations, Other data - student surveys, lesson plan reviews, teacher self-assessments, and more. (Hull, 2013, p. 9)

There is significant research pointing to the importance of considering a number of sources that can be utilized to better characterize and create context for the assessment of how a teacher is performing in the classroom. Finnegan (2013) stated, "Truly effective evaluation models produce feedback to teachers, in addition to, professional dialogue between teacher and administrator and among peers and colleagues" (p. 23). Engberg and Mihaly (2012), with RAND Educational, provided a document entitled *Multiple Choices – Options for Measuring Teaching Effectiveness* that identifies a variety of sources used to measure performance. They stated, "Student test scores are one indicator of teaching effectiveness" (p. 1). They added that classroom observations serve to measure teacher performance directly and that surveys of students can provide valuable insight into levels of engagement and the quality of student-teacher relationships (Engberg & Mihaly, 2012, p. 1). Further examination of classroom observation as a viable method for measuring teacher performance is also identified in the work of Cohen and Goldhaber (2016),

> Classroom observations, on the other hand, are used nearly universally to assess teachers. They have high levels of face validity because they assess teaching practices that teachers

> themselves can observe. For those striving to become better practitioners, this information can provide timely and actionable formative feedback. (p. 378)

Anderson et al. (2016) stated, "Among the most central design tasks for districts in this study was designing observation rubrics to measure classroom practice and planning how to collect and analyze student achievement data to assess teacher impact on student performance". They provided further evidence on what many districts are using as a basis for their current instruments that serve to observe teachers in the classroom,

> Charlotte Danielson's *Framework for Teaching* (FFT) was the exclusive basis for the design of the observation rubrics in half the districts included in the study. The other four districts developed their own rubrics to examine instructional practice, but drew from existing frameworks such as the FFT and Robert Marzano's *Classroom Instruction that Works* as sources of reference. (Anderson et al., 2016, p. 13)

Professional development has also had significant notoriety when considering teacher evaluation instruments. Danielson (2016), explained that "Personnel policies for the teachers not practicing below standard—approximately 94 percent of them—would have, at their core, a focus on professional development, replacing the emphasis on ratings with one on learning" (p. 20). Danielson (2016) expanded the idea of professional development for teachers adding more specificity with regard to the context in which it is most effective by describing that

> Most teachers report that they learn more from their colleagues than from an 'expert' in a workshop. When teachers work together to solve problems of practice, they have the benefit of their colleagues' knowledge and experience to address a particular issue they're facing in their classroom. (p. 20)

When considering teacher evaluation, current emphasis in the educational arena is aimed at the development of a balanced system that takes into consideration the skill level of the individual teacher. Danielson (2016) described this concept:

> They must construct a differentiated system, including designing and supporting a mentoring program; selecting teacher leaders and determining their compensation, support, and supervision; and designing collaborative evaluation procedures for novice and experienced teachers and training for evaluators. (p. 20)

With regard to evaluation development, the federal tone certainly emphasized the use of standardized testing results where consideration is given to how this should take place. Anderson et al. (2016) stated, "Other important steps in system design included involving stakeholder groups, particularly teacher unions, in the process, and assigning weights to each measure of teacher performance" (p. xiii). However, in the Anderson et al. (2016) report designed in partnership with the United States Department of Education, it was stated:

> Districts included in this study generated varied and changing strategies for measuring those contributions, suggesting that the ways in which teachers contribute to student learning are still difficult to measure. Nevertheless, measuring teachers' contributions using standardized assessments has become common practice. Still, some teachers expressed concern about using student assessments to measure their performance when those assessments do not align well with the curriculum they teach. (p. 65)

It becomes clear that while observational practices appear more standard across schools, the process for measuring teacher performance using student achievement data is far less clear.

> There is also little consensus on how the profession should define "good teaching." Many state systems require districts to evaluate teachers on the learning gains of their students.

> These policies have been implemented despite the objections from many in the measurement community regarding the limitations of available tests and the challenge of accurately attributing student learning to individual teachers. (Danielson, 2016, p. 20)

When considering the current state of teacher evaluation, it is important to make note of the recent events surrounding this topic. The PBS News Hour (2013) published *A Brief Overview of Teacher Evaluation Controversies*, identifying a number of events that involved issues related to teacher evaluation development in the United States. Seemingly, RTTT legislation spurred negative debate associated to teacher evaluation.

> The competition, known as RTTT, distributes funding to states that meet specific requirements and set up concrete plans to improve their schools. One key area of reform, as laid out by the law, is teacher evaluations. As such, the contest sparked a whole host of reforms, many of which have led to a number of conflicts between unions and government officials. (PBS News Hours, 2013, p. 1)

As a result, a number of events took place that provided evidence of the disagreement regarding how best to evaluate teachers.

> On Aug. 14, 2010 the Los Angeles Times publishes teacher scores despite resistance from teachers' unions…publishes value-added scores derived from seven years of data looking at 6,000 elementary school teachers in the Los Angeles Unified School District. The following month, Rigoberto Ruelas, who had taught fifth grade for 14 years, commits suicide. His family blames the publication of his 'average' and 'less effective' ratings for raising students' standardized test scores, and United Teachers Los Angeles urges the newspaper to remove the database from their website. (The PBS News Hour, 2013, p. 1)

Other issues around the United States also emerged regarding teacher evaluation. In February of 2012, the New York Post revealed New York City's "Worst Teachers" where it published names and salaries of teachers and included a link to value-added scores of teachers throughout the city (The PBS New Hour, 2013). Given the research and relative instability of value-added scores, the categorization of teachers in this manner received criticism. In Freedberg's 2012 Huffington Post article, he said, **"**Bill Gates came out strongly against the practice in New York. Publicly ranking teachers by name will not help them get better at their jobs or improve student learning" (p. 1). In an effort to continue to tie student achievement to teacher evaluation grew in the United States, further issues emerged:

> On September 10, 2012, the Chicago Teacher Union strikes for 7 Days where the city's teacher's union and its 26,000 members vote to go on strike, preventing more than 350,000 children from going to school…discontent stems from newly imposed and significantly tightened teacher evaluation requirements. (The PBS New Hour, 2013)

On January 17, 2013, New York City was unable to reach an agreement regarding new teacher evaluation policies; the teacher's union and government officials could not come to a consensus about the instrument. The PBS New Hour (2013) reported, "the union expressed concerns over making student test scores account for 40-percent of teacher evaluation grades, an increase of 20-percent" (p. 1). Much of the discontent involved in these events surrounded the prospect of teacher evaluation reform and the desire to connect it to the use of student achievement data as an indicator of teacher performance.

Arizona Teacher Evaluation

An examination of teacher evaluation in the state of Arizona reveals a variety of information. When first considering Arizona, the work of Jim Hull, Senior Policy Analyst for

the Center for Public Education, helped provide a more general picture of how educator evaluation is handled. Hull categorized levels of involvement with regard to how involved states are in the development and management of evaluation systems.

> States vary by how involved they were in the design process, High involvement: 13 states mandated the requirements and components of the evaluation system and required districts to implement them with little flexibility. Medium: 17 states provided model evaluation systems that districts could either adopt or develop their own. For example, a state may mandate the use of student growth models and weights but still allow districts to decide the other features or choose alternative models as long as they meet the state criteria. Low: 21 states required each district to design their own system with state approval. The state may provide guidance but plays a small role in implementation. (Hull, 2013, p. 6)

Hull's (2013) work noted that Arizona was considered to be "low" when considering the range of involvement (p. 8). His work also identified Arizona as a state that "requires multiple measures in teacher evaluation and requires or recommends that student achievement indicators comprise half of a teachers' evaluation" (p. 12), and that Arizona is a state that utilizes a statistical model in order to link student achievement and teachers which, in this case, was stated to be student growth percentiles (p. 19). Lastly, Hull (2013) provided information about Arizona as a state that does not require aggregate teacher evaluation data to be shared publicly (p. 29). These highlights provided a background for some of the more general approaches taken in Arizona with regard to teacher evaluation.

The ADE provided more detailed information with regard to teacher evaluation in the state of Arizona in an ADE document entitled *Teacher Evaluation Process. An Arizona Model for Measuring Educator Effectiveness* (2014).

> Arizona Revised Statute §15-203 (A) (38) was passed by the legislature in 2009. This statute required that the State Board of Education on or before December 15, 2011 adopt and maintain a *framework* for a teacher and principal evaluation instrument that includes quantitative data on student academic progress that accounts for between thirty-three percent and fifty percent of the evaluation outcomes and best practices for professional development and evaluator training. School LEAs and charter schools were directed to use an instrument that meets the data requirements established by the State Board of Education to annually evaluate individual teachers and principals beginning in school year 2012-2013. As a result, the State Board of Education appointed an 18-member Task Force to develop the *Arizona Framework for Measuring Educator Effectiveness*. (p. 1)

In further reviewing the work of this task force, there are number of other developments that align with many of the national trends on teacher evaluation. One specific factor is that the Arizona Framework was developed in conjunction with The Charlotte Danielson Groups (ADE, 2014, p. 2).

> The Charlotte Danielson Framework is the basis for the Teaching Performance Component of the model. A research-based set of components of instruction, aligned to the Arizona Professional Teaching Standards, and grounded in a constructivist view of learning and teaching. The framework consists of 22 components (and 76 smaller elements) clustered into four domains of teaching responsibility: Planning and

> Preparation, Classroom Environment, Instruction and Professional Responsibilities. (ADE, 2014, p. 5)

The ADE (2014) document is comprised primarily of rubrics entitled "The Danielson Framework Rubric" (p. 15). The model is composed of varied weights for other aspects of the instrument, "the teaching performance component makes up 50%, student academic progress is 33% and survey component is 17%" (ADE, 2014, p. 3). The document also consists of professional teaching standards. When considering the student academic progress section of the instrument, it stated,

> The total of school/grade/classroom-level data elements accounts for 33% of the evaluation outcome. If available, AIMS data must be used as at least one of the classroom-level data elements. Student growth data constitutes 20%, or 24 points, of the total evaluation outcome. (ADE, 2014, p. 4)

What becomes evident is the movement toward mandating the use of standardized testing data as a component of the evaluation instrument. The rationale of Teacher Evaluation Process is also included.

> This teacher evaluation model was created to provide process, templates, observation rubrics, and a rating system for measuring teacher performance. All components align and comply with Arizona State Board of Education's adopted *Framework for Measuring Educator Effectiveness*. The *framework* provides the legal parameters and state requirements for the teacher evaluation process statewide. (ADE, 2014, p. 2)

A final aspect of the Arizona evaluation system is the rating levels that are assigned to teachers as result of a final evaluation. The Teacher Evaluation Process document stated,

> When evaluating teaching performance, student level data, and survey results, the four performance classifications described below will be used. The following descriptors were adopted by the Arizona State Board of Education in January, 2013, and cannot be modified. (ADE, 2014, p. 13)

The ratings reflect an update from work done in previous years. They include:

> Highly Effective: The teacher consistently demonstrates the listed functions and other actions reflective of the teaching standards that are above and beyond stated expectations. Teachers who perform at this level exceed goals and targets established for student performance and survey data indicates high levels of satisfaction. A Highly Effective rating means that the only areas for growth would be to expand on the strengths and find innovative ways to apply it to the benefit of the school and LEA. Specific comments (i.e., evidence, explanation) are required for rating a teacher as Highly Effective. A Highly Effective *classification* means that performance is excellent. The employee is a top performer in all areas of teaching performance, student achievement, and academic progress in the perception of others.
>
> Effective: The teacher demonstrates the listed functions reflective of the teaching standards most of the time and meets goals and any targets established for student performance and survey data. Performance in this area is satisfactory and similar to that of others regarded as good performers. The indicator of performance delivered when classifying one as *Effective* is that performance is very good. While there are areas remaining that require further development to be considered an excellent performer in

this standard, an Effective classification is indicative of a valued teacher. Expectations for this level will be determined at the initial teacher conference with the evaluator.

Developing: The teacher sometimes demonstrates the listed functions reflective of the teaching standards and meets some of the goals and targets established for student performance and survey data. A *Developing* classification indicates that the employee performs well at times but requires more consistent performance overall. The teacher demonstrates potential, but must focus on opportunities for improvement to elevate the performance in this standard.

Ineffective: The teacher rarely demonstrates the listed functions and meets few goals and targets for student performance and survey data. The demonstrated performance of this teacher requires intervention. A classification of ineffective indicates that performance is unsatisfactory and the teacher requires significant improvement. Specific comments (i.e., evidence, explanation) are required when applying this classification (p. 13).

Summary

Chapter 2 covers the wide range of literature that exists in relation to how accountability has evolved over the year in education, thus impacting the approach taken with regard to teacher evaluation. A review of the literature in connection with the historical development of teacher evaluation through the 21st century is provided in this chapter. Additionally, the research and literature associated with national trends in teacher evaluation as well as current practices in Arizona are included in connection with the performance evaluations of educators.

CHAPTER 3

Research Design and Methodology

Introduction

This chapter includes the design for this research study as well as the methodological approach. A mixed method approach was applied in order to gain insight into the attitudes of high school mathematics teachers and administrators regarding the use of student standardized test results in teacher performance evaluations. The data collection for both the quantitative and qualitative analysis of survey results and one-on-one interviews will be described in this chapter. The study was broken down into eight sections which serve to provide a detailed description of the study: (a) Restatement of the Problem, (b) Restatement of Research Questions, (c) Research Design, (d) Population and Sample, (e) Instrumentation, (f) Validity and Reliability, (g) Data Collection Procedures, and (h) Data Analysis Procedures.

Restatement of the Problem

The problem examined in the study involved the attitudes of teachers and principals. More specifically, what are teacher and principal attitudes towards standardized testing results when they served as an indicator that measured an individual teacher's performance? As educational organizations attempt to tie teacher effectiveness to student achievement, a great deal of research indicates that the measures currently utilized to accomplish this task are not considered stable enough to do so. VAMs and SGPs are two instruments that have been widely utilized for this purpose. Each has been studied and identified as insufficient for the purpose of measuring teacher performance in connection with student achievement data. For example, Corcoran's (2010) work provided examples with regard to VAMs in that "given the extent of uncertainty in teacher value-added scores, it would not be surprising if these estimates fluctuated

a great deal from year to year" (p. 6). Corcoran specifically identified evaluation, promotion, compensation, and dismissal of teachers as outcomes for using value-added systems. Often times, this research provides an indication that these accountability systems are unstable or require further examination. Additionally, Lash et al. (2016) noted that "the findings indicate that even when computed as an average of annual teacher-level growth scores over three years, estimates of teacher effectiveness do not meet the level of stability that some argue is needed for high-stakes decisions about individuals" (p. 7). While these measures are under debate about their relative effectiveness, the problem arises in that they have been used as a piece of teacher performance evaluations. It is important to note that a need exists with regard to further research in this field when considering the use of student standardized test results as a measure of teacher performance in teacher evaluations.

Restatement of the Purpose of the Study

The purpose of this study was to examine attitudes of both high school math teachers and administrators regarding whether or not student standardized test results were seen as an effective measure of instructional performance.

Restatement of Research Questions and Hypotheses

The research questions and hypotheses that guided this study included:

1. What are the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?
    a. What are the attitudes of teachers instructing accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

b. What are the attitudes of teachers instructing non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

c. What are the attitudes of teachers that instruct both accelerated and non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

2. Is there a statistically significant difference between the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

   $H_0$2. There is no statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

   H2. There is a statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

3. What are the attitudes of high school administrators regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?

4. Is there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?

   $H_0$4. There is no statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of math teacher instructional performance in the classroom.

   H4. There is a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of math teacher instructional performance in the classroom.

Research Design

The research design for this study was characterized as a mixed-methods methodology. According to Creswell (2015), mixed method research is "an approach to research in the social, behavioral, and health sciences in which the investigator gathers both quantitative (close-ended) and qualitative (open-ended) data, integrates the two, and then draws interpretations based on the combined strengths of both sets of data to understand research problems" (p. 16). In order to develop a comprehensive picture of both administrator and teacher attitudes towards the use of student standardized test results on evaluation instruments as a measure of performance, it was essential to gather both quantitative and qualitative data. The design subtype is an additional factor that must be addressed in this study. The design subtype was a sequential mixed methods design in that survey data was collected first. One-on-one interview participants were identified through the implementation of a quantitative survey. Therefore, interview data was collected in

a sequence of quantitative first and qualitative second. Plano Clark & Ivankova (2015) described sequential mixed method design as

> Sequential timing refers to situations when researchers collect and analyze quantitative and qualitative data in sequence—one following or dependent on the other. For example, researchers can collect and analyze quantitative data first and then use these results to inform the follow-up qualitative data collection. (pp. 64-65)

Check and Schutt (2012) elaborated further in that "a common reason for mixing both quantitative and qualitative methods in one research project is to take advantage of the unique strengths of each methodological approach when engaged in different stages of the research process" (p. 239). In order to gather data that effectively described attitudes of both teachers and administrators toward the use of standardized test results in teacher evaluations, the implementation of a survey and interviews were necessary steps in the study. Tashakkori and Creswell (2007) noted that

> we have broadly defined mixed methods here as research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or a program of inquiry. (p. 4)

The data garnered from these sources served to investigate the attitudes of these two important groups of educators. Survey data were used quantitatively to guide qualitative coding of one-on-one interview data.

Census and Sample

According to Check and Schutt (2012), population is described as "the entire set of individuals or other entities to which study findings are to be generalized" (p. 92). This study

took place in a large, suburban school district in Arizona named District A that is a unified, K-12 school district. District A has 30 elementary schools, seven middle school serving 7th and 8th grade students, and five high schools for grades 9-12. In this study, the population included two important groups within District A's education system. The first was that of high school math teachers. This population was further defined to include all teachers who instruct an accelerated math course that served students in grades 9-12, a non-accelerated math course or both. Accelerated courses were defined as including a prefix such as honors or AP (advanced placement) and that move at a more rapid pace with regard to the delivery of instruction. Therefore, any teacher responsible for instructing courses that were exclusively identified as honors or AP were assigned to the accelerated group. The non-accelerated teachers were responsible for instructing courses that were standard, survey courses within District A in mathematics and did not include any courses that moved at a more rapid pace. The both group consisted of teachers that included both AP/Honors courses and non-accelerated courses as part of the teaching assignment.

Mathematics courses utilized to identify teachers were listed in District A's high school course catalog that were consistent across all five high schools. Courses that were specific to individual high schools, but were not listed for all schools, were not included for the purpose of teacher selection. Special education courses, online courses, and courses offered by the alternative school in District A were not included in the study. Research question number one included three different groups of teachers that instructed accelerated, non-accelerated or both types of courses and the survey used in the study included all of the mathematics courses identified under the criteria of non-accelerated, accelerated, and courses offered across the five high schools in the study district.

The second group can be defined as administrators who served students and teachers at the high school level and addressed research question three. This group included administrators who evaluate and/or train the mathematics teachers who provide instruction to students in grades 9-12. Under this criteria, the group of administrators included individuals who served in administrative positions at both school sites and the district level. Positions included at the school sites were principals and assistant principals. Therefore, the group of administrators identified in question three included district administrators, high school principals, and high school assistant principals.

The manner in which the population was identified in this study can be described as a census. The study investigated the attitudes of high school math teachers and administrators in a specific district and due to the fact that both groups are familiar with the standardized tests that their students are required to take, they served as an ideal population that helped to best describe attitudes toward the use of standardized tests as a measure of performance. A census was identified as the tool to use for the study as the population of math teachers and administrators at the high school level has a limited size. According to Daniel (2012) "the smaller the population, the more favorable it is to choose to take a census" (p. 56). Due to the small size of this particular population, the use of census becomes a legitimate tool when surveying both high school mathematics teachers and administrators. High School mathematics teachers and administrators served as specific groups with a limited size within District A and as distinct groups that possessed key information with regard to teacher evaluation and the use of standardized test results.

Sample size of survey participants consisted of a total of sixty-five high school mathematics teachers and twenty-two administrators. These figures served as the total population

within District A, which also supported the rationale for canvassing this population. The three groups of teachers included six accelerated teachers, twenty-four non-accelerated and twenty-two in the both category. When considering the administrator group, there were five district administrators, five high school principals, and ten high school assistant principals. Lastly, interview participants totaled ten teachers and five administrators. Interview participants were identified using a question on the survey that inquired about interest in participating in the study further through an interview. Participants were able to indicate their willingness to participate in an interview and provided contact information for the researcher. It is important to note that sample size for interview participants remained consistent even though similar patterns were identified with regard to qualitative analysis. According to Marshall (1996),

> An appropriate sample size for a qualitative study is one that adequately answers the research question. For simple questions or very detailed studies, this might be in single figures; for complex questions large samples and a variety of sampling techniques might be necessary. In practice, the number of required subjects usually becomes obvious as the study progresses, as new categories, themes or explanations stop emerging from the data (data saturation). Clearly this requires a flexible research design and an iterative, cyclical approach to sampling, data collection, analysis, and interpretation. This contrasts with the stepwise design of quantitative studies and makes accurate prediction of sample size difficult when submitting protocols to funding bodies. (p. 523)

Instrumentation

Survey. A variety of sources contributed to the data gathered for this study. In order to develop the survey for both teachers and administrators, the Teacher Evaluation Profile (TEP) questionnaire was adapted for use in this study. The TEP questionnaire was developed by

Stiggins and Duke (1988). Slim (2004) further revised the questionnaire in order to complete a study focusing on teacher evaluation entitled *The Influence of Governance Structure on Teacher Evaluation Practice*. While the questionnaire in this study maintained the same format with regard to question stems, organization, and use of a five-point Likert scale, questions on the survey were adapted to specifically focus on the use of student standardized test results as an evaluation tool rather than a general focus on attitudes towards an overall evaluation instrument and process. As a result, a survey was developed that included six questions (A-F) focused on gathering basic demographic information from participants regarding experience, gender, and position taken directly from the TEP. An additional 28 questions, based off of the TEP questionnaire by Stiggins and Duke (1988) and Slim (2004) was created in order to gather data about teacher and administrator attitudes in direct connection with student standardized test results. Survey questions were then categorized by the primary investigator into three themes. Theme 1 questions considered the concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness. Theme 2 questions focused on attitudes towards standardized test results and the degree of trust that participants had/did not have with regard to standardized test results. Theme 3 questions considered the actual process of teacher evaluations within the organization. Table one provides a breakdown of survey question alignment to each of the three themes. Further, the survey can be seen in Appendix A.

Interviews. An additional source of data were gathered during the study by conducting one-on-one interviews. Nine open-ended interview questions were adapted from Slim's (2004) study on Governance and Teacher Evaluation with a focus on the use of standardized testing results as a component of teacher evaluation (see Appendix B). An extensive search for the researcher Slim, who was responsible for developing the interview questions, yielded no results,

thus he was unable to be reached for permission to adapt the interview questions for this study. The researcher then piloted the interview questions. Interviewing teachers and administrators served as a means for gathering qualitative data. According to Check and Schutt (2012), "intensive interviewing engaged researchers more actively with subjects than standard survey research does" (p. 202). Additionally, "intensive or depth interviewing is a qualitative method of finding out about people's experiences, thoughts, and feelings" (Check & Schutt, 2012, p. 201). The implementation of interviews for this study served as a tool to dig deeper into the thoughts of both teachers and administrators, and served as a viable source of qualitative data that better described the attitudes of participants. Interview questions were categorized by three themes similar to the survey questions. Theme 1 one questions considered the concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness. Theme 2 questions focused on attitudes towards standardized test results and the degree of trust that participants had/did not have with regard to standardized test results. Theme 3 questions considered the actual process of teacher evaluations within the organization. Table one provides a breakdown of interview question alignment to each of the three themes.

Table 1

Theme Alignment of Survey and Interview Questions

Theme 1 Questions		Theme 2 Questions		Theme 3 Questions	
Survey Questions	Interview Questions	Survey Questions	Interview Questions	Survey Questions	Interview Questions
SQ1	IQ1	SQ17	IQ3	SQ6	IQ4
SQ2	IQ2	SQ18	IQ7	SQ9	IQ8
SQ3	IQ5	SQ21		SQ10	IQ9
SQ4	IQ6	SQ22		SQ11	
SQ5		SQ23		SQ12	
SQ7		SQ24		SQ13	
SQ8		SQ25		SQ16	
SQ14		SQ26		SQ19	
SQ15		SQ27		SQ20	
SQ28					

Validity and Reliability

Check and Schutt (2012) noted that "measurement validity refers to the extent to which measures indicate what they intended to measure" (p. 81). Additionally, reliability is described as "a measurement procedure that yields consistent scores (or that the scores change only to reflect actual changes in the phenomenon" (Check & Schutt, 2012, p. 83). When considering the reliability and validity of the instrument utilized to survey teachers and administrators in this study, the survey instrument was adapted from Stiggins and Duke's (1988) Teacher Evaluation

Questionnaire (TEP). The TEP focused on attributes that the teacher brought to the evaluation experience, perceptions of their evaluator, attributes of the evaluation procedures, and attributes of feedback (Stiggins & Duke, 1988, p. 94). Stiggins and Duke (1988) noted that "Four studies are described in the order in which they were conducted to share the evolution of our thinking on the revision of the evaluation process" (p. xi). As these studies evolved, they explained that "The third focused on the evaluation experiences of a few teachers who benefited from successful, growth-producing evaluations" (Stiggins & Duke, 1988, p. xi). In their study, the researchers implemented the TEP Questionnaire. The researchers described the methods employed and results obtained with regard to the TEP's reliability. For example, "It is worthy to note that the internal consistency reliability of the total fifty-five item instrument was .93, suggesting that the questionnaire asks a highly cohesive set of questions about the evaluation process" (Stiggins & Duke, 1988, p. 99). The information suggested that the TEP meets more than acceptable criteria for being considered a reliable measure. When considering validity of the TEP,

> Our research results and the TEP are based on a limited number of cases. The external validity of those case study results has not yet been established. The TEP may profit from further technical analysis. The confidence we have in the evaluation process and its impact on teacher improvement will grow as we accumulate corroborating evidence. (Stiggins & Duke, 1988, p. 121)

Additionally, the TEP has been adapted and implemented in a variety of other studies. Slim (2004), in the study entitled *The Influence of Governance Structure on Teacher Evaluation Practice,* noted that "The TEP questionnaire is found to be an instrument of high reliability and

validity and guides the selection of the key domains associated with the teacher evaluation process articulated and confirmed during the pilot study" (p. 81).

A series of nine interview questions were developed for the purpose of one-on-one interviews in order to gather qualitative data. The nine questions were adapted from the interview questions developed in Slim's (2004) study; however, the adapted questions for this study served to emphasize a focus on student standardized testing results and the connection of those results to the teacher evaluation process. The primary investigator sought expert validation of the qualitative interview questions by consulting committee members Dr. Richard Wiggall and Dr. Mary Dereshiwsky. The consultation with experts was intended to increase the reliability and validity of the interview questions prior to use in the study. The intent of interviews was to develop a more intensive description of participant attitudes towards the use of standardized testing results used as an indicator of performance in an evaluation system. Slim (2004), maintained specific steps in order to increase validity and reliability during the interview process, noting that

> The pre-constructed interview instrument corresponds to a standardized open-ended interview design. The exact wording and sequence of questions were predetermined and served as the focal point in each interview session. Prior to the commencement of each interview, respondents were presented with a copy of the questions sequenced according to the chronology in which they were presented. All respondents were asked the same questions in the predetermined order. (p. 82)

The aforementioned steps were replicated in this study for the purpose of enhancing the reliability of the interview process. Additionally, interview questions were pre-screened using a randomly selected group of administrator from District A. The administrators in this district

were asked to review the questions for clarity of wording and an understandable format. The administrators were selected randomly and were not participants in the study. Suggested revisions were applied to the interview questions prior to initiation of the overall study.

Data Collection Procedures

Data collection procedures began with first applying for and receiving IRB approval from the Northern Arizona University (NAU) (Appendix C). Once IRB approval was secured and the Informed Consent was approved (Appendix D), the process for research approval was then sought out for District A. District A, when considering studies that involve employees, required that the Assistant Superintendent be contacted via written letter so that the specific details of the study can be presented to the superintendent for approval (Appendix E). The next step was for the researcher to contact the high school principals (Appendix F) in order to secure teacher names that taught the specific mathematics courses at each high school outlined in research question number one of this study. The teachers were then sent a cover letter (Appendix G) explaining the purpose and nature of this study accompanied by an electronic survey of the adapted TEP questionnaire survey. Informed consent was also described in the cover letter prior to participants completing the survey. Administrators were also sent the TEP survey with a cover letter that included the informed consent. The survey instrument was developed using Google Surveys. The Google platform is commonly used in District A so participants were familiar with format and access so as to more conveniently access the survey for completion. One week reminders were sent to participants each of the first three weeks of the study as a follow up reminding them about completion of the survey. All survey data were exported into a password protected excel spreadsheet for security purposes and stored on a password protected laptop owned by the researcher.

One-on-one interviews were also scheduled with ten teachers and five administrators for the purpose of securing greater insight into attitudes as related to the use of student standardized test results in evaluation instruments in order to identify themes. Interviewees were provided with the questions in advance of the interview in order to allow time for reflection and participants were questioned about any clarifications that were needed in order to comfortably respond to interview questions. All interviews were conducted in the same format; for example, question order remained the same for each interview with no time limit. After saying the informed consent, interviewees were assured of confidentiality and permission to record and ultimately transcribe each interview was requested of each participant. Interviewees were asked to provide candid, open and honest responses and reassured that no personal information would be taken as a result of the interview. The explanation served as an additional step to ensure the confidentiality of participants.

In anticipation of the possibility that interviews ran longer than the one hour/sixty minutes of allotted time, a contingency plan was developed to address this issue. For example, interviewees were provided a follow-up interview via phone at a convenient time so as to complete the full interview process. Therefore, the researcher was still able to collect full qualitative interview data yet honor the time of interview participants. The researcher also sought feedback on the interview questions for clarity from participants. All interview data were stored in a locked drawer for security purposes. Per university guidelines/requirements, data will be stored for five years and then disposed of at that time.

Data Analysis Procedures

TEP survey data were analyzed utilizing SPSS software for the purpose of examination and summarization. Descriptive statistics were applied for each research question in part one of

the survey in order to provide insight into the demographics of participant responses on survey items. Additional quantitative analysis was necessary when considering survey data. According to Martin and Bridgmon (2012)

> Nonparametrics use either exact probability or excellent approximations for large samples. Thus, the accuracy of probability statements does not rely on the shape of the population…nonparametric statistics can be used for dependent variable scores that are inherently in the form of ranks (ordinal) or categories (nominal). (p. 348)

Therefore, the use of a five point Likert scale on part two of the survey instrument warranted the use of nonparametric statistics. Two specific statistics were applied for the purpose of analysis. When considering research question two, there were three groups of teachers. For example, the teachers of mathematics courses listed were categorized into three groups: accelerated, non-accelerated, or both. Teachers of basic survey courses such as Geometry 1-2, Algebra 1-2, Algebra 3-4 served as examples of nonn-accelerated and Honors/AP courses were categorized as accelerated. A group of teachers also assumed responsibility for instructing both accelerated and non-accelerated courses, creating a third group. In this case, the Kruskal-Wallis One-way Analysis of Variance was applied for the purpose of quantitative analysis for research question two. Martin and Bridgmon (2012) noted that "the purpose of the Kruskal-Wallis (K-W) test is to compare mean rank differences among two or more groups. The K-W test is a nonparametric alternative to the one-way ANOVA" (p. 70). Each survey question on part two of the instrument was analyzed using the Kruskal-Wallis Test in order to determine significance. An alpha level of .05 was used to determine significance. Research question four included an additional group of administrators. Overall administrator and the teacher groups were analyzed applying the Mann Whitney U Test as two groups being compared in the study and the MWU was applied to each of

the part two survey questions. Martin and Bridgmon (2012) noted the Mann Whitney U Test is a "…focus of analysis. The purpose of the Mann Whitney U (MWU) test is to compare mean rank differences between two groups. The MWU test is the nonparametric alternative to the independent t-test" (p. 71). Additionally, when considering the qualitative data obtained from one-on-one interviews of both administrators and teachers, the focus was on the use of descriptive coding in order to identify specific themes that existed about attitudes toward the use of standardized tests results as an evaluative tool for research questions one, two and three. According to Richards and Morse (2013), "Descriptive coding is used to store things known about data items…the researcher can then access this factual knowledge about the respondent, the setting, or context, when seeking patterns, explanations, and theories" (p. 154).

The data taken from both surveys and one-on-one interviews were compared in order to identify whether the quantitative survey data yielded similar patterns identified from the descriptive coding as a result of qualitative interviews. The analysis served to identify whether similar attitudinal perspectives were evident in survey data as well as interviews from both teachers and administrators. The connection of the two data sources served as a means for data triangulation in order to identify whether or not the sources converged with regard to the use of student standardized test results on teacher evaluation as an effective indicator of performance.

Table 2 contains the match-up of the research questions to corresponding sources of information and data analysis/reporting procedures. Table 3 illustrates the multimethod convergence information.

Table 2

Match-up of Research Questions to Corresponding Sources of Information and Data Analysis/Reporting Procedures

Research Question	Corresponding Source(s) of Information	Corresponding Data Analysis/Reporting Procedure(s)
1. What are the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom? a. …of teachers instructing accelerated math courses… b. …of teachers instructing non-accelerated… c. of teachers instructing both accelerated and non-accelerated…	Adapted TEP Survey and interview questions of mathematics teachers.	1. Summary descriptive statistics per survey questions (% of response distribution across Likert scale and means) 2. Qualitative procedures (descriptive coding in order to identify themes across interview participants)
2. Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated math courses, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?	Adapted TEP Survey of teachers part two.	1. Kruskal-Wallis Test

Table 2 (continued)

Research Question	Corresponding Source(s) of Information	Corresponding Data Analysis/Reporting Procedure(s)
3. What are the attitudes of high school administrators regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?	Adapted TEP Survey and interview questions for administrators.	1. Summary descriptive statistics per survey questions (% of response distribution across Likert scale and means) 2. Qualitative procedures (descriptive coding in order to identify themes across interview participants)
4. Is there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?	Adapted TEP Survey of administrators part two and teachers.	1. Mann-Whitney – U Test

Table 3

Multimethod Convergence Information

Question	Quantitative Results	Qualitative Results	Comparison/ Convergence
1. What were the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom? a. …of teachers instructing accelerated math courses… b. …of teachers instructing non-accelerated… c. of teachers instructing both accelerated and non-accelerated…	Descriptives in SPSS were applied in order to analyze survey questions and identify trends in teacher survey responses.	Interview responses on each question were descriptively coded in order to identify specific themes that may exist about teacher attitudes towards the use of standardized tests results as an evaluative tool.	Descriptive data and theme identification will be combined in order to form a basis for results and conclusions with regard to teacher attitudes.

Table 3 (continued)

Question	Quantitative Results	Qualitative Results	Comparison/ Convergence
2. Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated math courses, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?	Kruskal-Wallis Test was applied as a nonparametric statistic in order to compare to survey responses of mathematics teachers (non-accelerated and accelerated course instructors).	Quantitative Only	Quantitative Only
3. What were the attitudes of high school administrators regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?	Descriptives in SPSS were applied in order to analyze survey questions and identify trends in administrator survey responses.	Interview responses on each question were descriptively coded in order to identify specific themes that may exist about administrator attitudes towards the use of standardized tests results as an evaluative tool.	Descriptive data and theme identification will be combined in order to form a basis for results and conclusions with regard to administrator attitudes.

Table 3 (continued)

Question	Quantitative Results	Qualitative Results	Comparison/ Convergence
4. Was there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?	Mann Whitney U Test was applied as a nonparametric statistic in order to compare survey responses of mathematics teachers and administrators.	Quantitative only	Quantitative Only

Summary

The purpose of this study was to analyze both quantitative survey results and qualitative interview results in order to better understand teacher and administrator attitudes toward the use of student standardized tests results as an indicator of performance on teacher evaluations. The research design was mixed methods where sequential timing was applied as a subset methodology design in order to gather survey data and interview data simultaneously. The instrumentation used to gather this data was described and details concerning the sources for how the instrumentation was identified and applied were shared in this chapter. Procedures for data collection and analysis were established for the purpose of presenting findings, analysis, and summary of the quantitative and qualitative in Chapter Four. Lastly, Chapter Five presents the summary of the study, conclusions, implications for practice, and recommendations for further study.

Chapter 4

Findings

Introduction

The purpose of chapter four is to present the findings included in this study as related to the attitudes of high school math teachers and administrators regarding how standardized test results served as an effective measure of their instructional performance. Chapters one through three provided an introduction to the topic, review of the pertinent literature on teacher evaluation, and the design methodology for the study. The present chapter makes use of both quantitative and qualitative data for the purpose of presenting findings. Descriptive statistics were applied in order to describe three groups of high school mathematics teachers who instruct accelerated (Group A), non-accelerated (Group B), or those who teach both types of courses (Group C). Similar descriptive statistics were applied in order to describe the group of administrators. The administrative groups are comprised of district level administrators (DA), high school principals (HP), and high school assistant principals (AP). The present chapter uses non-parametric statistics because the data did not meet assumptions of parametric statistics of normality, equal interval measurement and homogeneity of variance. Therefore, results of the Kruskal-Wallis Test were applied as a nonparametric statistic in order to analyze survey responses of mathematics teachers within the three groups of educators. The Mann-Whitney U Test was applied as a nonparametric statistic in order to compare survey responses of mathematics teachers and administrators. Lastly, interview questions of both mathematics teachers and administrators were transcribed and thematically analyzed in order to provide greater depth towards the description of both teacher and administrator attitudes.

A description of the two groups that were surveyed helps in understanding the composition of both the teacher and administrator population in District A. More specific descriptive analysis of the participants will be provided in subsequent research questions. A total of 65 mathematics teachers are employed in District A who fall into one of three categories of mathematics teachers. There were 52 teachers who responded to the survey instrument provided by the researcher. The end result was a response rate of 80%. The groups, based upon survey responses about specific courses taught, were assigned to either an accelerated, non-accelerated or both category of high school mathematics teachers in District A. The accelerated group (Group A) was made up of six teachers, the non-accelerated group (Group B) included 22 teachers, and the both category (Group C) included 24 teachers. The administrative group was composed of a total of 22 administrators who serve at either the district level, are high school principals, or high school assistant principals. Surveys were sent out to the entire population of district administrators, high school principals, and high school assistant principals. This population included 22 individuals and 20 responses were obtained, yielding a response rate of 90.91%. Once survey responses were collected, groups of administrators were assigned to the three groups based upon the current position. There were five district administrators, five high school principals, and 10 assistant principals included in the survey results.

When considering the survey instrument used to analyze teacher and administrator responses, part one of the survey yielded data that provided a description of each of group with regard to gender, education level, and years of experience. Part two of the survey included 28 Likert- scaled survey questions that had values ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). All questions for part two of the survey were categorized into three different themes. Theme 1 (Concept of standardized test results as an indicator used to measure teacher

performance and/or effectiveness), Theme 2 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness.), and Theme 3 (Actual process of teacher evaluations).

Research Question 1

What are the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

Research Question 1a Findings

What are the attitudes of teachers instructing accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

Quantitative RQ1a findings: Group A. This section will provide summary descriptive statistical information on the group of teachers identified as instructors of high school accelerated mathematics courses in District A. Additionally, data will be presented summarizing mean scores for this group on each of the 28 Likert-type scale survey questions according to Themes 1, 2, and 3.

Table 4, provides the accelerated group composition with regard to gender breakdown, highest degree earned, years in the field of education, and years within the District A. The group of high school accelerated mathematics teachers was comprised of six teachers as taken from part one of the survey instrument. When looking at gender, there were four females, which equaled twice as many participants (66.7%) than the two male (33.3%) participants.

The majority of teachers in this group responded as having earned a master's degree (5, 83.3%), with only one (16.7%) holding a bachelor's, and none having a doctorate. Respondents also possessed varying levels of experience with regard to years in the profession: there was one

teacher in each of the 11-15, 16-20, 26-30, and >30 years categories, and two teachers had taught from 6-10 years. Of the six teachers, two each had taught in the District A 6-10 years and 11-15 years while one each taught in the district 26-30 and >30 years. None of the accelerated teachers reported having an experience level of less than six years in both total years of teaching experience and total years of teaching experience within the schooling organization.

Table 4

Group A: Gender, Degree, Total Years Teaching, and Years Teaching in District A

Gender

Male		Female		Total	
#	%	#	%	#	%
2	33.3	4	66.7	6	100

Degree

BA		MA		PhD		Total	
#	%	#	%	#	%	#	%
1	16.7	5	83.3	0	0	6	100

Years Total Teaching Experience

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	0	0	2	33.3	1	16.7	1	16.7	0	0	1	16.7	1	16.7	6	100

Table 4 (continued)

Years Total Teaching Within District A																	
0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	0	0	2	33.3	2	33.3	0	0	0	0	1	16.7	1	16.7	6	100

Quantitative RQ1a, theme 1 findings: Group A. The survey questions that addressed Theme 1 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) included: SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, and SQ28. As shown in Table five, each SQ is listed with the number of respondents for each category on the Likert scale. Additionally, Table five summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ1. SQ1 (The use of standardized test results is an effective tool for measuring teacher performance) responses for the accelerated group reflected disagreement for all respondents. Three respondents Strongly Disagree and three Disagree. The responses yielded a mean of 1.5. All respondent's responses reflected an attitude of disagreement towards the specific statement made by SQ1.

SQ2. SQ2 (I feel confident that the use of standardized test results can improve teacher performance in the district) produced similar results to SQ1 in that all respondent's responses reflected an attitude of disagreement; however, five of the six respondent's indicated an attitude of Disagree on the Likert-scale, whereas only one respondent selected Strongly Disagree. While responses on both SQ1 and SQ2 reflected disagreement, there were more respondents who Strongly Disagreed with SQ1 and more that Disagree with SQ2.

SQ3. SQ3 (Student standardized test scores should be a component of the teacher evaluation process) presented similar results to that of SQ1 and SQ2 in that all responses of the accelerated group of teachers reflected an attitude of disagreement. SQ3 responses included four Disagree and two Strongly Disagree. Both SQ2 and SQ3 yielded a mode of 2.00, meaning that for both survey questions the most common response was Disagree. SQ1 yielded a mode of 1.00 which reflected a stronger attitude of disagreement to SQ1 over SQ2 and SQ3.

SQ4. When considering SQ4 (Student standardized test scores are accurate in assessment of teacher performance), analysis of Likert-scaled questions reflected the highest levels of disagreement in that five of the six responses picked Strongly Disagree and one Disagree. SQ4 yielded the lowest mean of all Theme 1 questions at 1.17.

SQ5. SQ5 (Student standardized test scores reflect a teacher's knowledge of teaching practices) presented slightly different results in comparison to SQ1 through SQ4 in that one response fell in the Neutral range. There were two responses in the Disagree category and three in the Strongly Disagree. While the mode for SQ5 reflects similar attitudes to that of SQ1 and SQ4 because the most frequent response was Strongly Disagree, attitudes of respondent's in SQ5 presented some level of neutrality in that one accelerated teacher responded as such. However, 83% (five out of six) of respondents for SQ5 still reflected an attitude of disagreement.

SQ7. SQ7 (Student standardized test scores influence a teacher's future teaching performance) produced similar results to that of SQ2 in that disagreement was evident in respondent's attitudes. Five out of six responses to SQ7 chose Disagree, while one picked Strongly Disagree.

SQ8. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance) yielded results that were similar to that of SQ1 in the responses

were evenly divided between Strongly Disagree and Disagree with three respondent's for each. SQ8 yielded a similar pattern of disagreement to SQ1, SQ2, SQ3, SQ4, and SQ7 where all responses fell into the two categories of Strongly Disagree and Disagree.

SQ14. SQ14 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance) produced the most unique results in comparison to other survey questions in that it was one of two Theme 1 SQs that included a response that reflected agreement. For example, SQ14 was one of two questions with a mean above two (2.17) and had one Group A response of Agree. Additionally, another respondent selected Neutral. The last four responses were equally divided between Strongly Disagree (2) and Disagree (2).

SQ15. SQ15 (Students' performance on standardized tests is an effective tool that can be used to measure classroom performance) also produced results that were different from previous SQs, but similar to SQ14. SQ15 was the only other Theme 1 question to produce a mean above two (2.50) and reflected higher levels of neutrality. For example, two accelerated teacher responses were Neutral and one was Agree. There was one response at the Disagree level and two at the Strongly Disagree level. However, 50% of responses from accelerated teachers demonstrated attitudes of either Neutral or Agree on SQ15.

SQ28. The final question, SQ28 (I am confident that student's standardized test results accurately measure teaching effectiveness), produced results similar to Theme 1 SQs apart from SQ14 and SQ15. SQ28 responses reflected attitudes of Strongly Disagree in that four Group A teachers picked Strongly Disagree and two Disagree. SQ28 yielded the second lowest mean across Theme 1 questions at 1.33, only higher than SQ4 mean of 1.17.

Quantitative summary RQ1a, theme 1: Group A. There were eight Theme 1 questions (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ28), within Group A responses, where mean scores fell into the rating of one (or Strongly Disagree) on the Likert scale. Two questions (SQ14 and SQ15) had a mean score within the two threshold (Disagree) on the scale with 2.17 and 2.50 respectively. Therefore, results suggest that mean responses for this group represented varying levels of disagreement. SQ14 and SQ15 both emphasized the concept that student standardized test results and student performance on standardized tests served as tools to provide an accurate measure of classroom performance. However, when examining Theme 1 questions outside of SQ14 and SQ15, which emphasized the concept of using student standardized test results on teacher evaluations in order to measure teacher effectiveness, knowledge of teaching practices, and performance, means were slightly lower and fell in the range of 1.17 to 1.83 (Strongly Disagree). The findings presented in Theme 1 questions for the accelerated teacher group reflected consistent levels of disagreement. Overall, across all SQs in Theme 1, there were two respondents with an attitude of agreement (SQ14 and SQ15) and four that were Neutral (SQ5, SQ14 and SQ15). The remaining responses for the accelerated group of teachers all fell within the Likert scale categories of Strongly Disagree and Disagree.

Table 5

RQ1a, Theme 1: Group A Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)

	SD	D	N	A	SA	Mean	Mode
1. The use of standardized test results is an effective tool for measuring teacher performance.	3	3	0	0	0	1.5	1.00
2. I feel confident that the use of standardized test results can improve teacher performance in the district.	1	5	0	0	0	1.83	2.00
3. Student standardized test scores should be a component of the teacher evaluation process.	2	4	0	0	0	1.67	2.00
4. Student standardized test scores are accurate in assessment of teacher performance.	5	1	0	0	0	1.17	1.00
5. Student standardized test scores reflect a teacher's knowledge of teaching practices.	3	2	1	0	0	1.67	1.00
7. Student standardized test scores influence a teacher's future teaching performance.	1	5	0	0	0	1.83	2.00
8. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	3	3	0	0	0	1.50	1.00
14. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	2	2	1	1	0	2.17	2.00
15. Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.	1	2	2	1	0	2.50	3.00
28. I am confident that student's standardized test results accurately measure teaching effectiveness.	4	2	0	0	0	1.33	1.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1a, theme 1 findings: Group A. There were four interview questions that aligned to Theme 1 (IQ1, IQ2, IQ5, IQ6). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. There were 10 total teachers interviewed for qualitative purposes in the study. Of the 10 teachers, three in the group were specifically categorized as accelerated teachers. The

interview responses for the three teachers were used to identify themes for Group A and are labeled as T2, T3 and T7.

IQ1. As found in Table 6, two common themes surfaced for IQ1 (What do you believe is the intended purpose of using student standardized test results as a component of the evaluation tool within your schooling organization?) related to Theme 1. The first theme that emerged emphasized the idea that the intended purpose of the using standardized tests results as a component of the teacher evaluation system was to verify that teachers were doing what they were supposed to be doing. More specifically, the purpose served as an accountability measure to ensure that teachers were teaching effectively as related to the state standards. For example, T2 stated, "I think the intent is related to to just making sure teachers are doing what they're supposed to be doing; making sure the students are learning and that they're growing." T2's quote reflects both themes in that the interviewee specifically mentions teacher accountability and references that the focus should be on student growth and learning. T7 noted "I think the purpose is to see if the teacher is effective in teaching the standards as outlined in the common core standards." T7 also reflects an attitude that standardized test results are used in the evaluation system in order to be sure that teachers are accountable. The second theme that emerged focused on the idea that the purpose of the evaluation instrument does not focus on the growth of students. T3 clearly highlights the second theme and stated "I don't think it is a good idea. I think that the instrument does not look at how much growth the students have from the previous year."

The first theme for IQ1 appears to run contrary to quantitative survey disagreement about the use of standardized test results in evaluations. Interviewee responses indicated that teacher accountability was an intended purpose of using standardized test results in evaluations and

disagreement towards this concept did not surface; however, when considering the second theme for IQ1, which focused on the instrument's inability to effectively measure student growth, a connection to the quantitative survey results emerged. Quantitative survey results expressed consistent disagreement towards using standardized tests results on evaluations and IQ1 interview results also expressed disagreement in that standardized test results do not effectively show student growth.

Table 6

Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component

Themes
Teacher accountability in order to make sure teachers are doing what they are supposed to be doing.
Results do not focus on student growth.

IQ2. As indicated in Table 7, IQ2 (Do you believe that your schooling organization's procedure for utilizing student achievement data as an indicator of performance supports the intended purpose of teacher evaluation? How so?) one common theme arose which was a "No" response that was attributed to a lack of accountability for students on the state exam. For example, T7 responded "No. If students had to be held accountable and the was some way we could hold them accountable, then I'm all for it. But until that happens, no." Additionally, T3 stated

> I don't like it. I don't think that you should evaluate a teacher on how their students do on a standardized test. There's too many factors and students are not held accountable.

They don't get a grade and the test is not a graduation requirement so they might not take it seriously.

Responses to IQ2 were aligned to the disagreement indicated on survey responses and provided more detailed elaboration in that teachers in the accelerated group believed that the results were not a good indicator of performance because the students had little reason to take the test seriously due to lack of accountability. For this reason, teachers appeared to Disagree with the use of standardized test results on teacher evaluations.

Table 7

Student Achievement Data Supporting Intended Purpose of Teacher Evaluation

Theme
No
A lack of accountability with students on the state exam.

IQ5. When reviewing the data in Table 8 related to IQ5 (Describe what you consider to be an effective method of teacher evaluation using standardized testing results?), there was one common theme which was discussed most heavily: Until student accountability is addressed on the state exam, there is not a viable method to use standardized test results on teacher evaluations. T2 stated

I feel like students shouldn't be able to move on if they don't pass it (AzMerit test). So if there was a student component that the expectations on the student were there, then yeah. I think it would be a good evaluation of teachers.

T3 added further confirmation of the accountability theme and stated,

> No. Because if the you have a kid that has a bad attitude and there is no reason to do well it doesn't really measure how well a teacher teaches. If it was for a grade or graduation, then yes.

T7 added further depth and stated,

> Ok, there has to be an accountability piece on there, I definitely think for me, I always go back and look at my students and I know the students who try and don't try. In an administrative way, I don't know how effectively they could be used at this point.

Teacher responses on IQ5 reflected an inability to identify an effective way to use standardized test results to evaluate teachers because many responded that before results could be looked at, high school students needed some form of accountability in order to take the test seriously. The responses supported Theme 1 survey results in that students results were not seen as a viable tool to evaluate the performance of the accelerated group of teachers because the results were not connected to an accountability piece like grades or graduation.

Table 8

Effective Method of Teacher Evaluation using Standardized Testing Results

Theme
Student accountability on the state exam must be addressed prior to identifying a method

IQ6. As revealed in Table 9, there was a single most common theme provided by IQ6 (Do you believe that student standardized testing results serve as an indicator of teacher effectiveness? Why or why not?), which included a response of "No." However, when asked why/why not, two responses emerged. The first was attributed the fact that growth was difficult

to measure across math courses and that students lacking skills had no baseline exam to see how much they grew, even if scores on the test were still low. For example, if a student takes an end of course exam in geometry, there is an inability to use the results to look at growth on an algebra 1-2 math exam given the previous year. The inability to look at annual growth, appeared to create a response of disagreement to IQ6. T2 stated "No, because I can't tell where they came in skill-wise off of a standardized test results. I have my formative assessments to measure their growth so that is why I would trust those results more." Additionally, the concept of a lack of accountability for students emerged among those responding with a No as well. T7 stated

> No. Because there's no accountability. Even in my honors classes, I mean, those are usually high achieving kids, but because they know it doesn't mean anything it's really hard to determine if a teacher is effective based upon the results.

Table 9

Standardized Testing Results Serve as an Indicator of Teacher Effectiveness

Theme
No Annual growth is difficult to identify as the content measured is different for each year. Students are not accountable for the results and many do not take the exam seriously, thus making it difficult to rely on results as a measure of teacher performance.

Qualitative summary RQ1a, theme 1: Group A. Qualitative results with regard to IQ1, IQ2, IQ5 and IQ6 for Theme 1 provided more detail that appeared to support the consistent disagreement toward Theme 1 SQs. However, the theme connected to IQ1 indicated that teachers did understand that the purpose of using standardized tests was to verify that teachers were accountable for teaching standards. This appeared to be contrary to survey results;

however, there were also indications that teachers held beliefs that the evaluation instrument did not meet the intended purpose when considering standardized test results because it was unable to truly measure student growth. IQ2 and IQ5 interview questions revealed more consistent findings to that of survey results for Theme 1. The most common theme that emerged from interviews was the idea that teachers in the accelerated group expressed concern about trusting the results of standardized tests because students are not held accountable in a way that would cause them to take the AzMerit test seriously. Teachers indicated that the test was not a part of a grade for students nor was it a graduation requirement. This attitude may serve to confirm why Theme 1 SQs represented high levels of disagreement toward the use of standardized tests as an effective tool to measure teacher performance or effectiveness in an evaluation instrument. IQ6 yielded similar expressions of concern over student accountability when considering standardized test results as a measure of performance; however, IQ6 also uncovered the idea that the accelerated teachers seemed to feel that annual growth was difficult to measure annually.

Quantitative RQ1a, theme 2 findings: Group A. The survey questions that addressed Theme 2 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) included: SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27. As shown in Table 10, each SQ is listed with the number of respondents for each category on the survey. Additionally, Table 10 summarizes the mean and mode for each of the Theme 2 SQs on part two of the survey instrument implemented for this study.

SQ17. SQ17 (Teachers trust the use of student's performance on standardized tests as a part of the evaluation process) was the first question for Theme 2 on the survey and reflected disagreement toward the statement in this SQ. Five out of six responses fell in the category of Disagree and one in Strongly Disagree. No one selected Neutral, Agree, or Strongly Agree.

SQ18. SQ18 (Administrators trust the use of student's performance on standardized tests as a part of the evaluation process) results were somewhat different from SQ17. Four accelerated teachers were Neutral when considering administrator's trust in the use of standardized test results as a part of the evaluation process. SQ18 yielded four Neutral responses and two Disagree responses. Additionally, the mode for SQ17 (2.00) and SQ18 (3.00) indicated that the most frequent responses for each of the two questions marked a difference between Disagree and Neutral.

SQ21. SQ21 (Results on standardized tests identifies specific areas for professional learning) produced reduced levels of disagreement; of six respondents one teacher Strongly Disagreed and one Disagreed. Three chose Neutral and one picked Agree.

SQ22. SQ22 (Standardized tests help to clarify which learning goals are most important) produced similar findings to that of SQ21 in that there was more agreement to the statement. Both SQ21 and SQ22 yielded a mean of 2.67. However, SQ22 had two responses that fell in the Agree category, one in Neutral, two that were Disagree, and one that was Strongly Disagree. The mode for SQ22 was four, but it is also important to note that 50% of SQ22 responses were either Strongly Disagree or Disagree. The remaining responses were Neutral (one response) and Agree (two responses).

SQ23. SQ23 (Teachers can influence substantially how well their students do on standardized tests) yielded results were quite different from all other Theme 2 SQs. For example, five out of six responses on SQ23 were Neutral. Only one response was in the category Disagree. SQ23 yielded the most Neutral responses in relation to all other Theme 2 SQs.

SQ24. SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area) produced results where four out of six responses were Disagree, one was Strongly Disagree, and one Neutral. The mean as well as the mode for SQ24 was 2.00. Apart from the lone Neutral response, SQ24 yielded more Disagreement.

SQ25. SQ25 (Testing creates a lot of tension for teachers and/or students) responses reflected more agreement than most Theme 2 SQs. For example, SQ25 responses yielded two in the Agree category, two in the Strongly Agree category and two in the Disagree category. Some 67% of Group A respondents demonstrated some level of agreement with the idea that standardized tests cause tension.

SQ26. SQ26 (I expect my students to perform well on tests) replies reflected responses of agreement; two Strongly Agree and two Agree. SQ26, in comparison with other Theme 2 questions, produced the highest level of agreement as well as mean (4.33).

SQ27. SQ27 (Standardized testing is helping schools improve) yielded results that were aligned to SQ17 in that both questions had a mean of (1.83); however, it yielded the most Strongly Disagree responses, out of all Theme 2 questions, being chosen by three Group A teachers.. Of the others, one chose Disagree and two chose Neutral.

Quantitative summary RQ1a, theme 2: Group A. The Theme 2 questions included a wider range of responses focused on teacher attitudes towards the notion that results of standardized tests can/cannot be trusted, provide feedback about instruction, clarify learning goals, and trust towards the use of standardized results in the evaluation process. The range of mean scores is more inclusive of responses that reflect disagreement and neutrality. Theme 2 questions such as SQ21, SQ22, SQ25, and SQ26 yielded responses that reflected agreement of

some accelerated teachers towards using results for professional learning, clarifying learning goals, tension created by standardized tests, and expectations for students.

Table 10

RQ1a, Theme 2: Group A Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)

	SD	D	N	A	SA	Mean	Mode
17. Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	1	5	0	0	0	1.83	2.00
18. Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	0	2	4	0	0	2.67	3.00
21. Results on standardized tests identifies specific areas for professional learning.	1	1	3	1	0	2.67	3.00
22. Standardized tests help to clarify which learning goals are most important.	1	2	1	2	0	2.67	4.00
23. Teachers can influence substantially how well their students do on standardized tests.	0	1	5	0	0	2.83	3.00
24. Standardized tests give me important feedback about how well I am teaching in each curricular area.	1	4	1	0	0	2.00	2.00
25. Testing creates a lot of tension for teachers and/or students.	0	2	0	2	2	3.67	4.00
26. I expect my students to perform well on tests.	0	0	0	4	2	4.33	4.00
27. Standardized testing is helping schools improve.	3	1	2	0	0	1.83	1.00

Qualitative RQ1a, theme 2 findings: Group A. There were two interview questions that aligned to Theme 2 (IQ3 and IQ7). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for Group A of math teachers. There were 10 total teachers interviewed for qualitative purposes in the study. Of the 10 teachers, three in the group were specifically categorized as accelerated teachers (Group A). The interview

responses for the three teachers were used to identify themes for this group and are labeled T2, T3 and T7.

IQ3. As found in Table 11, one common theme surfaced for IQ3 (Do you believe that the use of student standardized testing results on a teacher evaluation instrument is a valid measure of teacher competency?) related to Theme 2: Student Accountability. Interview responses for IQ3 elicited two common themes. The theme emphasized the concept of student accountability. For example, teachers discussed ideas that students are not accountable for standardized test results because the exam is not connected to any accountability mechanisms such as graduation or grades. T2 noted "No. At this time, with students not being held accountable for their standardized tests, I do not think teachers can be held accountable if students are not being held accountable." Additionally, T7 explained, "If students had to be held accountable and there was some way we could hold them accountable, then I am all for it. But until that happens, no."

The interview responses align with SQ responses in that the teachers elaborated on why a lack of trust in the results is present. Responses did not state that the measurement was the problem, but rather that there were not systems in place to raise the level of importance for students when taking the test. Elaboration through interview responses helped to develop a better understanding of Theme 2 SQs regarding why disagreement existed on the part of teachers towards trusting the results of standardizes tests.

Table 11

Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation

Theme
No, because students are not held accountable for the results.

IQ7. As indicated in Table 12, IQ7 (Do you trust the results of student standardized tests as a measure of performance? Why or why not?), one common theme arose: student accountability. When considering student accountability, a number of teachers continued to elaborate on the idea that for results to be meaningful, students need some type of accountability factor that gets them to take the exam seriously. For example, T7 noted "I have kids who open the computer and select A for every single option and shut it in two minutes. There's no accountability as I mentioned before." T2 indicated that

> At this time, no, because it doesn't affect them at all. Obviously, I try to tell my students to try their best and I bring up the fact that maybe one day colleges will look at these test scores. But at this point, there's nothing being held over the students so it's just not the best way to evaluate performance.

The SQs also appeared to align to the IQ statements about the validity of standardized tests in that Group A teachers communicated disagreement with trust in the results; however, the theme of student accountability appeared again in that teachers, during interviews, did not criticize the measure but instead questioned the validity of students results due to a lack of accountability.

Table 12

Trust Student Standardized Tests as a Measure of Performance

Theme
No
Student accountability is lacking.

Qualitative summary RQ1a, theme 2: Group A. Theme 2 qualitative results were very clear as a result of interviewing teachers. The common idea that student accountability is

essential, in the eyes of teachers, surfaced frequently across both IQ3 and IQ7. Group A teachers communicated the concern that results were not valid due to the absence of some method that serves to elevate the importance of the test for students. Previous to the AzMerit era of testing, students were expected to take AIMS as a graduation requirement. Interviews of teachers using Theme 2 questions served to accentuate the concern over valid results without some form of accountability that was present in the years before AzMerit testing.

Quantitative RQ1a, theme 3 findings: Group A. The survey questions that addressed Theme 3 (Actual process of teacher evaluations) included: SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20. As shown in Table 13, each SQ is listed with the number of respondents for each category on the survey. Additionally, Table 13 summarizes the mean and mode for each of the Theme 3 SQs on part two of the survey instrument implemented for this study.

SQ6. SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students) yielded a range of responses. Responses for SQ6 were predominantly comprised of Agree responses with a total of four. Two Group A teachers Disagree, three were Neutral, and one Agreed.

SQ9. SQ9 (Traditional teacher evaluation process [pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator] is an effective tool that can be used to measure classroom performance.) reflected more agreement than that of SQ6. There was also one Neutral, one Disagree, and four Agree in SQ9.

SQ10. SQ10 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance) resulted in higher levels of agreement toward the concept of self-evaluation as a tool to measure performance. There were four out of six responses that Agreed,

one that Strongly Agreed, and one that was Neutral. SQ10 responses, except for one Neutral respondent, all reflected some degree of agreement.

SQ11. SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance) was similar to SQ10 in terms of format, but asked a slightly different question, focusing on student evaluation as a tool to measure performance. Four out of six respondents were Neutral on this statement while two Agreed with SQ11. It is of note that both questions (SQ10 and SQ11) demonstrated no direct disagreement or agreement with only two Agree responses and four Neutral responses.

SQ12. SQ12 (Peer teacher evaluation is an effective tool that can be used to measure classroom performance.) yielded three responses Neutral and three Agree. There were no responses for SQ12 that were in the Disagree or Strongly Disagree categories.

SQ13. SQ13 (Parent evaluation is an effective tool that can be used to measure classroom performance) produced somewhat different results as compared to previous SQs related to different forms of evaluating teacher performance. SQ13 considered parental evaluations and produced three Neutral and three Disagree responses. While all other SQs related to specific forms of evaluation and produced less Disagree responses, SQ13 yielded a result of 50% of respondents disagreeing with the concept of parent evaluation.

SQ16. SQ16 (Professional teaching portfolios [collection of reflections, critiques, lesson plans, samples of student work] is an effective tool that can be used to measure classroom performance.) findings showed that three respondents Agreed and one Strongly Agreed with the statement. There was also one Neutral and one Disagree response. The mode for SQ16 was four, which was indication that Agree was the most frequent response.

SQ19. SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers) produced results that were quite different from previous Theme 3 SQs in that all responses were in either the Strongly Disagree or Disagree range. Four responses were Disagree and two were Strongly Disagree.

SQ20. SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers), like SQ19, produced results that were quite different from previous Theme 3 SQs in that all responses were in either the Strongly Disagree or Disagree range. Both SQs had four responses that were Disagree and two that were Strongly Disagree.

Quantitative summary RQ1a, theme 3: Group A. While Theme 3 questions emphasized the process of teacher evaluation within the schooling organization. A focus on eliciting attitudinal responses from this group of educators around ideas such as traditional processes used in teacher evaluation such as pre-observation, classroom observation, written reports by evaluators, self-evaluation, professional portfolios, peer evaluation, and parent evaluation were included in these survey questions. A review of the mean scores for the Theme 3 survey questions included a range of 1.67 (Strongly Disagree) to 4.00 (Agree) in terms of means scores on individual questions. However, SQ19 and SQ20 specifically focused on including student standardized testing results as an objective of the evaluation process. SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, and SQ16 involved gathering attitudinal responses on more traditional methods of teacher evaluation. The mean scores for these questions were 2.5 (Disagree) to 4.0 (Agree). The responses suggest a more negative response or disagreement among participants in this group to questions that included students standardized tests as an objective of the teacher evaluation instruments.

Table 13

RQ1a, Theme 3: Group A Accelerated Teachers (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)

	SD	D	N	A	SA	Mean	Mode
6. The teacher evaluation process includes a discussion on student standardized test results for students.	0	2	3	1	0	3.17	3.00
9. Traditional teacher evaluation process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	0	1	1	4	0	3.50	4.00
10. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	0	0	1	4	1	4.00	4.00
11. Student evaluation is an effective tool that can be used to measure classroom performance.	0	0	4	2	0	3.33	3.00
12. Peer teacher evaluation is an effective tool that can be used to measure classroom performance.	0	0	3	3	0	3.50	3.00
13. Parent evaluation is an effective tool that can be used to measure classroom performance.	0	3	3	0	0	2.50	3.00
16. Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	0	1	1	3	1	3.67	4.00
19. Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	2	4	0	0	0	1.67	2.00
20. Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	2	4	0	0	0	1.67	2.00

Qualitative RQ1a, theme 3 findings: Group A. There were three interview questions that aligned to Theme 3 (IQ4, IQ8, IQ9). Each question was analyzed in order to identify themes

that support or do not support the quantitative findings for the accelerated group of math teachers. There were 10 total teachers interviewed for qualitative purposes in the study. Of the ten teachers, three in the group were specifically categorized as accelerated teachers. The interview responses for the three teachers were used to identify themes for the accelerated group and are labeled as T2, T3 and T7.

IQ4. As found in Table 14, two common themes surfaced for IQ4 (Do you believe that your schooling organization's teacher evaluation process results in an accurate measure of a teacher's ability to teach? Why or why not?) related to Theme 3: "No", with sub items: administrator time to conduct thorough evaluations and student accountability is required in the instrument; and, "Yes, with sub-items: use of pre-/post-formative assessment data and student accountability.

The first theme that emerged from interviews was that the evaluation process did not result in an accurate measure to teacher ability. Responses focused on the fact that administrators do not have enough time to observe a teacher regularly enough to accurately evaluate performance. For example T2 stated

> I feel like one observation per year is not really a valid measure of our ability. I don't think that's realistic that an admin can come into my class every single day during the school year. There's too many things happening. But the more presence that you have, its better because then admin could support teachers.

Additionally, T3 referenced the frequency of observations and noted "I think it, again, it just gives a snapshot in time. It might be better not to schedule, not to have formal evaluations, but just to do a few walkthroughs."

The second theme that emerged was "Yes" that the evaluation process does yield an accurate measure of teaching ability. For example, T7 explained "Yeah. I think that we do because we are using pre-/post-test data. I can hold students accountable for that. I'm not using a standardized test so yes, I think the data is accurate and effective." The statement referenced that the requirement of the overall process, which allows teachers to use classroom data, served to create better measurements of teacher performance. Both themes identified for IQ4 served to substantiate SQs for Theme 3 in that more traditional forms of teacher evaluation and specific processes such as using student data from the class were more favorable ways to assess performance.

Table 14

Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach

Themes
No
Administrator time to observe enough to get a good perspective of performance.
Yes
The instrument requires use of pre/post formative assessment data. Students can be held accountable for that so data reflects quality of instruction.

IQ8. As indicated in Table 15, replies to IQ8 (Do student's standardized testing results serve as a tool that can influence teacher performance in the classroom? How so?) generated two common themes. The first was "No" and uncovered two sub-items: timely assessment results and use of other assessments. The second theme, "Yes", revealed one sub-theme item: data can be used to identify gaps.

The first theme of "No", revealed in IQ8 was related to the timeframe for when teachers receive standardized tests results. Teachers indicated that because results are delayed, it proves difficult to use them to impact teacher performance for the specific group of students tested. T7 noted "The problem is we don't get the results until we don't have the kids anymore. SO can we affect the kids that took the test? No." An additional sub-item response to the "No" theme indicated that teachers see alternatives forms of assessment as better measures than standardized test results. For example, T2 explained

> No, we have district tests that we've been doing. We actually input them as a grade. We have the kids do a pre and post, and I think that is something we can measure, but the ones the state is giving, they can't be measured because there is no effect on the kids. We treat the district ones as a pre-final that we take a week or two before the finals, and it affects their grade, so yeah, the kids do care.

The second theme, "Yes", emerged in IQ8 discussions in that teachers did see the standardized test results as a tool to identify possible games in instruction and impact classroom performance. For example, T3 stated,

> But if you see, if you really study them and you see that there's a trend, for example, like on the geometry, if all of my kids did poorly on dilations, then I would say I probably didn't do a great job on that. I need to look at that for next year.

Additionally, T7 articulated a similar idea that connected to this theme and stated "I definitely go back and look at that to determine things I need to change in my classroom to make sure that I'm not missing something that's being covered or that's tested." The responses served to support Theme 3 SQs in that standardized test results, according to the accelerated teacher group, can

serve a purpose to influence performance and instruction; however, when used as an evaluative tool there was less agreement toward that idea.

Table 15

Standardized Testing Results Serves as a Tool that can Influence Teacher Performance

Themes
No
Results are not received in a timely manner.
Use of district assessments is better because accountability for students can be connected to them (i.e. used as a grade)
Yes
Data can be used to identify possible gaps in instruction.

IQ9. When reviewing the data in Table 16 related to IQ9 (Do student's standardized testing results serve as a tool that can influence a teacher's professional growth? How so?) one common theme was discussed most heavily: "Yes", standardized test results can be used to influence professional growth. Interview responses for IQ9 uncovered the theme where teachers explained that standardized test results can serve to impact professional growth. The results were reported as a tool that can help to identify possible gaps in instruction which could then influence training options that teachers sought out. For example, T7 stated,

> Yeah. For instance, things on some of the upper level math portions that require calculators that I didn't think about, I, myself might need to figure out how to do them on a calculator. Maybe that's something I need to review so that I can make sure that the things that are being tested are things that they know.

T3 noted "Life if they want to go back and revisit things that they feel that the students did not do well on, and the look to see how I can teach that better, I can be more effective."

Table 16

Standardized Testing Results Serves as a Tool that can Influence Professional Growth

Theme
Yes Results can serve as a tool to inform instruction and seek out training.

Qualitative summary RQ1a, theme 3: Group A. Theme 3 findings identify a number of ideas that support quantitative data obtained in survey questions. Theme 3 IQ questions revealed that teachers view the use of standardized test results as a tool that can inform instruction and even guide professional growth; however, there were also issues that emerged with the timeliness of receiving results and the lack of impact on current students. Additionally, the accelerated group provided evidence through interview responses that classroom data can be used more effectively as a tool to measure performance because there were accountability measures that could be connected to the assessments such as grades and being used as a final. The SQs for Theme 3 also indicated that standardized test results have a purpose, but were not looked upon with agreement when used as a tool for teacher evaluation.

Summary for RQ1a: Group A

The consistency of responses among the accelerated teachers when considering the use of standardized tests to evaluate teachers was evident. Teacher interview responses were indicative that standardized test results were not something they agreed with, which aligns with mean survey responses. During the interview process it became apparent that the teachers were not opposed to the use of standardized tests to measure performance, but consistently expressed that standardized tests lacked the ability to assess teacher effectiveness in an evaluation.

Research Question 1b Findings

What are the attitudes of teachers instructing non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

Quantitative RQ1b findings: Group B. This section will provide summary descriptive statistical information on teachers identified as instructors of high school non-accelerated (Group B) mathematics courses in District A. Data will be presented summarizing mean scores for this group on each of the 28 Likert- scaled survey questions according to Themes 1, 2, and 3.

Table 17 provides the non-accelerated group composition with regard to gender breakdown, highest degree earned, years in the field of education, and years within District A. The group of high school non-accelerated mathematics teachers was comprised of 24 teachers as taken from part one of the survey instrument. When looking at gender, there were 16 females, which equaled twice as many participants (66.7%) than the eight male (33.3%) participants.

The teachers in this group responded equally as having earned a master's degree (16, 50.0%) or holding a bachelor's degree (16, 50.0%). There were no teachers in this group that reported as having earned a doctorate. Respondents also possessed varying levels of experience with regard to years in the profession: there was one teacher in each of the 0, 21-25 and 26-30 years categories, and two teachers had each taught in the categories of 6-10 years and >30 years. Three teachers taught 16-20 years while seven taught 1-5 years. Of the 24 non-accelerated teachers, three had taught a full year in District A, ten had taught in the district 6-10 years, two for 6-10 years, four for 11-15 years, three for 16-20 years, and one each for 26-30 and >30 years. The greatest percentage of non-accelerated teachers reported having an experience level of 1-5

and 11-15 (both at 29.2%). The greatest percentage of teachers within District A reported as having 1-5 years of experience (10, 41.7%).

Table 17

Group B: Gender, Degree, Total Years Teaching, and Years Teaching in District A

Gender

Male		Female		Total	
#	%	#	%	#	%
8	33.3	16	66.7	24	100

Degree

BA		MA		PhD		Total	
#	%	#	%	#	%	#	%
12	50	12	50	0	0	8	100

Years Total Teaching Experience

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
1	4.2	7	29.2	2	8.3	7	29.2	3	12.5	1	4.2	1	4.2	2	8.3	24	100

Years Total Teaching Within District A

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
3	12.5	10	41.7	2	8.3	4	16.7	3	12.5	0	0	1	4.2	1	4.2	24	100

Quantitative RQ1b, theme 1 findings: Group B. The survey questions that addressed Theme 1 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) included: SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28. As shown in Table 18, each SQ is listed with the number of non-accelerated respondents for each category on the survey. Additionally, Table 18 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ1. SQ1 (The use of standardized test results is an effective tool for measuring teacher performance) yielded a wide range of responses, with the highest being ten teachers that Disagreed, while eight were Neutral. There were five teachers who Strongly Disagreed and only one that Agreed.

SQ2. SQ2 (I feel confident that the use of standardized test results can improve teacher performance in the district) produced a wider range of Likert responses than SQ1 in that there was at least one response for every category of the survey. The largest group of respondents fell within the Disagree category (10). Two selected Strongly Disagree, six Neutral, five Agreed, and one Strongly Agreed.

SQ3. SQ3 (Student standardized test scores should be a component of the teacher evaluation process), when compared to the previous two SQs, demonstrated higher levels of strong disagreement because seven respondents Strongly Disagree. Each of the first three SQs all had a mode of 2.00, which indicated that Disagree was the most frequent response to these survey questions. Additionally, the first three SQs all had a mean within the two range, representing disagreement.

SQ4. SQ4 (Student standardized test scores are accurate in assessment of teacher performance) presented different findings when compared to the previous three survey questions. For example, SQ4 had 10 respondents each in the categories of Strongly Disagree and Disagree, while only four were Neutral. The mode for SQ4 was 1.00 and mean was 1.75. This was the first survey question with a mean under two, therefore representing stronger levels of disagreement.

SQ5. SQ5 (Student standardized test scores reflect a teacher's knowledge of teaching practices) followed a similar pattern of disagreement when analyzing the mean (1.79) and mode (1.00). There were 12 Strongly Disagree and seven Disagree responses for SQ5. Three responses were Neutral and two were Agree. SQ5 had no responses in the Strongly Agree category, but there was still a small percentage of agreement on the part of Group B teachers which was similar to SQ4 findings.

SQ7. SQ7 (Student standardized test scores influence a teacher's future teaching performance) demonstrated a similar profile to that of SQ3. SQ7 yielded four Strongly Disagree, six Disagree, eight Neutral, and six Agree responses. The mean was 2.67, with a wide range of responses similarly distributed across the first four survey categories (Strongly Disagree to Agree).

SQ8. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance) followed a unique pattern in that there were seven responses for each of the categories of Strongly Disagree, Disagree and Neutral. Only three responses fell in the Agree category.

SQ14. SQ14 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance) represented the highest number of Neutral responses

(10) in comparison with all other survey questions. Three responses were Agree and the remaining responses fell into the Strongly Disagree (5) and Disagree (6) categories. SQ14 also had a mode of 3.00 and was the only Theme 1 question to have a mode of three other than SQ7.

SQ15. SQ15 (Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.) presented similarly to SQ3 in that similar numbers of responses were distributed across the four survey categories. SQ15 had six Strongly Disagree, eight Disagree, six Neutral, and four Agree responses.

SQ28. SQ28 (I am confident that student's standardized test results accurately measure teaching effectiveness) results were exactly the same as SQ4 in that 20 responses were equally divided across the categories of Strongly Disagree (10) and Disagree (10), and four were Neutral.

Quantitative summary RQ1b, theme 1: Group B. As shown in Table 18, when considering the analysis of mean survey responses on each question of part two of the survey, the non-accelerated group demonstrated a range of mean scores beginning at 1.75 (Strongly Disagree) and going as high as 2.71 (Disagree) for Theme 1. The non-accelerated group demonstrated a wider range of average mean scores across the 10 Theme 1 questions. The three questions (SQ4, SQ5 and SQ28) had mean scores at rating levels 1.75, 1.79, and 1.75 (Strongly Disagree) and seven questions (SQ1, SQ2, SQ3, SQ7, SQ8, SQ14, SQ15 and SQ28) that fell within the range of two (Disagree). SQ4 and SQ28 yielded the lowest means among the group, both at 1.75 (Strongly Disagree). The highest mean score was on SQ7 at 2.67 (Disagree) and SQ14 at 2.46 (Disagree). SQ7 focused on an attitudinal response about the statement "Student standardized test scores influence a teacher's future teaching performance," while SQ14 "Standardized Tests administered to students is an effective tool that can be used to measure classroom performance" sis the same. Overall mean scores remained in the 1-2 range, as

averages, which is an indication that Theme 1 survey responses fell within the Strongly Disagree/Disagree Likert-scaled survey ranges.

Table 18

RQ1b, Theme 1: Group B Non-Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)

	SD	D	N	A	SA	Mean	Mode
1. The use of standardized test results is an effective tool for measuring teacher performance.	5	10	8	1	0	2.21	2.00
2. I feel confident that the use of standardized test results can improve teacher performance in the district.	2	10	6	5	1	2.71	2.00
3. Student standardized test scores should be a component of the teacher evaluation process.	7	8	7	2	0	2.17	2.00
4. Student standardized test scores are accurate in assessment of teacher performance.	10	10	4	0	0	1.75	1.00
5. Student standardized test scores reflect a teacher's knowledge of teaching practices.	12	7	3	2	0	1.79	1.00
7. Student standardized test scores influence a teacher's future teaching performance.	4	6	8	6	0	2.67	3.00
8. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	7	7	7	3	0	2.25	1.00
14. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	5	6	10	3	0	2.46	3.00
15. Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.	6	8	6	4	0	2.33	2.00
28. I am confident that student's standardized test results accurately measure teaching effectiveness.	10	10	4	0	0	1.75	1.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1b, theme 1 findings: Group B. There were four interview questions that aligned to Theme 1 (IQ1, IQ2, IQ5, IQ6). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. Three teachers from the non-accelerated group were interviewed and responses are coded as T1, T5 and T6.

IQ1. As found in Table 19, two common themes surfaced for IQ1 (What do you believe is the intended purpose of using student standardized test results as a component of the evaluation tool within your schooling organization?) related to interview questions: monitor student growth and compliance purposes.

Group B teachers, when interviewed, described the idea that student growth was a specific reason for including standardized testing results as a component of the teacher evaluation system. There was an indication that the purpose was to use a tool to identify if students were progressing and learning, and the scores could provide that information. T5 stated "I believe our organization uses the student achievement data to make sure that we're showing a component of student growth." Additionally, T6 confirmed the same theme and stated "I think it's just numbers for the district to have, to publish whether students are growing or not."

The second theme that emerged was that of compliance. Teachers indicated that state compliance was the purpose for using the data and that it was required so there was not a choice on the part of the district. For example, T1 noted "So I believe the district, at this point, has been using it in a limited facility, as much as anything for compliance purposes." Additionally, T6 stated "Obviously there is a requirement to do it, so district has to compliant."

While many teacher responses on survey questions indicated disagreement with the idea of using standardized test results as a tool to evaluate performance, the interview responses

represented far less disagreement. Teachers focused on the use of standardized tests as a viable measurement, as long as it considered growth of students over time rather than one, isolated score. It became clear that when discussing standardized tests, there was more than one way in which the Group B teachers viewed the results. When considering it as a means to monitor growth, attitudes were more agreeable.

Table 19

Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component

Themes
Monitor Student Growth
Compliance Purposes

IQ2. As indicated in Table 20, IQ2 (Do you believe that your schooling organization's procedure for utilizing student achievement data as an indicator of performance supports the intended purpose of teacher evaluation? How so?) focused on the most common theme that arose: student growth. Teachers indicated that a primary purpose of teacher evaluation is to serve the needs of the students and the use of standardized test results was a way to verify if students were growing or not. For example, T5 stated

> Yes. Yes. I think student growth is a big part of what we do. I think it's just a part of what we do to show growth and make sure that it will move me in a positive direction with our students.

T6 provided a similar response "I think that it looks at student growth and uses it to see how much they grow." Teacher responses provided support for the identification of the theme as student growth was mentioned specifically by non-accelerated teachers. The responses added

important detail in that the non-accelerated group demonstrated an attitude that reflected the purpose of using standardized tests as a tool that District A uses to measure student growth.

Table 20

Student Achievement Data Supporting Intended Purpose of Teacher Evaluation

Theme
Yes
Monitor student growth.

IQ5. When reviewing the data in Table 21 related to IQ5 (Describe what you consider to be an effective method of teacher evaluation using standardized testing results?), the primary theme discussed most heavily was student growth. For example, T5 noted:

> I think the way we use it—looking at student growth, looking at just what a score is, it's hard to dig. You need to see where the student has been, where the student is going to really look at the effectiveness of teaching.

T1 also responded in a similar fashion. "You pretty much have to do it as a growth model. You have to do it as is this teacher improving scores over time rather than just looking at one score." The theme of student growth emerged regularly with the non-accelerated group. IQ1 and IQ2 elicited responses that focused on the use of standardized tests in evaluations as a way to monitor student growth. When asked more directly about the most effective way to use standardized tests in evaluations, the concept of student growth also emerged on a consistent basis. This would support survey data in that the non-accelerated group expressed disagreement, but there were also responses that were Neutral and Agree when considering Theme 1. This might suggest

more willingness to look at standardized testing data as viable tool within the context of performance evaluation.

Table 21

Effective Method of Teacher Evaluation using Standardized Testing Results

Theme
Student Growth

IQ6. As revealed in Table 22, there was one common theme provided by IQ6 (Do you believe that student standardized testing results serve as an indicator of teacher effectiveness? Why or why not?): "Yes", that included two sub-items. The first sub-item was that standardized test results can serve as an indicator of teacher effectiveness if used to compare similar teachers. For example T5 stated "I think it can be, teacher effectiveness, definitely. I think there is a way to compare similar teachers." T6 noted "I believe it does a little bit because there's going to be basic knowledge that students are going to learn that help them with questions. Math teachers can all look and compare results to see that basic knowledge." T1 explained

> One of the things you can do is you can look at results from teacher A, B, C, and D teaching the same subject on the same site and that will give you some indication right there. At the very least, when you're comparing homogenous groups of teachers teaching the same content area in the same manner, it can.

The interview responses align to survey questions for the non-accelerated teachers as Theme 1 responses consisted of more Neutral and Agree within survey categories. It is also important to note while there were more responses reflecting agreement, overall Theme 1 survey questions

reflected disagreement. Therefore, the interview responses do appear to run contrary to the overall survey responses.

The second theme identified was that there are other factors that impact results, but the tests can still provide insight into teacher effectiveness as long as there is consideration of other variables that impact student performance. T5 noted, "There's a lot of other variables that affect the outcome of a standardized test. I deal on a daily basis with a lot of kids that get anxiety when they're taking tests. You need to look at that too." Additionally, T1 explained "You have to watch the kids too though. They could be having a bad day, fought with a parent, or just gave up. Those kinds of things impact student and how they do on tests." The responses to IQ6 align more closely with the consistent disagreement that was voiced through survey responses. The notion that many variables impact student performance suggests that non-accelerated teachers still see that issue as impacting how they might be evaluated using standardized testing results.

Table 22

Standardized Testing Results Serve as an Indicator of Teacher Effectiveness

Theme
Yes
It can serve to compare similar teachers.
Consideration of variables that impact student learning

Qualitative summary RQ1b, theme 1: Group B. Qualitative results with regard to IQ1, IQ2, IQ5 and IQ6 for Theme 1 provided detail that appeared to support the mild disagreement towards Theme 1 SQs. The qualitative themes connected to IQ1 and IQ2 indicated that teachers understand that the purpose of using standardized tests was to monitor how students are progressing and learning. This appeared to be contrary to survey results as there was still

disagreement, but also supported the idea that the non-accelerated group had higher mean scores for a number of Theme 1 survey questions. The most common theme that emerged from interviews was the idea that teachers in the non-accelerated group noted that student growth is a reasonable approach in measuring student learning and teacher effectiveness.

Quantitative RQ1b, theme 2 findings: Group B. The survey questions that addressed Theme 2 (Attitudes towards standardized test results and the degree of trust that participants have/do not have with regard to standardized test results.) for the non-accelerated, Group B, of teachers included: SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27. As shown in Table 23, each SQ is listed with the number of respondents for each category on the Likert scaled survey. Additionally, Table 23 summarizes the mean and mode for each of the Theme 2 SQs on part two of the survey instrument implemented for this study.

SQ17. SQ17 (Teachers trust the use of student's performance on standardized tests as a part of the evaluation process) provided a substantial number of responses that fell into the Strongly Disagree category (14). There were also five Disagree responses, four Neutral, and one Agree.

SQ18. SQ18 (Administrators trust the use of student's performance on standardized tests as a part of the evaluation process) results were quite different from SQ17. Group B teachers yielded a high number of Neutral responses for a total of 14. There were also three Agree and one Strongly Agree. Some disagreement did exist with SQ18 in that only two teachers responded Strongly Disagree and three Disagree.

SQ21. SQ21 (Results on standardized tests identifies specific areas for professional learning) represented more responses of agreement. There were 10 teachers who responded to

SQ21 with Agree and 11 Strongly Agree. Four chose Neutral, three Disagree, and two Strongly Disagree.

SQ22. SQ22 (Standardized tests help to clarify which learning goals are most important.), produced the same mode as SQ21, but had a wider range of responses. For example, six teachers Agreed and three Strongly Agreed; however, four each represented responses of Strongly Disagree and Disagree. Seven responses were Neutral. The profile of SQ22 was quite different from the previous Theme 2 SQs in that there was not one or two survey categories that had a much larger concentration of responses.

SQ23. SQ23 (Teachers can influence substantially how well their students do on standardized tests.) yielded similar findings to others in that three responses each were identified in the categories of Strongly Disagree, Disagree, and Agree. Seven teachers responded Neutral and eight responded Agree.

SQ24. SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area) had a concentration of ten Neutral responses, with seven in the Disagree category. There were no Strongly Agree responses and four Agree. Lastly, there were three Strongly Disagree responses. The mean for SQ24 was 2.63 which still reflected a more consistent level of disagreement, similar to that of SQ18.

SQ25. SQ25 (Testing creates a lot of tension for teachers and/or students) produced the highest mode and yielded ten Strongly Agree responses and eight Agree responses. Lower levels of disagreement were evidenced with one Strongly Disagree response and two Disagree responses.

SQ26. SQ26 (I expect my students to perform well on tests) also was similar to SQ25 in that high levels of agreement were evident. SQ26 had the highest mean (4.25) of all Theme 2

questions with 13 Agree and nine Strongly Agree responses. One respondent chose Neutral and one Disagree.

SQ27. SQ27 (Standardized testing is helping schools improve.) findings were similar to a number of other survey questions in that there was more disagreement reflected in responses. For example, four Strongly Disagree, nine Disagree, nine were Neutral, and two Agree.

Quantitative summary RQ1b, theme 2, Group B. Also as presented in Table 23, the means for Theme 2 questions fell across a wide range from 1.67 to 4.25. It is noteworthy that more disagreement was reflected in SQ17, SQ27, SQ24, and SQ18. SQ25 and SQ26 focused on the tension that may be caused by standardized testing and teacher expectations of student to perform well on tests, both of which produced more agreement.

Table 23

RQ1b, Theme 2: Group B Non-Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)

	SD	D	N	A	SA	Mean	Mode
17. Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	14	5	4	1	0	1.67	1.00
18. Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	2	3	14	3	1	2.91	3.00
21. Results on standardized tests identifies specific areas for professional learning.	2	6	4	10	11	3.17	3.00
22. Standardized tests help to clarify which learning goals are most important.	4	4	7	6	3	3.00	3.00
23. Teachers can influence substantially how well their students do on standardized tests.	3	3	7	8	3	3.21	4.00
24. Standardized tests give me important feedback about how well I am teaching in each curricular area.	3	7	10	4	0	2.63	3.00
25. Testing creates a lot of tension for teachers and/or students.	1	2	3	8	10	4.00	5.00
26. I expect my students to perform well on tests.	0	1	1	13	9	4.25	4.00
27. Standardized testing is helping schools improve.	4	9	9	2	0	2.38	2.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1b, theme 2 findings: Group B. There were two interview questions that aligned to Theme 2 (IQ3 and IQ7). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. Three teachers from the non-accelerated group were interviewed and were coded T1, T5 and T6.

IQ3. As found in Table 24, two common themes surfaced for IQ3 (Do you believe that the use of student standardized testing results on a teacher evaluation instrument is a valid measure of teacher competency?) related to Theme 2. The most common was "No", with the sub-item: if the standardized test is the only measure. The second was "Yes" with the sub-item: if the measure is combined and/or compares similar teachers.

The first theme, "No", provided the reason for being chosen as "if the standardized test is used in isolation." T5 noted "It's tougher with teacher competency. By itself, I would say no." Additionally, T1 explained "There are a lot of ways to use standardized test results, and some are more valid than others. If you only use those it is hard to determine competency." The interview responses helped to provide more understanding of Theme 2 survey questions that reflected higher levels of disagreement on Group B teachers' trust of standardized test results. Interview responses reflected an attitude that results can be trusted, but not as an isolated tool on an evaluation.

The second theme, "Yes", points to the opposite response of the most common theme in that when combined with other factors and used in a fair comparison, standardized test results can serve to help inform an evaluator about competency. For example T1 explained:

> If you are looking at teachers year over year in the same content areas and the same subject areas at the same ability levels, I do think you can look for signs of growth. Or even teacher in the same subject in different districts, I think that works a lot more fairly than just results.

Additionally, T5 stated "That by itself, absolutely not. But with several other measurements, I think it is fine to be used." The interview responses provided insight into the degree of trust that the non-accelerated group had with regard to standardized test results. The interviewees clearly

articulated that as an isolated measure, the results are not reliable. Survey responses served to highlight the disagreement, while interview responses provided more detail about the nature of the disagreement.

Table 24

Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation

Themes
No
If measure is in isolation.
Yes
If measure is combined with other factors/fairly applied to similar teachers.

IQ7. As indicated in Table 25 regarding IQ7 (Do you trust the results of student standardized tests as a measure of performance? Why or why not?), one common theme arose: "No" regarding the current test (AzMerit). Teachers in the non-accelerated group indicated that there was not a high level of trust in the current standardized test used in Arizona. According to T5:

> AzMerit, no. The AIMS test that we had in the past did a really good job of breaking down scores. So they were kind of good to see where a student's weaknesses were. It didn't just give an overall score. It actually broke it down to different components of our standards and helped me look at—I could look the results and say this is an area where a big group of my students are struggling and is something we can focus on.

Tl elaborated in a similar manner "I am not a huge fan of AzMerit as a valid testing instrument. I trusted AIMS a little bit more than I trust AzMerit. There was more information with results and a lot more motivation with the kids."

IQ7 provided more insight into the nature of trust/distrust that non-accelerated teachers had with regard to standardized testing results. While survey questions communicated disagreement towards the trust in results, more detail is provided in that AzMerit, according to interviews, did not provide enough detail about student performance and thus, was not a trustworthy source.

Table 25

Trust Student Standardized Tests as a Measure of Performance

Theme
No
AzMerit does not provide enough information.

Qualitative Summary RQ1b, theme 2: Group B. Theme 2 qualitative results were very clear as a result of interviewing the non-accelerated teachers. The common idea that more detail with regard to the reporting of results is essential for this group of teachers. The non-accelerated group communicated that there was more trust in AIMS results because more information about student performance relative to specific standards was included in reports. Additionally, survey results combined with interview responses suggested that a lack of trust exists on the part of non-accelerated teachers with standardized tests results used in isolation as a tool to measure performance. However, if results were combined with other measures and used to compare similar teaching situations, then more agreement towards using the results as a measure of performance existed on the part of non-accelerated interviewees.

Quantitative RQ1b, theme 3 findings: Group B. The survey questions that addressed Theme 3 included: SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20. As shown in

Table 26, each SQ is listed with the number of Group B respondents for each category on the survey. Additionally, Table 26 summarizes the mean and mode for each of the Theme 3 SQs on part two of the survey instrument implemented for this study. As delineated in Table 26, Theme 3 questions yielded a span of means scores starting at 2.29 (Agree) and ending at 4.09 (Agree).

SQ6. Beginning with SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students.), the breakdown of responses indicated a similar split among three survey categories. Seven respondents picked Disagree, eight Neutral, and eight Agree.

SQ9. SQ9 (Traditional teacher evaluation process [pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator] is an effective tool that can be used to measure classroom performance) yielded results that reflected more agreement towards traditional forms of teacher evaluation. SQ9 produced thirteen Agree responses, four Strongly Agree, two Neutral, and five Disagree. The mode of 4.00 reflects the frequency of Agree responses.

SQ10. SQ10 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance) also yielded a mode of 4.00, indicative of more agreement with the survey question. Ten responses in agreement confirm the mode and there were an additional four Strongly Agree responses. Eight were Neutral and two Disagreed.

SQ11. SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance) findings were similar to SQ10 when analyzing mean (3.50) and mode (4.00). However, further analysis of SQ11 responses indicated less agreement to this survey item. For example, SQ11 yielded two Strongly Disagree responses, one Disagree, seven Neutral, eleven Agree, and three Strongly Agree. While the mean and mode are similar and there are

higher numbers of overall agreement, there was one more response in the categories that reflected Disagreement (Strongly Disagree and Disagree).

SQ12. SQ12 (Peer teacher evaluation is an effective tool that can be used to measure classroom performance) yielded the highest mean across all Theme 3 questions (4.09). These findings can be attributed to the 15 Agree responses by the non-accelerated group. Additionally, there were five Strongly Agree responses and three Neutral responses. Clearly, peer evaluation produced a favorable attitude by the non-accelerated group of teachers.

SQ13. When compared to SQ12, SQ13 (Parent evaluation is an effective tool that can be used to measure classroom performance.) produced less agreement. SQ13 had the lowest mean (2.67) and mode (2.00) across all survey items in Theme 2. SQ13 produced three Strongly Disagree, nine Disagree, six Neutral, five Agree, and one Strongly Agree responses.

SQ16. SQ16 (Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance) produced higher levels of agreement in that there were four Strongly Agree and eleven Agree responses. The question also yielded four Neutral, four Disagree, and one Strongly Disagree responses.

SQ19. SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers) yielded more disagreement as a result of seven Strongly Disagree responses and five Disagree. However, there were also nine Neutral responses and four Agree.

SQ20. SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.) yielded a number of responses with varying

levels of disagreement. There were seven Strongly Disagree, and five Disagree responses. Additionally, there were 10 Neutral and two Agree responses.

Quantitative summary RQ1b, theme 3: Group B. Questions involving specific aspects of traditional teacher evaluation resulted in higher mean scores. For example, SQ9, SQ10, SQ11, SQ12 and SQ16 had averages above three (Neutral). These questions addressed traditional teacher evaluation processes such as self-evaluation, peer evaluation, and professional portfolios. The ranges of mean scores for these questions spanned 3.50 (Neutral) to 4.09 (Agree). All had a mode of four as well, suggesting that more agreement existed with regard to these forms of evaluating teacher performance. However, SQ19, which focused on attitudes towards using standardized test results as an objective of teacher evaluations, had a much lower mean score for non-accelerated teachers as compared to more traditional forms of evaluation. The mean average for SQ19 was 2.50 (Disagree) with a mode of 3.00 (Neutral). Therefore, disagreement and Neutral responses appear to be more common to this question for the group of non-accelerated teachers.

Table 26

RQ1b, Theme 3: Group B Non-Accelerated Teaches (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)

	SD	D	N	A	SA	Mean	Mode
6. The teacher evaluation process includes a discussion on student standardized test results for students.	0	7	8	8	0	3.06	3.00
9. Traditional teacher evaluation process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	0	5	2	13	4	3.67	4.00
10. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	0	2	8	10	4	3.67	4.00
11. Student evaluation is an effective tool that can be used to measure classroom performance.	2	1	7	11	3	3.50	4.00
12. Peer teacher evaluation is an effective tool that can be used to measure classroom performance.	0	0	3	15	5	4.09	4.00
13. Parent evaluation is an effective tool that can be used to measure classroom performance.	3	9	6	5	1	2.67	2.00
16. Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	1	4	4	11	4	3.54	4.00
19. Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	5	6	9	4	0	2.50	3.00
20. Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	7	5	10	2	0	2.29	3.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1b, theme 3 findings: Group B. There were three interview questions that aligned to Theme 3 (IQ4, IQ8, IQ9). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. Three teachers from the non-accelerated group were interviewed and responses were coded as T1, T5 and T6.

IQ4. As found in Table 27, a single common qualitative theme surfaced for IQ4 (Do you believe that your schooling organization's teacher evaluation process results in an accurate measure of a teacher's ability to teach? Why or why not?) related to Theme 3: administrator time to conduct thorough evaluations. The theme that emerged for IQ4 was very clear in that non-accelerated teachers, during interviews, explained that the evaluation process was not an accurate measure of teacher performance due to the fact that administrators did not spend enough time to gather comprehensive data about performance. According to T1 "You don't get observed enough for them to have more than a snapshot idea of your abilities. You don't get observed enough, especially in this district." T5 echoed a similar thought "So our system is decent. But to be able to come into the classroom for a couple of visits and one formal evaluation may not be enough to talk about teacher competence." Additionally, T6 stated "The time and effort that administrators have probably isn't enough to do a perfect evaluation."

The responses to IQ4 do not seem to align to the overall survey questions, especially SQ9 which inquired directly about traditional forms of teacher evaluation that mirror the process in District A. SQ9 produced a mode of 4.00 (Agree) and 17 responses in either the Agree or Strongly Agree survey categories. The level of agreement on Theme 3 SQs does not appear to align with interview themes on IQ3.

Table 27

Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach

Theme
No Not enough time for administrators and determine teacher competence

IQ8. As indicated in Table 28 regarding IQ8 (Do student's standardized testing results serve as a tool that can influence teacher performance in the classroom? How so?), one common theme arose: "Yes" with the sub-item: anything that is a priority can become important in the organization. Each interviewee noted that if a focus exists within the organization on certain practice, then it can impact teacher performance in the classroom. For example, T1 noted "Anything can influence teacher performance in the classroom. I mean a principal can use whatever tool they want to influence teacher's performance in the classroom." Additionally, T5 noted

> Yes, we have used PLC's to see where our student are at. We are seeing where they are going, making changes to what we do. I've found myself going back and looking at about how we can do things differently.

Survey questions for Theme 3 reflected a number of means in the Neutral range of the survey. Additionally, a number of questions had high concentrations of Agree responses for the non-accelerated teachers. This level of agreement suggests that there is alignment with survey responses and IQ8 in that interviewees presented a level of trust that existed in the initiatives that can be focused on within an organization for the purpose of influencing classroom performance.

IQ8 interview responses suggested that use of standardized test results can be one focus within the organization that impacts classroom performance of teachers.

Table 28

Standardized Testing Results serves as a tool to Influence Teacher Performance

Theme
Yes If the organization prioritizes an initiative, the initiative will become important and influence performance.

IQ9. When reviewing the data in Table 29 related to IQ9 (Do student's standardized testing results serve as a tool that can influence a teacher's professional growth? How so?) one common theme was discussed most heavily: Yes, standardized test results can be used to influence professional growth. Interview responses were consistent in that teachers in the non-accelerated group communicated an attitude of agreement that standardized test scores are a tool that can assist with determining what to pursue in terms of professional growth. For example, T1 noted "Well, yes. I think teachers will take their professional growth in the areas that may be weak on testing." Additionally, T5 explained that

> I think that any good teacher should know where their weaknesses are. And looking at standardized tests is just, again, one tool that teachers can use to see where they're at and find out what they don't know because there's always something that we don't know or that we can get better at. And this is just another –and standardized tests are a tool.

Lastly, T6 stated "It might open our ways of teaching to different kids instead of doing our normal style." The higher concentration of Neutral and Agree responses across theme three SQ's suggests alignment to interview responses on IQ9.

Table 29

Standardized Testing Results serves as a tool to Influence Professional Growth

Theme
Yes Results can serve as a tool to inform instruction and seek out training.

Qualitative Summary RQ1b, theme 3: Group B. Theme 3 findings identify a number of ideas that support quantitative data obtained in survey questions. Theme 3 IQ questions revealed that teachers view the use of standardized test results as a tool that can inform instruction and even guide professional growth. The non-accelerated group provided evidence through interview responses that standardized testing data can be used as a tool to measure performance. The SQs for Theme 3 also indicated that standardized test results have a purpose, but were looked upon with some agreement and neutrality when used as a tool combined with other measures for teacher evaluation. Additionally, qualitative interviews provided more insight into attitudes about District A's evaluation tool. For example, survey responses reflected more agreement toward evaluation systems that were similar to District A's approach to teacher evaluation, but interviews revealed less agreement towards the instrument as an effective measure of teacher ability. Lack of time to conduct more observations and classroom visits was cited as a reason that the instrument was not seen as an effective tool to measure teacher ability.

Summary for RQ1b: Group B

The consistency of responses among the non-accelerated teachers when considering the use of standardized tests to evaluate teachers was evident. Teacher interview responses were indicative that standardized test results were not something they agreed with, which aligns with mean survey responses. During the interview process it became apparent that the teachers were not opposed to the use of standardized tests to measure performance, but consistently expressed that standardized tests lacked the ability to assess teacher effectiveness in an evaluation. Discrepancies did exist between survey data and interview questions that related to more traditional forms of teacher evaluation. Survey data communicated agreement towards these forms of evaluation as an effective measure of performance, while interviews revealed less confidence in the District A's evaluation tool. A noteworthy finding to report was that while high levels of disagreement still existed on the part of non-accelerated teachers toward the use of standardized tests on an evaluation instrument, the concept of student growth was presented as a viable tool to measure teacher effectiveness and appeared to be a reasonable tool to include in the evaluation process. However, qualitative interviews also revealed that even student growth should not be used in isolation and combined with other measures of performance.

Research Question 1c Findings

What are the attitudes of teachers that instruct both accelerated and non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

Quantitative RQ1c findings: Group C. The category of teachers who instruct both accelerated and non-accelerated will be referred to as "both" and Group C when describing the group as the teachers are responsible for instructing two types of courses, accelerated and non-

accelerated. Additionally, data will be presented summarizing mean scores for this group on each of the 28 Likert- scaled survey questions according to Themes 1, 2, and 3.

Group C was comprised of 22 teachers and Table 30 provides the group's composition with regard to gender breakdown, highest degree earned, years in the field of education, and years within District A. When considering gender, Group C included 14 females and eight males. The educational level included mostly teachers who earned a master's degree, which was 59.1% of the entire group. There was one teacher in the group that had earned a doctorate and eight that had received a bachelor's degree. Total experience level in the field of education for this group saw the highest percentages (22.7%) distributed across years 21-25. However, the ranges of 6-10, 11-15, and 16-20 included 18.2% each of the overall group with four teachers in each of the experience level categories. Experience ranges 1-5 years and >30 years each included one teacher. When considering experience levels in District A, both 11-15 years and 16-20 years had five (22.7%) each included the highest percentage of teachers. Levels of 1-5 and 21-25 years included three (13.6%) teacher each, while 6-10 years included four teachers (18.2%). Two teacher had >30 years of experience in the district.

Table 30

Group C: Gender, Degree, Total Years Teaching, and Years Teaching in District A

Gender

Male		Female		Total	
#	%	#	%	#	%
8	36.4	14	63.6	22	100

Degree

BA		MA		PhD		Total	
#	%	#	%	#	%	#	%
8	36.4	13	59.1	1	4.5	22	100

Years Total Teaching Experience

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	1	4.5	4	18.2	4	18.2	4	18.2	5	22.7	1	4.5	3	13.6	22	100

Years Total Teaching Within District A

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	3	13.6	4	18.2	5	22.7	5	22.7	3	13.6	0	0	2	9.1	22	100

Quantitative RQ1c, theme 1 findings: Group C. The survey questions that addressed Theme 1 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) included: SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, and SQ28. As shown in Table 31, each SQ is listed with the number of group C

respondents (accelerated and non-accelerated) for each category on the Likert scale. Additionally, Table 31 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ1. SQ1 (The use of standardized test results is an effective tool for measuring teacher performance) represented strong levels of disagreement (95.46%) across the group, yielding 11 Strongly Disagree and 10 Disagree responses. There was only one response in the Agree category.

SQ2. SQ2 (I feel confident that the use of standardized test results can improve teacher performance in the district) demonstrated similar findings as SQ1, but did have slightly more agreement. For example, there were 13 Strongly Disagree and four Disagree responses. Three responses were Neutral and two were Agree; however, 77.27% of responses on SQ2 represented some form of Disagreement.

SQ3. SQ3 (Student standardized test scores should be a component of the teacher evaluation process.) produced similar levels of overall disagreement to that of SQ1, in that 14 Strongly Disagree and six Disagree which represents 90.90% of responses. Additionally, two were Neutral, and no responses fell in the Agree category. When considering SQ1-SQ3, there were only three responses of agreement.

SQ4. SQ4 (Student standardized test scores are accurate in assessment of teacher performance) continued to produce similar levels of disagreement with 15 Strongly Disagree and six Disagree responses. Only one respondent chose Neutral.

SQ5. SQ5 (Student standardized test scores reflect a teacher's knowledge of teaching practices.) produced complete disagreement with all responses falling into one of two categories:

13 Strongly Disagree and nine Disagree. SQ5 was the only Theme 1 question that produced no Neutral or Agree responses among Group C survey participants.

SQ7. SQ7 (Student standardized test scores influence a teacher's future teaching performance) yielded results were indicative of a wider and more even distribution of responses across multiple survey categories. For example, SQ7 produced six Strongly Disagree, six Disagree, five Neutral, and four Agree responses. SQ7 was the most evenly distributed survey item across survey categories and had one of the highest means for Theme 1 (2.33).

SQ8. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance) aligned to previous SQs with a higher level of disagreement. There were 11 Strongly Disagree, nine Disagree, and only two Neutral responses.

SQ14. SQ14 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance) yielded an even number of responses in the categories of Strongly Disagree and Disagree, both at eight. There were four Neutral responses and two Agree.

SQ15. SQ15 (Students' performance on standardized tests is an effective tool that can be used to measure classroom performance) was similar in response to SQ14 in that there were seven Strongly Disagree, eight Disagree, five Neutral, and two Agree responses.

SQ28. SQ28 (I am confident that student's standardized test results accurately measure teaching effectiveness) was similar to previous SQs in that high levels of disagreement were evidenced; 13 Strongly Disagree and eight Disagree. Only one Neutral response was recorded.

Quantitative summary RQ1c, theme 1: Group C. SQ1, SQ2, SQ3, SQ4, and SQ5 all represented high levels of disagreement. When analyzing overall Theme 1 survey responses for Group C, only SQ15 produced a mode of 2, while all others questions had a mode of 1.

Additionally, only SQ7, SQ14, and SQ15 had means in the two range (Disagree). All other SQs yielded means in the one range (Strongly Disagree).

Table 31

RQ1c, Theme 1: Group C Accelerated and Non-Accelerated Teachers (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)

	SD	D	N	A	SA	Mean	Mode
1. The use of standardized test results is an effective tool for measuring teacher performance.	11	10	0	1	0	1.59	1.00
2. I feel confident that the use of standardized test results can improve teacher performance in the district.	13	4	3	2	0	1.73	1.00
3. Student standardized test scores should be a component of the teacher evaluation process.	14	6	2	0	0	1.45	1.00
4. Student standardized test scores are accurate in assessment of teacher performance.	15	6	1	0	0	1.36	1.00
5. Student standardized test scores reflect a teacher's knowledge of teaching practices.	13	9	0	0	0	1.41	1.00
7. Student standardized test scores influence a teacher's future teaching performance.	6	6	5	4	0	2.33	1.00
8. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	11	9	2	0	0	1.59	1.00
14. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	8	8	4	2	0	2.00	1.00
15. Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.	7	8	5	2	0	2.09	2.00
28. I am confident that student's standardized test results accurately measure teaching effectiveness.	13	8	1	0	0	1.45	1.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1c, theme 1 findings: Group C. There were four interview questions that aligned to Theme 1 (IQ1, IQ2, IQ5, IQ6). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the Group C math teachers. Four teachers who instruct both accelerated and non-accelerated courses were interviewed for the purpose of gathering qualitative data. The teachers were labeled T4, T8, T9 and T10.

IQ1. As found in Table 32, two themes surfaced for IQ1 (What do you believe is the intended purpose of using student standardized test results as a component of the evaluation tool within your schooling organization?). The most common theme was: Improve student learning/growth. One sub-item idea that was presented was that the purpose of using standardized tests was it to impact student learning and improve student performance. T4 clearly stated "I think the intent was to improve test scores."

The next most common theme was: link student performance. T10 was focused on students and noted, "They can look at their data as a group and determine what the needs of the students are and how they can improve their instructional practices." T9 also articulated that students were part of the reason for the use of standardized test results, but also included the teacher in that perspective. T9 explained "I think the intended purpose is to try to link how the students are doing with what the teacher's doing. I think that is the purpose." Lastly, T8 communicated a different idea and stated "I don't think we use them that much."

As a result of the analysis it became apparent that the non-accelerated/accelerated group, through interviews, identified improvement of student performance as the primary reason for using standardized test results, but also included the idea that it was a tool to connect student performance to teacher performance. When comparing these responses to the Theme 1 SQ responses, there does appear to be a connection in that teachers believed the purpose is to impact

students, but did not believe it was an effective method for evaluating teachers. SQs mean scores for Theme 1 all fell in the Disagree range when considering standardized test results as an appropriate tool for performance evaluation. IQ1 provides insight into what the group thought was the intended purpose of using standardized tests results.

Table 32

Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component

Themes
Improve Student Learning/Growth
Link student performance to teacher performance

IQ2. As indicated in Table 33 related to IQ2 (Do you believe that your schooling organization's procedure for utilizing student achievement data as an indicator of performance supports the intended purpose of teacher evaluation? How so?), the common theme that arose was "No". Teachers indicated that no, because the process does not measure the variety of other factors that impact students. For example, T4 stated

> And the problem is if my students don't buy into it, I'm not successful and it's evaluated upon me. If my parents don't put emphasis on it, then it's evaluated back on me. If the teachers don't put emphasis on it then it falls back on all of us. I like the challenge, but I need to have all three things (parents, students and teachers) under control. What frustrates me is that two of out three, sometimes I don't feel I have control over.

T8 noted that "I for one am not concerned with it, but if they start tying into you know, riffs or pay, then that's a problem." T10 noted "I say we have made great improvements, but using

formative data instead of standardized tests helps us more. So, no." The responses align with SQs in that both expressed levels of disagreement toward the use of standardized testing results.

Table 33

Student Achievement Data Supporting Intended Purpose of Teacher Evaluation

Theme
No
Does not consider other factors (i.e. parent involvement, student motivation)

IQ5. When reviewing the data in Table 34 related to IQ5 (Describe what you consider to be an effective method of teacher evaluation using standardized testing results?) the primary theme which was discussed most heavily was student growth using pre- and post-testing. Three of the four interview responses made specific mention of the practice using pre- and post-test data to measure student growth. For example, T10 explained

> I think obviously having some sort of pre- and post-tests, like we've tried to do in our PLC's, gives it more individuality instead of comparing teacher to teacher and class to class. You are looking at the students you have, and what they have learned, and how they improved over the course of the year.

The theme of pre- and post-test to measure student growth emerged regularly with the group C (accelerated and non-accelerated). T4 stated "I know the pre- and post-test is a standard go to line. I think those are ok." Additionally, T8 articulated

> The way they currently are, no. There would not be a method since its one time a year. We don't have, or they don't give us much data, and we don't have pre-tests and a post-test to see if they've increased knowledge, then that would be good.

IQ1 and IQ2 elicited responses that focused on the use of standardized tests in evaluations and the inability to measure performance based upon other variables that impact students. When asked more directly about the most effective way to use standardized tests in evaluations, the concept of pre- and post-tests emerged on a consistent basis. This would support survey data in that Group C (accelerated and non-accelerated) expressed disagreement towards using standardized test results as an evaluation tool.

Table 34

Effective Method of Teacher Evaluation using Standardized Testing Results

Theme
Student Growth using pre- and post-testing

IQ6. As revealed in Table 35, there was a common theme provided by IQ6 (Do you believe that student standardized testing results serve as an indicator of teacher effectiveness? Why or why not?), which was "No". The sub-items connected to the no response was that standardized test results are not a viable indicator of teacher effectiveness because there is a lack of student accountability when taking the test in Arizona. For example T4 stated,

> No, because students need to see a reward in it. Why do they want to do well on a standardized test? Everybody has to see value in it. If it's just because it makes me look good on my evaluation, that's not a valid reason.

T10 noted, "It is only a great indicator if for how students are performing as long as there is an incentive for them to do well. AP exams have meaning, but that doesn't really exist for kids with AzMerit." T8 explained that "No, there just needs to be buy in for students."

Table 35

Standardized Testing Results Serve as an Indicator of Teacher Effectiveness

Theme
No
Lack of student accountability

Qualitative Summary RQ1c, theme 1: Group C. Qualitative results with regard to IQ1, IQ2, IQ5 and IQ6 for Theme 1 provided more detail that appeared to support the stronger levels of disagreement towards Theme 1 SQs. The theme connected to IQ1 and IQ2 indicated that Group C teachers believe that the purpose of using standardized testing results was to monitor student growth and link it to teacher performance. This provided more background about why and how this group perceived the use of standardized tests results in evaluations and helped to explain the high levels of disagreement on Theme 1 SQs. The most common theme that emerged from interviews was the idea that Group C teachers was that the use of standardized tests results is not effective because they do not consider student growth and there is no accountability for students to see value in the test.

Quantitative RQ1c, theme 2 findings: Group C. The survey questions that addressed Theme 2 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) for the Group C accelerated and non-accelerated teachers included: SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27. As shown in Table 36, each SQ is listed with the number of respondents for each category on the Likert-scaled survey. Additionally, Table 36 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ17. SQ17 (Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.) had 95.46% of responses in the Strongly Disagree or Disagree range, with 15 Strongly Disagree and six Disagree. SQ17 produced only one Neutral response.

SQ18. SQ18 (Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.) was more balanced in response composition with five Strongly Disagree, six Disagree, eight Neutral, and three Agree. It is noteworthy SQ17 focused on teacher attitudes towards trust of standardized test results, while SQ18 emphasized administrator trust of results.

SQ21. SQ21 (Results on standardized tests identifies specific areas for professional learning.) produced a more balanced range of responses. There were five Strongly Disagree, eight Disagree, five Neutral, and four Agree.

SQ22. SQ22 (Standardized tests help to clarify which learning goals are most important) presented a similar profile to SQ21. SQ22 produced eight Strongly Disagree, seven Disagree, three Neutral, and four Agree responses.

SQ23. SQ23 (Teachers can influence substantially how well their students do on standardized tests.) yielded substantially more Neutral responses (10) than the previous Theme 2 questions. Two respondents Strongly Disagree, five Disagree, four Agree, and one Strongly Disagree. This was the first SQ in Theme 2 that included a Strongly Agree response.

SQ24. SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area.) included seven Strongly Disagree responses, nine Disagree, four Neutral and two Agree.

SQ25. SQ25 (Testing creates a lot of tension for teachers and/or students.) produced the highest mode of Theme 2 questions (5.00). There was only one Disagree response, two Neutral

responses and the remaining responses (19) fell into categories of agreement. There eight Agree and 11 Strongly Agree responses.

SQ26. SQ26 (I expect my students to perform well on tests.) resulted in seven Strongly Agree, nine Agree, four Neutral, and one response each for Disagree and Strongly Disagree. SQ26 also produced a high mode (4.00).

SQ27. Lastly, SQ27 (Standardized testing is helping schools improve.) produced higher levels of disagreement with twelve Strongly Disagree responses, five Disagree, and five Neutral. There were no responses in either of the Agree categories for SQ27.

Quantitative summary RQ1c, theme 2: Group C. SQ25 and SQ26 yielded modes that both fell into categories of agreement with 5.00 and 4.00 respectively. Additionally, SQ18 and SQ23 had Neutral (3.00) modes. All other questions reflected disagreement with modes of either 2.00 or 1.00.

Table 36

RQ1c, Theme 2: Group C Accelerated and Non-Accelerated Teachers (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)

	SD	D	N	A	SA	Mean	Mode
17. Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	15	6	1	0	0	1.36	1.00
18. Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	5	6	8	3	0	2.41	3.00
21. Results on standardized tests identifies specific areas for professional learning.	5	8	5	4	0	2.36	2.00
22. Standardized tests help to clarify which learning goals are most important.	8	7	3	4	0	2.14	1.00
23. Teachers can influence substantially how well their students do on standardized tests.	2	5	10	4	1	2.86	3.00
24. Standardized tests give me important feedback about how well I am teaching in each curricular area.	7	9	4	2	0	2.05	2.00
25. Testing creates a lot of tension for teachers and/or students.	0	1	2	8	11	4.32	5.00
26. I expect my students to perform well on tests.	1	1	4	9	7	3.91	4.00
27. Standardized testing is helping schools improve.	12	5	5	0	0	1.68	1.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1c, theme 2 findings: Group C. There were two interview questions that aligned to Theme 2 (IQ3 and IQ7). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. Four teachers that instruct both accelerated and non-accelerated courses were interviewed for the purpose of gathering qualitative data. The teachers were labeled T4, T8, T9, and T10.

IQ3. As found in Table 37, one common theme surfaced for IQ3 (Do you believe that the use of student standardized testing results on a teacher evaluation instrument is a valid measure of teacher competency?): "No", with sub-items: too many other variables that affect students and pre-/post- test use.

The first sub-item focused on the idea that there are too many other factors that impact students when considering performance, emerged frequently across interview participants. T10 explained

> I would like to say this would be a great way to measure teacher competency if all classrooms were the same, if they had the same number of of 504's and IEP's. There needs to be more attention given to to a more equal playing field for teachers in terms of their demographics in their class.

Additionally, T4 described the theme and stated I think it's part of it. It's not complete. There's too many circumstances that socioeconomics would have a play in it." T9 noted a similar idea

> That's where I don't think it is a valid measure. Because I see too many other variables with the students and how they respond to it and how they feel about it. And a teacher has no control over that.

The interview responses helped to provide more understanding of Theme 2 SQs that reflect higher levels of disagreement on the part of Group C teachers' trust of standardized test results. Interview responses reflected an attitude that results are not valid because there are too many student factors that can impact performance.

The second sub-item was only articulated by one teacher for IQ3, but connected to a similar idea identified in Theme 1 questions, specifically IQ5. The idea of pre- and post-testing emerged consistently in other interview responses, which made it noteworthy to include in

Theme 2. For example, T8 explained, "If we could have pre-tests at the beginning of the year and then a post-test then I think there is some merit to that."

Overall, these interview responses served to back up the level of disagreement on Theme 2 SQs, specifically SQ17 which had high levels of disagreement when considering trust toward standardized results. IQ3 responses add more depth in that teachers indicated that the results are not trustworthy because there are too many factors that impact student performance. Additionally, a common sub-item of the need to pre- and post-tests emerged in this group of teachers across multiple interview questions, which is why it was included as part of the analysis for IQ3.

Table 37

Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation

Theme
No
Too many variables that impact students
There needs to be a pre-/post-test.

IQ7. As indicated in Table 38 related to IQ7 (Do you trust the results of student standardized tests as a measure of performance? Why or why not?), one common theme arose from participants: "No"; however, the reason (sub-item) that each interviewee gave was different. While T4's initial response was "Yes", further analysis of the response still expresses a concern that there are a lot of variables to consider when looking at results. For example,

> I would say yes, but there's a lot of variation in that. I had a stomach ache that day, I don't do good on tests, I got lucky for some reason. I think my issue with it is that you can't control everything that impacts how a student will do on that specific day.

T4 indicated a level of trust, but quickly identified the factors that could impact results. T8 expressed a concern with attendance specifically when asked about the level of trust in standardized test results. T8 stated,

> Nope. If they could break it down to kids who've been here 90% of the time, okay. Then again, being we have one test to judge it by, we don't even see the actual ones, so we don't know if the practice test is applying to the actual test.

T9 also expressed a lack of trust and stated,

> Not completely. Some students, yes. Some students, no. Because the students who truly try, then yes. I do believe that it is a good measure of what they can do. The students who sit there and just bubble it in, it's not a true performance. It's not a true measure of their performance.

T10 included a similar "No" response and cited a lack of student accountability. "So if I were to take the AzMerit for example, I would have to say I would lean more towards no because there is nothing in place right now that students need to really put forth their best effort."

IQ7 provides more insight into the nature of trust/distrust that Group C teachers had with regard to standardized testing results. While SQs communicated disagreement towards the trust in result, more detailed IQ responses confirmed a similar response of disagreement towards trust in the results.

Table 38

Trust Student Standardized Tests as a Measure of Performance

Themes
No
Student factors impact performance
Poor attendance
Student effort
Student accountability

Qualitative summary RQ1c, theme 2: Group C. Theme 2 qualitative results were very clear as a result of interviewing teachers. Results from interview responses provide more detail and align with survey results. For example, there were high levels of disagreement on specific SQs involving trust of standardized test results (SQ17, SQ22, and SQ27). Additionally, IQs provided more insight into the nature of this disagreement. Results suggest that the lack of trust stems from an absence of student accountability and method for implementing pre- and post-tests to measure student growth. Additionally, the idea that there are numerous factors, such as out of the control of teachers, that can impact student performance emerged as a concept that provided more depth of understanding behind survey questions.

Quantitative RQ1c, theme 3 findings: Group C The survey questions that addressed Theme 3 (Actual process of teacher evaluations) included: SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20. As shown in Table 39, each SQ is listed with the number of respondents for each category on the Likert-scaled survey for the non-accelerated group. Additionally, Table 39 summarizes the mean and mode for each of the Theme 3 SQs on part two of the survey instrument implemented for this study.

SQ6. SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students.) produced six Strongly Disagree, six Disagree, seven Neutral, and two Agree responses. The Neutral responses were highest across the four categories, but SQ6 still had more than half of the responses within one of the Disagree categories at 54.55%.

SQ9. **S**Q9 (Traditional teacher evaluation process [pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator] is an effective tool that can be used to measure classroom performance.) findings indicate stronger levels of agreement. For example, three responses were Strongly Agree, 10 Agree, seven Neutral, and only two were Disagree.

SQ10. SQ10 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance.) presented higher levels of agreement than SQ9. For example, SQ10 yielded five Strongly Agree and 10 Agree responses. There was also a wider range of categories covered because there were also three Neutral, two Disagree, and one Strongly Disagree responses.

SQ11. SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance.) displayed the most Neutral responses with nine. Additionally, there were seven Agree, four Disagree, and two Strongly Disagree responses. The findings of SQ11 indicate a wide range of responses.

SQ12. SQ12 (Peer teacher evaluation is an effective tool that can be used to measure classroom performance.) produced high levels of agreement similar to SQ9. There were twelve responses in the Agree category and three in the Strongly Agree category. Additionally, there were five Neutral responses. There was one response in each of the Disagree and Strongly Disagree categories. SQ12 also produced the highest mean (3.68) of all Theme 3 questions.

SQ13. SQ13 (Parent evaluation is an effective tool that can be used to measure classroom performance.) represented more disagreement in that five respondents Strongly Disagree and eight Disagree. Six were Neutral and three chose Agree. The mode of 2.00 served to confirm that there was more disagreement associated with this survey question.

SQ16. SQ16 (Professional teaching portfolios [collection of reflections, critiques, lesson plans, samples of student work] is an effective tool that can be used to measure classroom performance.) had three responses in each of the Strongly Disagree, Disagree, and Neutral categories. There were also 11 Agree and two Strongly Agree responses.

SQ19. SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.) reflect stronger levels of disagreement in that there were 14 responses that Strongly Disagreed and five that Disagreed. There were only three Neutral responses, and no respondent chose Agree

SQ20. SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.) was very similar to SQ19, yielding higher levels of disagreement; 12 Strongly Disagree and seven Disagree, which represented 86.63% of all responses. Only two responses were Neutral and none fell into either of the Agree categories.

Quantitative summary RQ1c, theme 3: Group C. SQ19 (1.50) and SQ20 (1.52) produced the lowest mean scores that both fell into the Strongly Disagree category. Additionally both of the SQs focused on the idea of including standardized test results as an objective of the teacher evaluation process. However, SQ9, SQ10, SQ12, and SQ16 all had mean scores in the Neutral range with 3.64, 3.76, 3.68, and 3.27 respectively. These specific questions focused on participant attitudes toward more traditional aspects of teacher evaluation such as peer evaluation and classroom observations. SQ13 addressed parent evaluation and was met with disagreement

(*M*=2.32). Findings suggest that more neutral attitudes were prevalent with regard to forms of evaluation that were more common within District A's evaluation tool. However, when considering standardized tests results and evaluation, varying levels of disagreement existed on the part of survey respondents.

Table 39

RQ1c, Theme 3: Group C Accelerated and Non-Accelerated Teachers (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)

	SD	D	N	A	SA	Mean	Mode
6. The teacher evaluation process includes a discussion on student standardized test results for students.	6	6	7	2	0	2.24	3.00
9. Traditional teacher evaluation process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	0	2	7	10	3	3.64	4.00
10. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	10	5	3.76	4.00
11. Student evaluation is an effective tool that can be used to measure classroom performance.	2	4	9	7	0	2.95	3.00
12. Peer teacher evaluation is an effective tool that can be used to measure classroom performance.	1	1	5	12	3	3.68	4.00
13. Parent evaluation is an effective tool that can be used to measure classroom performance.	5	8	6	3	0	2.32	2.00
16. Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	3	3	3	11	2	3.27	4.00

Table 39 (continued)

19. Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	14	5	3	0	0	1.50	1.00
20. Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	12	7	2	0	0	1.52	1.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ1c, theme 3 findings: Group C. There were three interview questions that aligned to Theme 3 (IQ4, IQ8, IQ9). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the accelerated group of math teachers. Four Group C teachers who instruct both accelerated and non-accelerated course were interviewed for the purpose of gathering qualitative data. The teachers were labeled T4, T8, T9, and T10.

IQ4. As found in Table 40, a single common theme surfaced for IQ4 (Do you believe that your schooling organization's teacher evaluation process results in an accurate measure of a teacher's ability to teach? Why or why not?) related to Theme 3: "No" with sub-items of: different reasons. The theme that emerged for IQ4 was very clear in that Group C teachers, during interviews, explained that the evaluation process was not an accurate measure of teacher performance due to a variety of reasons. For example, T4 stated, "No, because many of us will not try something new because we don't want it to go poorly in front of our administrators." Additionally, T9 explained:

> There are pieces that give accurate information, and there are pieces that do not. It gives a snapshot, a very, very small snapshot. The testing and using that as part of the

evaluation, that I don't feel is completely accurate. I mean, there might be some truth in it, but it's not completely accurate.

T10 expressed doubt about the accuracy of District A's evaluation process as well; however, cited a different reason as to why this was the perspective.

I know, when you have a rubric or an evaluation tool, that it is supposed to be seamless. However, I have seen where it is filled out differently depending on who the evaluator is. I don't know that it truly measures how effective you are as a teacher.

The interview responses in IQ4 seemed to run contrary to survey responses on traditional forms of teacher evaluation. District A's evaluation process includes all of the components listed in SQ9, and the question produced higher levels of disagreement with a mode of 4.00. However, the interview responses expressed less confidence when asked about the current process used in District A.

Table 40

Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach

Theme
No
Will not try new approaches, want to be successful in observation
Small snapshot, doesn't provide full picture of performance/competency
Inter-rater reliability of administrators

IQ8. As indicated in Table 41, IQ8 (Do student's standardized testing results serve as a tool that can influence teacher performance in the classroom? How so?) found one common theme, "Yes", standardized test results can be used for improvement. Each interviewee noted

that the data can inform them on areas that might be weak and thus, performance can be improved. For example, T4 stated "Absolutely, absolutely. I like the idea of finding areas of weakness and figuring out how I can improve. T8 explained "I think it can, I think you can check yourself to what you did the previous year, and you know the strength of your students." T9 provided a similar statement. "Those results can definitely help influence. It goes back to the ones that we know did truly try on the test, but yes." Lastly, T10 explained

> I would definitely say yes because I think if a teacher at all cares about continuous improvement or a teacher wants to know what they can do better, then absolutely. I mean, I would think you would look at how your students are doing to determine.

The consistent idea that emerged for Theme 3 during interviews was that standardized test results can serve to inform instruction and improve performance. The information helps provide more information about Group C's agreement with standardized tests as a tool. However, Theme 3 survey questions still reflected disagreement towards incorporating the results in an evaluation instrument. Interview responses provide information about attitudes toward the results as a tool to improve, but survey question results suggest that this group was unsupportive of using standardized tests on evaluations.

Table 41

Standardized Testing Results Serve as a Tool to Influence Teacher Performance

Theme
Yes
Continuous improvement/tool to improve instruction

IQ9. When reviewing the data in Table 42 related to IQ9 (Do student's standardized testing results serve as a tool that can influence a teacher's professional growth? How so?), one common theme was discussed most heavily: "No", standardized test results are not to be used to influence professional growth. Interview responses were consistent in that teachers in Group C communicated an attitude of agreement that standardized test scores are a tool that can assist with performance, but are not specific enough to identify areas of professional growth or what to pursue in terms of training. For example, T4 noted, "No, because it would too much emphasis on something that is not that important. My colleagues and materials will help guide my growth." Additionally, T9 stated, "I hope not. The professional growth is more of an individual teacher focus and some need more than others based upon student needs."

Interview responses for IQ9 aligned closely to survey data in that evaluation processes that were more traditional and were not inclusive of standardized test results produced more agreement. Additionally, the interview responses supported survey responses when considering higher levels of disagreement toward the use of standardized testing in teacher evaluation.

Table 42

Standardized Testing Results Serve as a Tool to Influence Professional Growth

Theme
Yes.
Results can serve as a tool to inform instruction and seek out training.

Qualitative summary RQ1c, theme 3: Group C. Theme 3 findings identify a number of ideas that support quantitative data obtained in survey questions. Theme 3 IQs revealed that teachers view the use of standardized tests results as a tool that can inform instruction and but

not guide professional growth. Group C provided evidence through interview responses that standardized testing data can be used as a tool to measure performance. The SQs for Theme 3 also indicated that standardized test results have a purpose, but were not looked upon with agreement when used as a tool combined with other measures for teacher evaluation. Additionally, qualitative interviews provided more insight into District A's evaluation tool and ran contrary to survey results. For example, survey responses reflected more agreement toward evaluation systems that were similar to District A's approach to teacher evaluation, but interviews revealed less agreement toward the instrument as an effective measure of teacher ability. Inter-rater reliability in scoring evaluations and student demographics were cited as factors that could impact standardized test results and ultimately teacher performance evaluation results.

Summary for RQ1c: Group C

The consistency of responses among the Group C (accelerated and non-accelerated) teachers when considering the use of standardized tests to evaluate teachers was evident. Teacher interview responses were indicative that standardized test results were not something they agreed with, which aligns with mean survey responses. During the interview process it became apparent that the teachers were not opposed to the use of standardized tests to measure performance, but consistently expressed that standardized tests lacked the ability to assess teacher effectiveness in an evaluation. Discrepancies did exist between survey data and interview questions that related to more traditional forms of teacher evaluation. Survey data communicated agreement toward these forms of evaluation as an effective measure of performance, while interviews revealed less confidence in the District A evaluation tool. Additionally, while teachers in this group supported the idea of using standardized test results as

tool to improve classroom performance, the respondents did not see it as a viable tool to inform professional growth.

Overall Summary for RQ1

RQ1 inquiries about the attitudes of three teacher groups toward the use of standardized tests results as a measure of performance. Similarities existed across the three groups when considering Theme 1 questions that directly involved the concept of using standardized test results as a measurement of performance. Averages of Theme 1 mean scores on survey questions served to better define the similar attitudes of each group in that there was disagreement toward the use of standardized test results as a tool for evaluation. Each group demonstrated a similar spectrum of responses when considering the lowest to highest mean scores for SQs in Theme 1. Each group had ranges that all fell within the 1-2 rating (Strongly Disagree and Disagree) on the survey. This suggests that means scores may reflect consistent disagreement for each of the three groups.

Theme 2 survey items presented a similar pattern among the three groups in that responses were distributed across ratings of one (Strongly Disagree) to four (Agree*)*. However, for each group there were subtle differences with mean scores for specific questions. SQ21 (Results on standardized tests identify specific areas for professional learning.) yielded a mean score within the rating of Disagree for the Group A and Group C. Group B produced a mean score within the Neutral rating. In reviewing the mean scores, it could suggest that attitudes toward this question were more Neutral for Group B teachers in comparison to Group A and Group C teachers. However, it is also important to consider the mode for the three groups. Group B produced a mode of 4.00, which means that respondents in the Agree rating were most frequent for SQ21. Group A yielded a mode of 3.00 (Neutral) and Group C produced a mode of

2.00 (Disagree). The information presented identifies that there were frequent Neutral responses among Group A teachers with regard to SQ21. SQ22 (Standardized tests help to clarify which learning goals are important.) and SQ23 (Teachers can influence substantially how well their students do on standardized tests.) produced similar results to that of SQ21 in that Group B teachers also yielded mean scores in the Neutral rating. Group A and Group C teachers produced mean scores that fell within the Disagree rating.

Theme 3 questions yielded results that reflected both similarities and differences among the three groups with regard to analysis of mean score survey responses. SQ9 (Traditional teacher evaluation process [pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator] is an effective tool that can be used to measure classroom performance.), SQ10 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance.), SQ12 (Peer teacher evaluation is an effective tool that can be used to measure classroom performance.), and SQ16 (Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work.) is an effective tool that can be used to measure classroom performance.) all produced similar mean scores that fell within the same ratings of Neutral. SQ13 (Parent evaluation is an effective tool that can be used to measure classroom performance.) produced a Disagree rating for all three groups of teachers. SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance.) reflected differences among groups when analyzing mean scores. Group A and Group C teachers yielded mean scores that were Neutral while Group B teachers produced mean scores that fell within the rating of Agree.

Teacher interviews confirmed similar ideas that were represented in survey responses. While SQ responses were indicative of disagreement, IQ responses communicated a similar

attitude in that teachers did not support the idea of using standardize test results as an evaluation method. Additionally, two specific themes emerged as a result of the interviews. Multiple teachers expressed concern with using results because the most recent standardized test given in Arizona lacks accountability for students. High School students took previous standardized tests as a graduation requirement but no longer do so. Teacher concerns emerged from a perceived lack of accountability for students when taking current standardized tests, and that results were, at times, seen as invalid. An additional theme identified through interviews was the idea that a standardized test is a snapshot of performance and many variables can impact results such as student attendance and socioeconomic status. Further, student growth as measured by pre-/post-testing was provided a more effective way to measure teacher effectiveness.

Research Question 2 Findings

Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

$H_0$2. There is no statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

H2. There is a statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

RQ2 was analyzed using the non-parametric Kruskal Wallis Test to compare the three groups of teachers on responses to questions on part two of the survey instrument. The focus of SQs was on the use of standardized test results as an indicator of performance implemented in teacher evaluations. As shown in Table 43, the results of the Kruskal Wallis Test determined if there was a statistically significant difference among the three teacher groups or if it was necessary to retain the null hypothesis ($H_0 2$). If the null hypothesis ($H_0 2$) was retained, then there was not enough evidence to reject the null hypothesis suggesting that there was not a large enough statistical difference. In order to make the determination of whether or not to retain the null hypothesis, an alpha level of .05 was used in the study. Probability levels that were greater than the alpha level of .05 suggested that there was not enough evidence to support a significant difference among the three groups for each survey item. Review of Kruskal Wallis Test results indicated that SQ4, SQ5, SQ7, SQ9, S10, SQ11, SQ12, SQ13, SQ14, SQ15, SQ16, SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, and SQ28 were not statistically significant at the .05 alpha level. This suggests that there is a random sampling/measurement error. The data for all survey questions are presented in Table 43, but the aforementioned survey items will not be discussed as there is no significant difference among the three groups.

A review of the Kruskal Wallis Test results indicated that SQ1, SQ2, SQ3, SQ6, SQ8, SQ19, SQ20, and SQ27 demonstrated a probability level that was less than the alpha of .05, thus suggesting that there was a statistically significant difference among the three groups for these specific survey questions. Table 43 includes the Kruskal Wallis Test results as well as the median for each group. SQ1 (The use of standardized test results is an effective tool for measuring teacher performance.) reveals that there is a significant difference that existed among the three groups due to the fact that $\chi^2 = 8.53$, $p = .014$. The median (2.00) for Group B appeared

to be higher than Group A (1.50) and Group C (1.50). The median of 2.0 for Group B suggests that these teachers believed that there is slightly more agreement when considering the use of standardized test results as an effective tool for measuring teacher performance; however, all of the medians represent a score on that represents disagreement. Similar analysis can be taken from Table 43 for each of the questions that produced a probability level less than the .05 alpha. SQ2 (I feel confident that the use of standardized test results can improve teacher performance in the district.) produced $x^2 = 11.30$, $p = .004$. In this case, the medians were different for Group A (2.00), Group B (2.50), and Group C (1.00). SQ3 (Student standardized test scores should be a component of the teacher evaluation process.) yielded $x^2 = 7.46$, $p = .024$ with a median of 2.00 for Group A and Group B, while the median for Group C was 1.00. SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students.) yielded $x^2 = 8.13$, $p = .017$. Group A and Group B both produced a median of 3.00, while Group C had a median of 2.00. The data represents Neutral responses for Group A and Group B, and disagreement with Group C. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.) produced $x^2 = 6.11$, $p = .047$. In this case the Group B median was 2.00 while Group A and Group C were 1.50; however, there is still an indication that all groups disagreed with the statement. SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.) findings were $x^2 = 12.12$, $p = .002$. The medians reflected difference among the three groups. For example, Group A (median, 2.00) and Group C (median, 1.00) both reflected disagreement; however, the Group B yielded a median of 3.00 which highlighted a neutrality for SQ19. SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.) findings were $x^2 = 7.71$, $p = .021$. The median for Group A was 2.00 and for Group B was 2.5.

The median for Group C was 1.00. Although there were differences across medians for each group, all demonstrated varying degrees of disagreement. SQ27 (Standardized testing is helping schools improve.) produced $\chi^2 = 6.72$, $p = .035$ and medians of 1.50 (Group A), 2.00 (Group B) and 1.00 (Group C). Similar to SQ20, all medians for SQ27 were different but still reflected varying levels of disagreement.

The results demonstrate the differences that exist among the groups with regard to responses on eight total survey questions. Additionally, 20 survey questions did not meet the alpha level of .05 in order to be considered statistically significant. This means that there were 20 total SQs where the null hypothesis ($H_0 2$) was retained. The data suggest that responses on approximately 71% of survey questions yielded enough evidence to retain the null hypothesis. Additionally, of the SQs that represented a significant difference, SQ6 and SQ19 yielded differences in medians among the three groups which represented a neutrality and disagreement. For example, SQ6 produced medians of 3.00 for both Group A and Group B, whereas Group C yielded a median of 2.00. The data suggest that Group C Disagree in SQ6 finding, but Group A and Group B were Neutral. SQ19 followed a similar pattern in that two groups (Group A and Group C) represented a level of disagreement on the survey, where Group B was Neutral.

Table 43

Kruskal Wallis Comparisons of Accelerated, Non-accelerated and Both Groups

SQ	μ	*df*	P (Asymp. Sig.)	Medians		
				ACCEL	NONACCEL	BOTH
SQ1	8.53	2	.014	1.50	2.00	1.50
SQ2	11.30	2	.004	2.00	2.50	1.00
SQ3	7.46	2	.024	2.00	2.00	1.00
SQ4	5.62	2	.060	1.00	2.00	1.00
SQ5	1.40	2	.496	1.50	1.50	1.00
SQ6	8.13	2	.017	3.00	3.00	2.00
SQ7	3.49	2	.175	2.00	3.00	2.00
SQ8	6.11	2	.047	1.50	2.00	1.50
SQ9	0.32	2	.852	4.00	4.00	4.00
SQ10	.904	2	.636	4.00	4.00	4.00
SQ11	4.20	2	.122	3.00	4.00	3.00
SQ12	4.68	2	.096	3.50	4.00	4.00
SQ13	1.14	2	.567	2.50	2.50	2.00
SQ14	2.66	2	.265	2.00	3.00	2.00
SQ15	1.07	2	.586	2.50	2.00	2.00
SQ16	0.65	2	.723	4.00	4.00	4.00
SQ17	3.61	2	.164	2.00	1.00	1.00
SQ18	3.14	2	.208	3.00	3.00	2.50
SQ19	12.12	2	.002	2.00	3.00	1.00
SQ20	7.71	2	.021	2.00	2.50	1.00
SQ21	5.53	2	.063	3.00	3.50	2.00
SQ22	5.22	2	.074	2.50	3.00	2.00
SQ23	2.02	2	.365	3.00	3.00	3.00
SQ24	5.39	2	.068	2.00	3.00	2.00
SQ25	1.46	2	.483	4.00	4.00	4.50
SQ26	1.37	2	.504	4.00	4.00	4.00
SQ27	6.72	2	.035	1.50	2.00	1.00
SQ28	2.76	2	.251	1.00	2.00	1.00

Research Question 3 Findings

What are the attitudes of high school administrators regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

Quantitative RQ3 findings: Administrators. The group of Administrators was comprised of 20 individuals; 10 were female and 10 were male, equaling an even split between gender. Some 17 members of this group earned a master's degree and three earned a doctorate degree. Administrative experience was mostly concentrated within two bands, 1-5 years with eight (40%) administrators and 11-15 years with five (25%) administrators.

Experience bands of 6-10 and 16-20 had three members each, and there was one administrator in the 26-30 experience band. When considering experience within District A, the largest percentage was consolidated in 1-5 years with nine (45%) people. There were five (25%) in the 6-10 year range, four (20%) in 11-15, and two (10%) in 16-20. Table 44 summarizes the gender, education and experience of the administrator group. Additionally, of the 20 administrators five served at the district level, five as high school principals, and 10 as highs school assistant principals.

Table 44

Administrators: Gender, Degree, Total Years in Administration, and Years as an Administrator in District A

Gender

Male		Female		Total	
#	%	#	%	#	%
10	50	10	50	20	100

Degree

MA		PhD		Total	
#	%	#	%	#	%
17	85	3	15	20	100

Years Total Administrative Experience

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	8	40	3	15	5	25	3	15	0	0	1	5	0	0	20	100

Years Total as Administrator in District A

0		1-5		6-10		11-15		16-20		21-25		26-30		>30		Total	
#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
0	0	9	45	5	25	4	20	2	10	0	0	0	0	0	0	20	100

Quantitative RQ3, theme 1 findings: Administrators. The survey questions that addressed Theme 1 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) for the administrator group included: SQ1, SQ2, SQ3, SQ4,

SQ5, SQ7, SQ8, SQ14, SQ15, SQ28. As shown in Table 45, each SQ is listed with the number of respondents for each category. Additionally, Table 45 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ1. SQ1 (The use of standardized test results is an effective tool for measuring teacher performance.) indicated that 50% of respondents Disagree with the statement. However, there were also eight Neutral, one Agree, and one Strongly Agree responses.

SQ2. SQ2 (I feel confident that the use of standardized test results can improve teacher performance in the district.) produced slightly more disagreement than SQ1. For example, SQ2 also had 10 responses in the Disagree category, but also had two in the Strongly Disagree category. Four chose Neutral and four picked Agree. Both SQ1 and SQ2 shared the same mode of 2.00 and similar means that fell in the Disagree category.

SQ3. SQ3 (Student standardized test scores should be a component of the teacher evaluation process.) had nine respondents Disagree and one Strongly Disagree. Five each picked Neutral and Agree.

SQ4. SQ4 (Student standardized test scores are accurate in assessment of teacher performance.) demonstrated slightly higher levels of disagreement across the administrator group. There were 14 Disagree, one Strongly Disagree, three Neutral, and two Agree responses.

SQ5. SQ5 (Student standardized test scores reflect a teacher's knowledge of teaching practices.) was almost identical to SQ4. Both questions had the same number of Disagree responses (14) and Strongly Disagree responses (1); however, SQ5 had two Neutral and three Agree responses.

SQ7. SQ7 (Student standardized test scores influence a teacher's future teaching performance.) represented a wider range of responses, for example, there were two Strongly Disagree, six Disagree, nine Neutral, and three Agree.

SQ8. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.) profiled in a similar manner with one Strongly Disagree, eight Disagree, seven Neutral, two Agree and one Strongly Agree.

SQ14. SQ14 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance.) was similar to SQ8 in that it produced one Strongly Disagree, eight Disagree, seven Neutral, and four Agree responses.

SQ15. SQ15 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance.) yielded one Strongly Disagree, ten Disagree, four Neutral, and five Agree.

SQ28. SQ28 (I am confident that student's standardized test results accurately measure teaching effectiveness.) resulted in five Strongly Disagree, seven Disagree, five Neutral, and three Agree.

Quantitative Summary RQ3, theme 1: Administrators. Theme 1 SQs were relatively consistent as all had a mean within the two range (Disagree). Modes were similar in that nine of the ten questions were 2.00, and one was 3.00 (SQ7).

Table 45

RQ3, Theme 1: Administrators (SQ1, SQ2, SQ3, SQ4, SQ5, SQ7, SQ8, SQ14, SQ15, SQ28)

	SD	D	N	A	SA	Mean	Mode
1. The use of standardized test results is an effective tool for measuring teacher performance.	0	10	8	1	1	2.65	2.00
2. I feel confident that the use of standardized test results can improve teacher performance in the district.	2	10	4	4	0	2.50	2.00
3. Student standardized test scores should be a component of the teacher evaluation process.	1	9	5	5	0	2.75	2.00
4. Student standardized test scores are accurate in assessment of teacher performance.	1	14	3	2	0	2.30	2.00
5. Student standardized test scores reflect a teacher's knowledge of teaching practices.	1	14	2	3	0	2.35	2.00
7. Student standardized test scores influence a teacher's future teaching performance.	2	6	9	3	0	2.65	3.00
8. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	2	8	7	2	1	2.60	2.00
14. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	1	8	7	4	0	2.70	2.00
15. Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.	1	10	4	5	0	2.65	2.00
28. I am confident that student's standardized test results accurately measure teaching effectiveness.	5	7	5	3	0	2.30	2.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ3, theme 1 findings, Administrators. There were four interview questions that aligned to Theme 1 (IQ1, IQ2, IQ5, IQ6). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the administrator group. There were five total administrators interviewed for qualitative purposes in the study. Of

the five administrators, one in the group was a district level administrator, two were high school principals, and two were high school assistant principals. The interview responses from the five administrators were used to identify themes for the overall group.

IQ1. As found in Table 46, two common themes surfaced for IQ1 (What do you believe is the intended purpose of using student standardized test results as a component of the evaluation tool within your schooling organization?) related to Theme 1: teacher accountability and monitor student growth.

The first theme that emerged stated that the purpose of using standardized test results was to measure teacher effectiveness by holding them accountable for teaching the standards. The district level administrator (DA) noted "I think one of the reasons we use standardized tests is to help hold teachers accountable for teaching all the standards, and it's a way we can monitor if they have taught their students to master the state standards." Additionally, a high school principal (HP1) stated "I believe the intended purpose is to measure teacher effectiveness, math teacher effectiveness." Further, an assistant principal (AP1) stated "So just ensuring that those teachers are fulfilling their curriculum map or their content standards." A clear theme emerged with regard to monitoring instruction of standards.

The second theme that emerged was in reference to monitoring student learning. For example, AP2 noted "I believe the intended purpose of using students' standardized test results is to monitor student growth within our district." HP2 explained a similar idea "I feel like standardized testing is to gauge the amount of content knowledge that the students have received from the beginning of the year until the end."

In reviewing survey data, it was clear that higher levels of disagreement existed for administrators with regard to standardized test results and measurement of performance. The

interview responses added depth to understanding this disagreement area in that they clearly articulated purposes that focused on teacher accountability.

Table 46

Belief of Intended Purpose of Student Standardized Test Results as Evaluation Tool Component

Themes
Teacher accountability relative to instruction of standards
Monitor student growth/learning

IQ2. As indicated in Table 47 related to IQ2 (Do you believe that your schooling organization's procedure for utilizing student achievement data as an indicator of performance supports the intended purpose of teacher evaluation? How so?), one common theme arose which was "No" regarding the belief that standardized test results do not consider other factors that impact student learning. Responses from each of the five administrators, to some degree, communicated this theme. For example, DA exclaimed:

> My initial reaction is no. Because we don't take into account a lot of other factors, such as the socioeconomic status of the student, the student's background, perhaps they just immigrated to our country, perhaps they just moved into our neighborhood or our school district. They're a lot of other factors that are beyond the control of the school or teacher, that can affect standardized test scores.

HP1 echoed a similar thought:

> I think there are many other tools that we can use to gauge student growth. I mean some kids just aren't good at it. Some get nervous at tests, I think there are very few children that are really good at taking those types of tests.

Additionally, AP2 stated "There's far more elements that go into teacher performance besides the performance of their particular students on the standardized test. I don't think it's an accurate tool and certainly can't be used in isolation." The high school assistant principals also expressed a similar idea. For example, AP1 said, "I think there are still some things we need to define with achievement. Students have so many issues that can impact performance, making it very difficult." Lastly, AP2 noted "There are a number of considerations to look at with students and sometimes that impacts how they do, so it is very difficult to do."

Table 47

Student Achievement Data Supporting Intended Purpose of Teacher Evaluation

Theme
No
Student factors impact results.

IQ5. When reviewing the data in Table 48 related to IQ5 (Describe what you consider to be an effective method of teacher evaluation using standardized testing results?), there was one common theme which was discussed most heavily: Student growth/pre-tests and post-tests. A clear theme emerged for IQ5 in that student growth and using pre-test/post-test model to measure growth was deemed as an effective way to use results. For example, DA noted, "Well if we look at it on an individual basis, by student and their growth rather than by what they do on a standardized test." Additionally, AP1 stated "I think when you look at pre- and post-tests, there's a lot of validity to that approach." HP1 said, "We need to look at it in terms of yearly growth and more than one test." The interview responses clearly explained why the there was

disagreement on Theme 1 survey responses, as the administrators communicated other alternatives over using standardized test scores to measure and evaluate performance.

Table 48

Effective Method of Teacher Evaluation using Standardized Testing Results

Theme
Student growth/pre- and post-tests

IQ6. As revealed in Table 49, there was a single most common theme provided by IQ6 (Do you believe that student standardized testing results serve as an indicator of teacher effectiveness? Why or why not?), which was "No". There a number of sub-item responses that accompanied the "no" response when administrators elaborated on IQ6, for example, noting that test results are only a snapshot and are unable to serve as an indicator of effectiveness. HP1 stated "No, not in and of themselves. Again, just using one indicator to determine teacher effectiveness, that's irresponsible." HP2 noted "No, I don't. It is just not a broad enough scope. It is not well-rounded tool and provides a snapshot." DA also responded "no" and explained, "I really don't, because there are too many other factors. There are so many things that could impact how students perform." AP1 also noted "There are too many student dynamics to use standardized tests as a good tool." The responses clearly aligned to levels of disagreement that were seen on survey responses.

Table 49

Standardized Testing Results Serve as an Indicator of Teacher Effectiveness

Theme
No
One snapshot of performance
Too many other variable that impact student performance

Qualitative summary RQ3, theme 1: Administrators. Qualitative results with regard to IQ1, IQ2, IQ5 and IQ6 for Theme 1 provided more detail that appeared to support the consistent disagreement toward Theme 1 SQs. Many of the concepts teachers identified were also communicated by administrators. Factors such as student growth and pre-/post-testing were identified as alternative methods that can be used for measuring growth. Additionally, many of the administrators responded similarly to that of the survey questions, noting that standardized tests provide a one-time, limited view of student performance. The information gathered from interviews supported Theme 1 survey responses.

Quantitative RQ3, theme 2 findings: Administrators. The survey questions that addressed Theme 2 (Concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness) for the administrator group included: SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27. As shown in Table 50, each SQ is listed with the number of responses for each category on the survey. Additionally, Table 50 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ17. SQ17 (Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.) yielded the lowest mean (1.80) and mode (1.00) across all Theme

2 SQs. There were nine Strongly Disagree, seven Disagree, three Neutral, and one Agree responses.

SQ18. SQ18 (Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.) produced slightly higher levels of agreement with two Strongly Disagree, nine Disagree, five Neutral, and four Agree responses.

SQ21. SQ21 (Results on standardized tests identifies specific areas for professional learning.) yielded higher levels of agreement with responses of one Strongly Disagree, four Disagree, six Neutral, eight Agree and one Strongly Agree. The mean was also one of the higher values across Theme 2 questions, but still reflected neutrality (3.20).

SQ22. SQ22 (Standardized tests help to clarify which learning goals are most important.) produced similar findings to that of SQ21. SQ22 had one Strongly Disagree, six Disagree, five Neutral, six Agree, and two Strongly Agree responses.

SQ23. SQ23 (Teachers can influence substantially how well their students do on standardized tests.) indicated stronger agreement than disagreement. Administrator responses were highest with 10 in the Agree and two in the Strongly Agree categories. Five were Neutral, and only two response fell into the Disagree category. There was only one Strongly Disagree response.

SQ24. SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area.) had no Strongly Disagree responses, but had four Disagree, nine Neutral, and seven Agree responses.

SQ25. SQ25 (Testing creates a lot of tension for teachers and/or students.) yielded the highest mean of 4.15 among Theme 2 questions. Responses were primarily concentrated in Strongly Agree (6) and Agree (12). There was only one response each for Neutral and Disagree.

SQ26. SQ26 (I expect my students to perform well on tests.) also reflected high levels of agreement. SQ26 had the single most Agree responses in Theme 2 with 19. There was also one Neutral response.

SQ27. SQ27 (Standardized testing is helping schools improve.) yielded responses of Strongly Disagree (1), Disagree (7), Neutral (7), and Agree (5).

Quantitative summary RQ3, theme 2: Administrators. SQ17 yielded the lowest mean (1.80) and mode (1.00) across all Theme 2 SQs, reflecting more disagreement. SQ18 produced slightly higher levels of agreement with a mean of 2.55 and mode of 2.00, however there were four Agree responses. Both SQ21 and SQ22 had a mode of 4.00 and similar means (SQ21, 3.20; SQ22, 3.10). SQ24 produced a mean of 3.15 and mode of 3.00. SQ25 and SQ26 yielded a mode of 4.00. SQ26 produced a mean that was just under four at 3.95, while SQ25 produced a mean of 4.15. SQ28 findings indicated that there while the mean was 2.80 (Disagree), the mode was 3.00 (Neutral).

Table 50

RQ3, Theme 2: Administrators (SQ17, SQ18, SQ21, SQ22, SQ23, SQ24, SQ25, SQ26, SQ27)

	SD	D	N	A	SA	Mean	Mode
17. Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	9	7	3	1	0	1.80	1.00
18. Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	2	9	5	4	0	2.55	2.00
21. Results on standardized tests identifies specific areas for professional learning.	1	4	6	8	1	3.20	4.00
22. Standardized tests help to clarify which learning goals are most important.	1	6	5	6	2	3.10	4.00
23. Teachers can influence substantially how well their students do on standardized tests.	1	2	5	10	2	3.50	4.00
24. Standardized tests give me important feedback about how well I am teaching in each curricular area.	0	4	9	7	0	3.15	3.00
25. Testing creates a lot of tension for teachers and/or students.	0	1	1	12	6	4.15	4.00
26. I expect my students to perform well on tests.	0	0	1	19	0	3.95	4.00
27. Standardized testing is helping schools improve.	1	7	7	5	0	2.80	3.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ3, theme 2 findings: Administrators. There were two interview questions that aligned to Theme 2 (IQ3 and IQ7). Each question was analyzed in order to identify themes that support or do not support the quantitative findings for the administrative group. There were five total administrators interviewed for qualitative purposes in the study. Of the five administrators, one in the group was a district level administrator, two were high school assistant principals, and two were high school assistant principals. The interview responses from the five administrators were used to identify themes for the overall group.

IQ3. As found in Table 51, one common theme surfaced for IQ3 (Do you believe that the use of student standardized testing results on a teacher evaluation instrument is a valid measure of teacher competency?) related to Theme 2: No response accompanied by elaboration the existence of significant factors that impact student performance on standardized tests. For example, DA stated,

> No, I don't. There are too many factors that can influence or affect a student's performance on a standardized test. I don't think one test once a year is a good indicator of what the student accomplished throughout the year.

Additionally, HP2 stated

> There are some phenomenal teachers that work very hard to get the growth that they do get and with students coming with diverse prior knowledge, and family situations, socioeconomic situations—I mean, there is just so much that comes in.

AP2 articulated a similar idea "No. I think there's so many variables that go into student testing that really are not components of how competent a teacher is in the classroom." The interview responses recorded for IQ3 demonstrate strong alignment to Theme 2 survey responses in that disagreement exists with regard to using standardized tests as a tool to measure performance.

Table 51

Belief of Student Standardized Testing Results as a Validity Measure on Teacher Evaluation

Theme
No
Numerous factors impact student performance on standardized tests.

IQ7. As indicated in Table 52 IQ7 (Do you trust the results of student standardized tests as a measure of performance? Why or why not?), one common theme arose: "No"; however, there were a number of different responses when asked why or why not. For example, issues such as results lacking detail was identified by DA. "They don't really break it down. They're so broad in general, that we can't see specifically where a kid's lacking in skills." AP1 articulated a similar idea "You are either proficient, minimally proficient and so forth. But what does that mean and try to define it to teachers and students." Another idea that emerged was that of the test being new. For example, AP2 noted, "I trusted the ones a little bit prior to the newer test. I think that when you're still piloting to some extent that you can't take the results too seriously." Lastly, student accountability was cited as a contributing to factor to a lack of trust in the results of standardized tests. For example AP2 noted "It was a graduation requirement or something that the kids actually took seriously, it's a whole different ballgame. But because it's not, I don't truly believe that I can trust the results." Clear distrust of the results existed and aligned with survey results as well; however, interviews provided more detail about the nature of that distrust.

Table 52

Trust Student Standardized Tests as a Measure of Performance

Theme
No
Results are not detailed
Test is new
Student Accountability

Qualitative Summary RQ3, theme 2: Administrators. Theme 2 qualitative results were very clear as a result of interviewing administrators. The common idea that student accountability, student growth, and student variables impact the faith that administrators have in standardized test results surfaced frequently across both IQ3 and IQ7. The administrator group communicated the concern that results were not valid due to the many factors that can impact student performance. Previous to the AzMerit era of testing, students were expected to take AIMS as a graduation requirement, which was also a way to elevate the importance of the test to students and the absence of this accountability measure impacted the level of trust administrators had towards the current test in Arizona.

Quantitative RQ3, theme 3 findings: Administrators. There were nine survey questions that addressed Theme 3 (Actual process of teacher evaluations) for the administrator group and included: SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20. As shown in Table 53, each SQ is listed with the number of responses for each category on the survey. Additionally, Table 53 summarizes the mean and mode for each of the Theme 1 SQs on part two of the survey instrument implemented for this study.

SQ6. SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students.) had five Disagree responses, while the remaining were responses were either Neutral (4) or reflected varying degrees of agreement (Agree, 8; Strongly Agree, 2).

SQ9. SQ9 (Traditional teacher evaluation process [pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator] is an effective tool that can be used to measure classroom performance.) yielded even less disagreement than SQ6 with only two Disagree and four Neutral responses. There were 12 Agree and two Strongly Agree responses.

SQ10. SQ10 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance.), similar to SQ6, had higher levels of disagreement. There were four responses of Disagree, five Neutral, 10 Agree, and one Strongly Agree.

SQ11. SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance.) produced responses in the Disagree category. Five were Neutral and the remaining responses were Agree (13) and Strongly Agree (2). SQ11 also had the second highest mean (3.85) across Theme 3 questions.

SQ12. SQ12 (Peer teacher evaluation is an effective tool that can be used to measure classroom performance.) yielded stronger levels of agreement with 12 Agree and two Strongly Agree responses; however, there two Disagree and one Strongly Disagree. SQ12 had one Neutral response.

SQ13. SQ13 (Parent evaluation is an effective tool that can be used to measure classroom performance.) reflected more disagreement on the part of administrators with one Strongly Disagree response and eight Disagree responses. There were also eight Neutral and three Agree responses.

SQ16. SQ16 (Professional teaching portfolios [collection of reflections, critiques, lesson plans, samples of student work] is an effective tool that can be used to measure classroom performance.) had little disagreement and the highest mean (4.05) across all Theme 3 questions with 11 Agree and six Strongly Agree responses. There were only two Disagree and one Neutral responses.

SQ19. SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.) had ten Disagree and one Strongly Disagree responses. There were also three Neutral and six Agree responses.

SQ20. SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.) had similar response profiles as SQ19. SQ20 had eleven Disagree and two Strongly Disagree responses. There were also two Neutral and five Agree responses.

Quantitative summary RQ3, theme 3: Administrators. Both questions had similar means SQ19 (2.70) and SQ20 (2.50) and each had a mode of 2.00. Theme 3 SQs yielded higher levels of agreement as evidenced by SQ6, SQ9, SQ10, SQ11, SQ12 and SQ16 all having modes of 4.00 (Agree).

Table 53

RQ3, Theme 3: Administrators (SQ6, SQ9, SQ10, SQ11, SQ12, SQ13, SQ16, SQ19, SQ20)

	SD	D	N	A	SA	Mean	Mode
6. The teacher evaluation process includes a discussion on student standardized test results for students.	0	5	4	8	2	3.37	4.00
9. Traditional teacher evaluation process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	0	2	4	12	2	3.70	4.00
10. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	0	4	5	10	1	3.40	4.00
11. Student evaluation is an effective tool that can be used to measure classroom performance.	0	0	5	13	2	3.85	4.00
12. Peer teacher evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	12	2	3.60	4.00
13. Parent evaluation is an effective tool that can be used to measure classroom performance.	1	8	8	3	0	2.65	2.00
16. Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	0	2	1	11	6	4.05	4.00
19. Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	1	10	3	6	0	2.70	2.00
20. Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	2	11	2	5	0	2.50	2.00

Note. SD=Strongly Disagree (1), D=Disagree (2), N=Neutral (3), A=Agree (4), SA=Strongly Agree (5).

Qualitative RQ3, theme 3 findings: Administrators. There were three interview questions that aligned to Theme 3 (IQ4, IQ8, IQ9). Each question was analyzed in order to

identify themes that support or do not support the quantitative findings for the administrative group. There were five total administrators interviewed for qualitative purposes in the study. The interview responses from the five administrators were used to identify themes for the overall group.

IQ4. As found in Table 54, two common themes surfaced for IQ4 (Do you believe that your schooling organization's teacher evaluation process results in an accurate measure of a teacher's ability to teach? Why or why not?) related to Theme 3: "Yes" was the most common followed closely by "No". The two themes that emerged from IQ4 were quite different. The first noted that the evaluation process in District A is a well-rounded instrument based in research that works well in measuring effectiveness. For example, HP1 stated, "I think our evaluation process is effective. It is based on the Charlotte Danielson model and it's specific enough for us to measure lots of areas of teacher performance." Additionally, DA noted, "I do think overall, our system is good in a teacher evaluation." AP1 described thoughts on the system in this way "I do think, for the most part, our system is very accurate." Interestingly, there was one administrator (HP2) that explained "I feel like often it's a snapshot. I often say observations are really just snapshots of one day, but I don't think it's a well-rounded picture of how well a teacher teaches." Four out of five administrators reported in a positive tone regarding District A's evaluation instrument. Qualitative data aligns well to survey responses for Theme 3 as SQs dealing with evaluation processes that were similar to District A's system reflected agreement towards use as a tool to measure performance.

Table 54

Schooling Organization's Teacher Evaluation Process Results as Accurate Measure of Teachers' Ability to Teach

Themes
Yes
Well-rounded tool to measure performance
No
The instrument is not well-rounded enough to measure performance

IQ8. As indicated in Table 55, IQ8 (Do student's standardized testing results serve as a tool that can influence teacher performance in the classroom? How so?), one common theme arose: "Yes", the results can influence classroom performance. Administrator interviews provided a clear level of agreement for the idea that standardized test results can impact classroom performance. For example, HP1 stated "I think you can look at it and maybe look to see if there's certain standards that you might not have been strong enough on." Additionally, AP2 explained "But I do think that teachers, good teachers, take pride in performance of their students. They can use it as a guide." The survey responses for Theme 3 reflected a consistent level of agreement that aligned with IQ8.

Table 55

Standardized Testing Results Serves as a Tool that can Influence Teacher Performance

Theme
Yes
Results can inform a teacher on strengths/weaknesses.

IQ9. When reviewing the data in Table 56 related to IQ9 (Do student's standardized testing results serve as a tool that can influence a teacher's professional growth? How so?) two themes were discussed: the first was "Yes", standardized test results can impact professional growth, and the second was "No", standardized test results do not influence professional growth. For example, DA noted, "I can use it to look at what the difference between the two teaching styles is, and help the teacher that's may not performing as well, improve their practice." Additionally, HP1 explained "You could use it to guide professional growth or offer suggestions of areas to improve." AP1 explained "I think by processing with other teachers and gaining different types of strategies to use in their classroom with standards that need remediation, I think that can definitely influence their professional growth." There were two administrators who responded no to IQ9. For example, AP2 stated "There's a lot to standardized tests, a lot out there…negatives. So I think teachers still look at it as teaching to the test." HP2 noted that "It just doesn't give us enough information for them to actually figure out what they would need to work on." Both responses make it more difficult to connect any one theme to survey responses due to the close mix of yes and no responses.

Table 56

Standardized Testing Results Serves as a Tool that can Influence Professional Growth

Themes
Yes
A provide info to guide professional growth
No
Connection to standardized test

Qualitative summary RQ3, theme 3: Administrators. Theme 3 findings identify a number of ideas that support quantitative data obtained in SQs. Theme 3 IQs revealed that teachers view the use of standardized tests results as tool that can inform instruction and even guide professional growth; however, there were also issues that emerged with the timeliness of receiving results and the lack of impact on current students. Additionally, the administrator group provided evidence through interview responses that District A' evaluation instrument was viewed as a good tool to measure teacher performance. The SQs for Theme 3 also indicated that standardized test results have a purpose, but only when considered as a tool used in an evaluation instrument.

Overall Summary for RQ3

An overall analysis of mean scores for survey responses of the administrators was conducted for the purpose of describing attitudes toward the use of standardized testing results as a measure of teacher performance on teacher evaluations. For example, Theme 1 responses for questions SQ7, SQ14 and SQ15 all produced similar mean scores. Additionally, each of these questions fell within the Disagree category on the survey. SQ7 (Student standardized test scores influence a teacher's future teaching performance.), SQ14 (Standardized tests administered to students is an effective tool that can be used to measure classroom performance.), and SQ15 (Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.) primarily focused on student standardized tests results as a tool to measure classroom performance.

Theme 2 mean scores included nine total SQs and five were in the Neutral range (SQ21, SQ22, SQ23, SQ24, and SQ26). SQ25 was in the Agree range and the remaining survey questions fell into a category that represented disagreement. Theme 3 mean scores were far less

scattered in terms of agreement/disagreement. SQ6, SQ9, SQ10, SQ11 and SQ12 all produced Neutral mean scores and involved traditional evaluation practices as a tool to measure teacher performance: SQ6 and SQ11resulted in mean scores that reflects neutrality with administrators. Both questions emphasized student standardized test results as a part of the teacher evaluation process. SQ19 and SQ20 reflected disagreement for administrators. Both questions considered student standardized tests results as an objective of the teacher evaluation process.

The consistency of responses among the administrators when considering the use of standardized tests to evaluate teachers was evident. Administrator interview responses were indicative of the standardized test results were not something they agreed with, which aligns with mean survey responses. During the interview process it became apparent that the teachers were not opposed to the use of standardized tests to measure performance, but consistently expressed that standardized tests lacked the ability to assess teacher effectiveness in an evaluation. Current use of District A's evaluation tool was mostly seen as a positive resource. A common theme emerged through the interviews, that numerous variables can impact a student's results and performance.

Research Question 4 Findings

Is there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?

$H_0$4. There is no statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

H4. There is a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results serve as an effective measure of their instructional performance in the classroom.

The Mann-Whitney U Test was used to compare administrators and teachers regarding their agreement or disagreement on the 28 SQs in the present study. The Mann-Whitney U was used because there were only two group to compare. Often times, administrators have different attitudes and opinions about the importance of standardized tests regarding the evaluation of teachers. The possibility of different attitudes between teachers and administrators served as a rationale for investigation into this particular research question.

As shown in Table 57, the results of the Mann-Whitney U Test determined if there was a statistically significant difference between the two groups or if the null hypothesis ($H_0$4) would be retained. If the null hypothesis ($H_0$4) was retained, then there was not enough evidence to reject the null hypothesis suggesting that there was not a large enough statistical difference for that particular survey item. In order to make the determination of whether or not to retain the null hypothesis, an level of .05 was used in the study. Probability levels that were greater than the alpha level of .05 suggested that there was not enough evidence to support a significant difference between the two groups for each survey item. After review of Mann-Whitney U Test results for each of the 28 SQs, it was determined that a number of survey items were not statistically significant using an alpha of .05 as compared to probability levels. SQ2, SQ7, SQ9, SQ10, SQ12, SQ13, SQ14, SQ15, SQ17, SQ18, SQ21, SQ22, SQ23, SQ25, and SQ26 were not statistically significant at the .05 alpha level. This is considered a random sampling or measurement error as there is not the presence of a statistically significant probability level that is less than the alpha level. Mann-Whitney U Test results are presented in Table 57, but

questions where there was not enough evidence to suggest a statistically significant difference on response results will not be discussed in this section.

SQ1 (The use of standardized test results is an effective tool for measuring teacher performance.) presents a significant difference that existed among the two groups. It was compared using the μ statistic that was equal to 264.00 with a probability of .001. The median (2.50) for the administrators appeared to be higher than the teachers (2.00). The median of 2.5 for the administrators suggests that the administrators believed that there is slightly more agreement when considering the use of standardized test results as an effective tool for measuring teacher performance than the teachers. However, both of the medians represent a score on the survey that represents a disagreement. A number of other questions followed a similar pattern in that there was a statistically significant difference due to the probability level value being less than the .05 alpha, yet median responses still suggested that both groups demonstrated some level of disagreement. SQ1, SQ3, SQ4, SQ5, SQ8 and SQ28 were the questions that illustrated this scenario. SQ3 (Student standardized test scores should be a component of the teacher evaluation process.) produced a $\mu = 252.50$, $p = 0.00$. The administrator median was 2.50 and the teacher median was 2.00; both representing disagreement. SQ4 (Student standardized test scores are accurate in assessment of teacher performance.) yielded $\mu = 233.50$, $p = .000$. Administrator median (2.00) and teacher (1.00), although different, represented varying degrees of disagreement. SQ5 (Student standardized test scores reflect a teacher's knowledge of teaching practices.) had a $\mu = 259.00$ and $p = .000$. Both teacher median and administrator median were identical to SQ4. SQ8 (Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.) produced a $\mu = 310.50$, $p = .006$. The median for administrators was 2.50 and for teachers was 2.00. SQ28 (I

am confident that student's standardized test results accurately measure teaching effectiveness.) yielded a $\mu = 310.00$, $p = .004$ with an administrative median of 2.00 and teacher median of 1.00. It is noteworthy to mention that of these questions all were considered to be Theme 1 questions that explicitly considered the concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness.

A number of other questions posed varying scenarios that deserve further analysis as a result of the Mann-Whitney U Test results. For example, SQ6 (The teacher evaluation process includes a discussion on student standardized test results for students.) and SQ11 (Student evaluation is an effective tool that can be used to measure classroom performance.) were identified as having a statistically significant difference based upon probability levels and .05 alpha level values. SQ6 had a $\mu = 317.00$ and probability level = .027. The administrator median (4.00) and teacher median (3.00) suggested that the administrative group Agreed with this statement while the teacher group appeared to take a Neutral stance. Often times, administrators are more familiar with all components of the teacher evaluation system and the question can be asked about whether or not teachers have as much familiarity with the District A evaluation instrument, thus the more Neutral response. SQ11 followed a similar pattern with $\mu = 337.00$, $p = .013$ and similar medians for administrators (4.00) and teachers (3.00). Question 11 (Student evaluation is an effective tool that can be used to measure classroom performance.) presents an interesting point of discussion in that the practice of utilizing student evaluation as part of teacher evaluation and performance decisions is not a regular practice in District A, yet conversations at the high school administrative level have encouraged principals to investigate avenues to increase student voice and the question of administrator agreement with this statement could stem from that sphere of influence.

There were three questions that presented a unique profile after analysis of the Mann-Whitney U Test results. SQ16 (Professional teaching portfolios [collection of reflections, critiques, lesson plans, samples of student work] is an effective tool that can be used to measure classroom performance.) produced a $\mu = 359.00$ and $p = .030$, yet medians for administrators and teachers were 4.00 (Agree). Additionally, SQ19 (Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.) resulted in a $\mu = 318.50$ and $p = .008$, yet still had identical medians at 2.00. Similarly, SQ20 (Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.) results were $\mu = 344.00$ and $p = .026$ with medians of 2.00 (disagree).

While SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area.) and SQ27 (Standardized testing is helping schools improve.) also demonstrated a statistically significant difference between responses of the two groups, the questions followed a different pattern that also merits further discussion. SQ24 had a $\mu = 266.50$ and $p = .001$ with administrative median of 3.00 (Neutral) and teacher median of 2.00 (Disagree). The implication is that one group (administrators) suggested a more Neutral response while the teacher group presented disagreement. The same pattern is demonstrated on SQ27 with a $\mu = 296.00$, $p = .003$ and the same median values for administrator and teacher. SQ24 stated "Standardized tests give me important feedback about how well I am teaching in each curricular area" and SQ27 stated "Standardized testing is helping schools improve." SQ24 is interesting in that teachers are directly connected to the curriculum and a defined group of students where administrators take a more global approach to curriculum through classroom observations. The possible difference could be illustrative of this phenomenon.

Overall, across the 28 question survey, 46.42% of the questions displayed a statistically significant difference between administrator and teacher responses. Additionally, 53.57% of the survey questions represented retention of the null hypothesis ($H_0$4) in that there was not enough evidence to suggest a statistically significant difference on response results. Table 57 presents the Mann-Whitney U test statics, *p* values, and medians for administrators and teachers for part two survey questions.

Table 57

Mann-Whitney U Test of Administrator and Teacher Survey Responses

SQ	μ Statistic	P value Asymp. Sig. (2-tailed)	Median	
			ADMIN	TEACHER
SQ1	264.00	.001	2.50	2.00
SQ2	421.00	.192	2.00	2.00
SQ3	252.500	.000	2.50	2.00
SQ4	233.50	.000	2.00	1.00
SQ5	259.00	.000	2.00	1.00
SQ6	317.00	.027	4.00	3.00
SQ7	443.50	.377	3.00	2.00
SQ8	310.50	.006	2.50	2.00
SQ9	505.00	.836	4.00	4.00
SQ10	403.00	.143	4.00	4.00
SQ11	337.00	.013	4.00	3.00
SQ12	456.00	.435	4.00	4.00
SQ13	467.00	.483	3.00	2.00
SQ14	383.00	.073	3.00	2.00
SQ15	406.50	.136	2.00	2.00
SQ16	359.00	.030	4.00	4.00
SQ17	442.00	.275	2.00	1.00
SQ18	462.00	.520	2.00	3.00
SQ19	318.50	.008	2.00	2.00
SQ20	344.00	.026	2.00	2.00
SQ21	409.00	.148	3.00	3.00
SQ22	400.00	.121	3.00	2.50
SQ23	373.50	.054	4.00	3.00
SQ24	266.50	.001	3.00	2.00
SQ25	495.00	.735	4.00	4.00
SQ26	404.50	.091	4.00	4.00
SQ27	296.00	.003	3.00	2.00
SQ28	310.00	.004	2.00	1.00

Summary of Findings

Chapter four addresses specific findings with regard to research questions one through four as identified in this study. Survey responses of teachers and administrators were presented

in the form of descriptive demographic data in order to describe each of these groups. Additionally, Likert-type scales where mean scores and modes were calculated in order to apply measures of central tendency for response analysis took place in order to suggest levels of agreement/disagreement for both groups. Survey questions addressed specific themes such as the use of standardized test results to evaluate teacher performance; attitudes towards teacher evaluation processes that include traditional methods and standardized test results; and trust levels in terms of the validity of standardized test results. Interviews were conducted with both administrators and teachers in order to add further depth of understanding to quantitative survey results. Interviews assisted in adding more information about why teachers and administrators demonstrated consistent disagreement with the idea of using standardized test results to evaluate teacher performance.

When considering the attitudes of high school mathematics teachers and administrators towards the use of standardized test results as tool to evaluate performance, it was clear that each group demonstrated disagreement regarding this idea. Additionally, when applying statistical analysis to the survey questions, the data suggest that responses on approximately 71% of survey questions yielded enough evidence to retain the null hypothesis (H_02) for research question two, reflecting similar attitudes of disagreement across the three groups of math teachers in response to survey questions.

An additional piece of this study looked at administrator and teacher attitudes in comparison to one another. Both groups took the same survey and statistical analysis was applied for the purpose of comparing the two groups. Findings suggest that 46.42% of the questions displayed a statistically significant difference between the administrator and teacher groups. Further, 53.57% of the survey questions represented retention of the null hypothesis

($H_0$4) for RQ4 in that there was not enough evidence to suggest a statistically significant difference on response results.

Table 58

Summary Findings of Survey Questions and Interviews for each Research Question

Research Question	Survey Questions	Interviews
1. What were the attitudes of high school math teachers regarding how student standardized test results served as an effective measure of their instructional performance in the classroom? a. …of teachers instructing accelerated math courses… b. …of teachers instructing non-accelerated… c. of teachers instructing both accelerated and non-accelerated…	Accelerated (Group A) • T1 – Disagreement • T2 – Disagreement • T3 – Agree/Neutral (traditional forms of evaluation); Disagree – standardized tests as objective of evaluation Non-Accelerated (Group B) • T1 – Disagreement T2 – Neutral (standardized test results as a tool to guide instruction); Disagree as a tool for evaluation of performance • T3 - Agree/Neutral (traditional forms of evaluation); Disagree – standardized tests as objective of evaluation Non-Accel/Accel (Group C) • T1 – Disagreement • T2 – Disagreement • T3 – Agree/Neutral (traditional forms of evaluation); Disagree – standardized tests as objective of evaluation	• Purpose of evaluation – Teacher accountability, monitor student growth, link teacher and student performance • Student accountability is lacking so results are not trustworthy • Outside Factors/Variables Impact Student Performance • Administrators need more time evaluate teachers • Standardized test results serve as a tool to guide instruction • AzMerit results arrive too late • AzMerit results are not specific • Inter-rater reliability negatively impacts evaluations • Pre/Post-tests are more effective measures of performance

Table 58 (continued)

Research Question	Survey Questions	Interviews
2. Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated math courses, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?	• $H_0$2 retained on 71% of SQ's (20 items) that there is no statistically significant difference among the attitudes of math teachers • $H_0$2 was not retained on SQ1, SQ2, SQ3, SQ6, SQ8, SQ19, SQ20, and SQ27.	N/A
3. Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated math courses, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?	• T1 – Disagree (All administrator mean fell in category of Disagree) • T2 – Neutral (standardized test results as a tool to guide instruction); Disagree as a tool for evaluation of performance • T3 – Neutral (traditional forms of evaluation); Disagree – standardized tests as objective of evaluation	• Purpose of evaluation – Teacher accountability, monitor student growth, link teacher performance to instruction of state standards • Outside Factors/Variables Impact Student Performance • Results are a snapshot only • Pre/Post-tests are more effective measures of performance • Student accountability is lacking so results are not trustworthy • AzMerit results are not specific • Standardized test results serve as a tool to guide instruction/professional growth

Table 58 (continued)

Research Question	Survey Questions	Interviews
4. Was there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?	• $H_0$4 retained for SQ2, SQ7, SQ9, SQ10, SQ12, SQ13, SQ14, SQ15, SQ17, SQ18, SQ21, SQ22, SQ23, SQ25, and SQ26 (53.57%) • $H_0$4 not retained for SQ1, SQ3, SQ4, SQ5, SQ6, SQ8, SQ11, SQ16, SQ19, SQ20, SQ24, SQ27, SQ28 (46.43%)	N/A

Summary

It was clear that each group demonstrated disagreement regarding the idea of using standardized test results as tool to evaluate teacher performance. Statistical analysis of survey questions suggests that responses on approximately 71% of items produced evidence that suggests retention of the the null hypothesis ($H_0$2) for research question two. Similar attitudes of disagreement, across the three groups of math teachers in response to survey questions, was a clear finding in the study.

A consideration for interview questions emerged in that there were a high number of neutral responses across groups in survey responses. Therefore, refining interview questions in order to uncover information from neutral responses would prove helpful in adding more descriptive, qualitative data to the study. Qualitative interview questions did, however, provide more detail about the nature of disagreement towards the use of standardized test results in performance evaluations. Themes identified from interview responses highlighted that there was disagreement towards the use of standardized test results in performance evaluations. Ideas that

emerged from interviews were lack of student accountability to take the exam seriously, student variables that can impact performance, student growth as a more viable approach to measure teacher performance, and lack of detail in AzMerit results.

The study focused on description of administrator and teacher attitudes in comparison to one another. Each group responded to the same survey items and statistical analysis was applied for the purpose of comparing the two groups. Findings indicated that 46.42% of the questions displayed a statistically significant difference between the administrator and teacher groups. Further, 53.57% of the survey questions represented retention of the null hypothesis ($H_0 4$) for RQ4 in that there was not enough evidence to suggest a statistically significant difference on response results.

Chapter 5

Summary, Conclusions, Implications, and Recommendations

Introduction

Chapter 5 presents a summary of the study as well as conclusions developed from the findings presented in Chapter 4. This chapter provides a discussion of the implications for practice in the educational world and recommendations for future studies. Concluding remarks about the study and a final statement are also included in this chapter.

Summary of Study

The concept of accountability has increasingly become an important aspect within the educational world. One specific area of education where there has been increased accountability is that of teacher evaluation and standardized testing. The focus has been on developing a means to measure individual teacher performance and determine effectiveness levels using standardized testing results. Value-added measures and student growth percentiles have been developed in order to attempt to use standardized test results as a means to evaluate the performance of teachers. While both of these approaches have been examined in order to determine if use is appropriate, a great deal of research points to the idea that they are unstable and inconsistent in measuring teacher performance. The inability of these tools to consistently measure teacher performance creates a significant concern in that increased accountability measures can impact educators in more profound ways. For example, high stakes decisions about compensation, tenure and dismissal could be influenced by tools that may not be stable enough to appropriately measure teacher performance. Additionally, legislative efforts to mandate the use of student achievement data on teacher evaluations have emerged across the United States, significantly impacting decisions about teacher performance through the evaluation process. Therefore, there

is great value in examining this concept further through the eyes of professionals in the field of education.

This study took place in a large, suburban K-12 school district with 30 elementary schools, seven middle schools, and five comprehensive high schools. The participants in this study included two groups within the district, teachers and administrators. The first group included three groups of high school math teachers. More specifically, all teachers in the district who instructed accelerated math courses (Group A), non-accelerated math courses (Group B) or those who taught both accelerated and non-accelerated math courses (Group C). Group A teaching assignments included only honors or AP (advanced placement) courses in the teacher schedule. The Group B non-accelerated teachers were responsible for teaching only classes that were considered survey courses and were not advanced in any way. The Group C accelerated/non-accelerated group consisted of teachers where both AP/Honors courses and non-accelerated courses were part of the teaching assignment. The second group was comprised of high school administrators who served at the district level and high school level. The administrator category was further defined as individuals who evaluate and/or train high school mathematics teachers.

The purpose of this mixed-method study was to examine the attitudes of high school mathematics teachers and high school administrators when considering the use of standardized test results as a measure of teacher performance in teacher evaluations. The researcher gathered both quantitative and qualitative data that served to describe the attitudes of both teachers and administrators. Participants were given surveys for the purpose of collecting quantitative data: he survey instrument consisted of two parts. Part one gathered demographic data about participant gender, education level, and experience. Part two included Likert-scaled survey

questions that produced responses varying from strong disagreement (1.00) to strong agreement (5.00). Additionally, survey items were categorized into three themes around the use of standardized test results and teacher evaluation. As a result of surveying participants, willing individuals were identified in order to participate in qualitative interviews that provided a deeper understanding of the attitudes for both teachers and administrators. Interview questions were also categorized into the same three themes as the survey questions.

Survey data were reported for each group of teachers (Group A, Group B, and Group C) in narrative form and tables. Data included the number of Likert-scaled responses for each individual in the group as well as the mean and mode for each survey question. Written narratives and tables were developed from interview questions for Group A, Group B, and Group C teachers. Interview transcriptions were analyzed for the purpose of identifying common ideas/themes that were provided during individual interviews. The same process was repeated in order to describe the attitudes of the administrative group. Survey questions were then statistically analyzed using the Kruskal-Wallis Test for the purpose of comparing responses across the three groups of teachers. Additionally, the teacher and administrator responses were compared for each survey question using the Mann-Whitney U Test in order to provide further statistical analysis.

Both quantitative and qualitative data emerged as a result of the study. Findings were presented in both table and narrative form in chapter four.

RQ1a (What are the attitudes of teachers instructing accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) findings suggested that the accelerated group of teachers (Group A) produced consistent disagreement with Theme 1 mean scores for survey items. Theme 1

questions primarily considered the concept of standardized test results as an indicator used to measure teacher performance and/or effectiveness. Theme 2 (use of standardized test results on evaluations) reflected consistent disagreement among the accelerated teachers. There were two Theme 2 questions, SQ25 and SQ26, which had a number of neutral responses. Additionally, there were also a number of agree responses. The questions dealt with testing causing tension and teacher expectations for students on standardized tests. Theme three (evaluation processes using standardized test results) produced more neutral responses toward forms of teacher evaluations that were included in the district's instrument such as pre-observation conferences, classroom observations, and post-conferences; however, consistent disagreement surfaced when asked if results on standardized tests should be an objective of an evaluation process.

RQ1b (What are the attitudes of teachers instructing non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) findings were similar to Group A in that all Theme 1 question mean scores reflected disagreement. Theme 2 data suggested that non-accelerated teachers took a more neutral stance when considering results on standardized tests impacting professional learning, clarifying learning goals, and teacher impact on student performance. However, responses were more consistently disagreeable when specifically identifying trusting standardized testing results as a part of an evaluation process. Theme 3 (evaluation processes using standardized test results) were similar in that more neutral responses were provided by Group B teachers when considering the effectiveness of familiar evaluation tools like teaching portfolios, classroom observations, and written evaluations by administrators. When asked about standardized tests as an objective of an evaluation process, Group B responses also reflected disagreement.

RQ1c (What are the attitudes of teachers that instruct both accelerated and non-accelerated math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) involved the Group C accelerated/non-accelerated teachers and produced survey responses that indicated disagreement toward Theme 1 (standardized test results as an indicator used to measure teacher performance and/or effectiveness) questions, similar to that of the other two teacher groups. Group C produced a similar profile to Group A with Theme 2 questions in that all means showed varying degrees of disagreement, apart from SQ25 (Testing creates a lot of tension for teachers and/or students) and SQ26 (I expect my students to perform well on tests). The survey response profile followed the same pattern as Group A which reflected agreement. SQ26 reflected more Neutral attitudes. Theme 3 findings for Group C reflected similar patterns of disagreement, neutrality, and agreement as the other two groups. However, there were two questions where Group C expressed disagreement when the other two groups were neutral. The questions considered tools used in an evaluation to assess teacher performance such as peer evaluation and student data.

Qualitative interviews produced a deeper understanding of teacher attitudes through analysis of question responses. For example, common ideas emerged from each of the three teacher groups as a result of qualitative interviews. Each group indicated an understanding that the purpose of using standardized test results was for teacher accountability and monitoring of student growth. Additionally, the idea that standardized test results are not a viable tool to measure teacher performance was evident as well. This feeling was attributed to the fact that there are many factors that impact student performance. Further, a lack of trust in the results was presented through interview responses because AzMerit results were reported as providing non-specific data that was not helpful in identifying learning gaps for students and not received in a

timely manner. Additionally, the idea that there is not an accountability tool for students on the AzMerit, such as attributing a grade for the test or a graduation requirement, was problematic for teachers in considering the results as a valid measure. The use of pre-/post-testing as a tool to measure student growth was considered to be a more valid means for measuring performance. Group B did respond more favorably toward the idea that standardized test results can be used to guide classroom performance, professional growth, and instruction; however, the other two groups expressed more disagreement towards the same ideas.

RQ2 (Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) served to analyze each survey item using the Kruskal-Wallis Test in order to compare responses across the three groups of teachers. Results indicated that 71% of the survey questions or 20 items retained the $H_0 2$ in that there was not a statistically significant difference across the three groups. $H_0 2$ was not retained for eight survey questions.

RQ3 (What are the attitudes of high school administrators regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) findings suggested that the administrative group produced consistent disagreement toward Theme 1 (use of standardized test results on evaluations) survey responses. Theme 2 (level of trust in the results on standardized tests) produced more neutral responses on the part of the administrative group. Theme 3 (evaluation processes using standardized test results) administrator responses were predominantly neutral when considering traditional forms of evaluation such as classroom observation, peer observation, and self-evaluation. However, when

considering survey items that focused on the use of standardized test results as an objective of teacher evaluation, consistent disagreement was expressed by the administrator group.

A number of common ideas emerged from interviews with administrators. For example, administrators indicated a clear understanding that the use of standardized test results on evaluations was meant as a tool to measure student growth, monitor teacher instruction as related to coverage of state standards, and as a teacher accountability tool. Additionally, similar to that of the teachers, administrators also reported that trust toward standardized test results was not high due to variables outside of the school that impacted student performance as well as the idea that results are only a snapshot of student performance. Administrators also expressed concerns with timeliness of receiving AzMerit results and lack of specificity with results. The group of administrators also communicated a concern with a lack of accountability for students to elevate the importance of AzMerit; however, the administrator group did express a more favorable attitude when considering the use of standardized test results as a mechanism to guide classroom instruction, identify strength/weaknesses, and impact professional growth of teachers.

RQ4 (Is there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) findings were based on results of the Mann-Whitney U Test in order to compare administrator and teacher responses on each of the survey questions. Some 15 questions produced a probability level where $H_0 4$ was retained and there was not a statistically significant difference between survey responses of the two groups. There were 13 questions where $H_0 4$ was not retained and a difference did exist. Of the 13 questions, there were two questions that produced a statistically significant difference where median responses reflected different levels of agreement/disagreement. For example, SQ6 (The teacher

evaluation process includes a discussion on student standardized test results for students) and SQ11 (Teacher self-evaluation is an effective tool that can be used to measure classroom performance) both yielded agreement for administrators and neutrality for teachers. SQ24 (Standardized tests give me important feedback about how well I am teaching in each curricular area) and SQ27 (Standardized testing is helping schools improve) both produced neutral medians for administrators and disagreement for teachers.

Conclusions

The study produced data that served to describe threes group of mathematics teachers and a group of administrators when considering attitudes towards the use of student achievement data in a performance evaluation.

RQ1a-c (attitudes of accelerated, non-accelerated and accelerated/non-accelerated towards the use of student standardized test results as an effective measure of their instructional performance in the classroom) produced similar trends in that teachers disagreed with the idea of using standardized test results as a tool in a performance evaluation. Teachers communicated the idea that standardized test scores do serve a purpose, but should not be included in performance evaluations. Teachers provided insight into the idea that standardized tests, along with a variety of others tool, can impact performance; however, there are multiple factors that negatively impact the potential for results to be consistently valid such as outside factors that impact student performance on the AzMerit test, lack of detail in test results, and timeliness of receiving results. Additionally, the methods teachers preferred for using data to measure performance focused on the pre-/post-testing in order to monitor student growth. Forms of evaluation such as classroom visits, written observations, and portfolios were seen as better indicators of teacher effectiveness.

RQ2 (Is there a statistically significant difference among the attitudes of math teachers that instruct accelerated math courses, non-accelerated or both types of math courses regarding how student standardized test results served as an effective measure of their instructional performance in the classroom?) results allow for the conclusion that almost three-quarters of the survey questions, specifically those that asked questions about the use of standardized tests as a part of teacher evaluation, were not significantly different across the three groups of teachers. Therefore, the result of analysis of survey questions better describe the attitude of disagreement toward use of standardized tests on teacher evaluations in that they were similar across the three groups of math teachers.

RQ3 (What were the attitudes of high school administrators regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?) presents a conclusion that was similar to that of the three teacher groups. The administrator group, through surveys and interviews, communicated an attitude of disagreement toward the use of standardized tests as a measure of teacher performance. Administrator interviews provided more information in that standardized test scores produced results that served only as an isolated measure that was impacted by a number of factors. For example, the idea that students were not held accountable in some way for results caused administrators to question the validity of standardized test results. Additionally, the idea that results did not provide enough specificity about student learning created a lack of trust in the results on the part of administrators.

RQ4 (Was there a statistically significant difference between teachers' and administrators' attitudes regarding how student standardized test results served as an effective measure of math teacher instructional performance in the classroom?) allows the conclusion to

be drawn that while there were a number of responses from teachers and administrators that differed, survey questions involving the specific idea of using standardized test results as a tool to measure teacher performance resulted in similar degrees of disagreement. Two questions produced a neutral median from administrators, while the same questions produced a median score that represented disagreement for teachers. The first question discussed standardized tests scores influencing teaching performance and the second question discussed standardized test scores as a tool to measure classroom performance; however, when that focus shifted to using standardized test scores as a tool on an evaluation instrument, it can be concluded that both groups disagreed with that idea.

Implications for Practice

Policy makers and administrators are encouraged to make note of the following implications for practice when considering the use of standardized testing results in teacher evaluations as well as when developing future legislation or policies that may involve the areas of performance evaluation of educators and student standardized test results.

1. Use caution when enacting or revising any legislation that impacts the educator evaluation instrument and requirements of using standardized test results as a component of performance measurement.
2. Take into consideration that educators are not apprehensive with regard to accountability, but believe that a system of pre-/post–testing for the purpose of measuring student growth is viewed as a more appropriate means to combine student data and teacher evaluation.

3. When evaluating teachers using standardized testing data, considerations must be made toward the educational level of the coursework being measured, understanding that differences may exist with student achievement levels.
4. Analyze further evaluation approaches using data as a tool to evaluate teacher performance apart from VAMs and SGPs.
5. Avoid using standardized testing results as a performance measurement tool that impacts high-stakes decisions involving compensation, tenure and dismissal.
6. District human resources departments must continue to examine effective methods for inter-rater reliability training for administrators on teacher evaluation instruments.
7. Examine ways to reduce turnaround time for when standardized test results are distributed to teachers.
8. Examine ways to break down standardized test result reports into more specific categories so as to better inform teachers about areas of strength and weakness regarding student performance.
9. Although legislation is already in place regarding the use of student achievement data, it is essential that legislators seek to identify a proven method for using standardized test results in teacher evaluations before requiring implementation.
10. Districts should consider policies that increase student accountability with regard to students taking standardized tests (i.e., taken for a grade, possible graduation requirement).
11. Districts should reference the current study or studies similar in nature prior to developing/revising teacher evaluation instruments.

12. Avoid implementation of compensation for educators based upon standardized test results without consideration of variables that impact student performance such as demographics and baseline data reflecting students' current academic performance/standing.

Recommendations for Future Studies

1. Replicate the study using revised interview questions that focus on clarification of neutral responses on survey questions. There were a number of neutral responses on survey questions. By doing so, additional information may be uncovered that further describes teacher attitudes. Gathering this data would also aid in understanding the necessary pieces of information that requires clarification for participants in order to elicit more thorough responses.
2. Replicate the study with a narrowed focus on specific survey questions that emphasize Theme 1 and increase the level of qualitative data through additional interviews and focus groups.
3. Replicate this study with a focus specifically on the accelerated subgroup of mathematics teachers, but expand the population to include the middle grades. Often times, there are mathematics teachers that deliver honors/advanced courses in mathematics where middle grade students can earn high school credit. Expanding the group to include grades seven and eight would provide a larger group of accelerated teachers in order to better describe attitudes.
4. Develop a future study that expands the population of mathematics teachers to include a larger group across multiple K-12 districts. A study of this nature could

provide a larger sample and warrant the addition of focus groups to gather more data regarding teacher attitudes with a larger population.

5. Develop a future study that targets different segments of the population of teachers which could add relevant insight into how attitudes toward standardized testing and teacher evaluation are similar or different across multiple subgroups. Ethnicity, gender, and experience levels of teachers could provide valuable data regarding how specific sub-groups of teachers view this topic.
6. Implement a future study that focuses on different teacher and administrator groups. While the data from the specific group of math teachers is valuable, it could prove meaningful that more information be gathered from a variety of other educational stakeholders. For example, high school teachers in the content areas of English Language Arts and Science also have a great deal of experience with standardized tests and certainly can provide valuable insight into teacher attitudes.
7. Implement a future study with elementary and middle school teachers. Teachers at these levels also have valuable experiences and perspectives with regard to state testing. Comparisons between these groups could serve to identify specific themes that exist across a wider population of teachers and administrators.
8. Implement a study that includes parents. Studying parent attitudes toward standardized testing would also be a valuable perspective that could help better understand parental attitudes toward the use of standardized tests to evaluate teacher performance.
9. Implement a future study that includes students. A common theme that emerged from qualitative interviews is that both teacher and administrator groups were concerned

about the lack of accountability that helped students to see the importance of testing. A study involving student attitudes could produce valuable data on what creates meaning for students when taking a standardized test.

10. A future study targeting legislators and members of the State Board of Education with regard to attitudes toward the use of standardized tests on performance evaluations could prove beneficial. Understanding legislative attitudes towards this topic could help to make comparisons to educator attitudes as well provide valuable information as different approaches are investigated for implementation.
11. A future study that takes yields a summative statistical figure for each theme would produced valuable data. For example, additional analysis of this nature allows for a summative figure to be produced for each of the three themes that were analyzed in this study using survey questions. Analysis of this nature creates the opportunity for application of parametric statistics and serves to add more information about teacher and administrator attitudes.

Concluding Remarks

This study began as the researcher's desire to better understand the impact of mandatory legislation towards the use of student achievement data and evolved into a formal study that was intended to add descriptive information for district administrators, site-based administrators, and teachers on the attitudes that an important group of educators had with regard to using standardized test results in a responsible, appropriate manner. The sincere hope is that the data presented in this study serve to influence future decisions when it comes to how best to manage accountability with the best interests of teachers in mind so that they can effectively meet the needs of students. The findings in this study suggest that teachers see a viable role for

standardized test results, which is one tool out of many that serves to provide insight into student learning. However, the results are not seen as an effective tool for measuring teacher performance within the context of an evaluation instrument that can impact pieces of professional standing such as tenure, dismissal, and compensation.

REFERENCES

Anderson, L. M., Butler, A., Palmiter, A., & Arcaira, E. (2016). *Study of emerging teacher evaluation systems* (pp. 1-147). United States Department of Education, Office of Planning, Evaluation, and Policy Development, Washington, D.C.

Arizona Department of Education (ADE). (2014). *The teacher evaluation process. An Arizona model for measuring educator effectiveness* (pp. 1-80). Retrieved from https://cms.azed.gov/home/GetDocumentFile?id=54b589481130c00dd469e8e1

Arizona Department of Education (ADE). (n.d.). *AzMerit Arizona's Statewide Achievement Assessment For English Language Arts and Mathematics* [Brochure]. Phoenix, AZ: Author. Retrieved from http://azmeritportal.org/core/fileparse.php/1972/urlt/AzMERIT-brochureHIGH-SCHOOL3.pdf.

ARS 15-203. (2012). *Arizona revised statute principal and teacher evaluation*, section (A)-38. Retrieved from https://www.azleg.gov/ars/15/00203.htm.

AzMerit Portal. (2017). *AzMerit Testing Conditions, Tools and Accommodations Guidance.* Retrieved from file:///Users/teacher/Desktop/Doctorate/Dissertation/GetDocumentFile.pdf.

Barrett, F. X., Burris, C., Cody, A., Koonlaba, A., Martinez, S. J., McKelvy, T., Meeks, J. L., Jr., & Nolan, L. (2016). *Teachers talk back: Educators on the impact of teacher evaluation* (pp. 1-21). The Network for Public Education. Retrieved from https://networkforpubliceducation.org/wp-content/uploads/2016/04/NPETeacherEvalReport.pdf

Barrett, J. (1986). *The evaluation of teachers. ERIC Digest.* Retrieved from ERIC database. (ED278657).

Bergin, C. (2015). Using student achievement data to evaluate teachers. *Network for Educator Effectiveness,* 1-6. Retrieved from https://nee.missouri.edu/documents/NEE_White_PaperStudent_Achievement_in_Teacher Evaluation2015_5_20.pdf.

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record The Voice of Scholarship in Education,116*(1), 1-21. Retrieved from http://www.tcrecord.org/content.asp?contentid=17293.

Blumberg, A. (1985). Where we came from: Notes on supervision in the 1840s. *Journal of Curriculum and Supervision, 1*(1), 56-65.

Center on Great Teachers and Leaders at the American Institute for Research (2013). *Databases on state teacher and principal evaluation policies*. Retrieved from http://resource.tqsource.org/stateevaldb/Compare50States.aspx.

Check, J., & Schutt, R. K. (2012). *Research methods in education*. Thousand Oaks, Calif: Sage Publications.

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher, 45*(6), 378-387.

Coleman, E. (1945). The "supervisory visit". *Educational Leadership: Supervision for Modern Schools, 2*(4), 164-167.

Corcoran, S. (2010). *Can teachers be evaluated by their students' test scores? Should they be?* Report for the education policy for action series. Providence, RI: Annenberg Institute for School Reform at Brown University. Retrieved from http://eric.ed.gov/?id=ED522164

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Thousand Oaks, CA: SAGE.

Daniel, J. (2012). *Sampling Essentials*. Thousand Oaks, CA: SAGE.

Danielson, C. (2016). It's time to rethink teacher evaluation. *Education Week, 35*(28), 20-24. Retrieved from https://www.edweek.org/ew/articles/2016/04/20/charlotte-danielson-on-rethinking-teacher-evaluation.html.

Danielson Group. (n.d.). *The framework.* Retrieved from http://www.danielsongroup.org/framework/.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2011). *Getting teacher evaluation right: A brief for policymakers*. Capitol Hill Research Briefing convened by the American Educational Research Association and the National Academy of Education. Washington, DC. Retrieved from https://files.eric.ed.gov/fulltext/ED533702.pdf.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research,53*(3), 285-328.

Dee, T., & Jacob, B. (2009). *The impact of no child left behind on student achievement.* National Bureau of Economic Research Working Paper No. 15531. Retrieved from http://www.nber.org/papers/w15531.

Dewey, J., & Simpson, D. J. (2010). *Teachers, leaders, and schools: Essays by John Dewey*. Carbondale, IL: Southern Illinois University Press.

Ellis, C. R. (2007). No child left behind-A critical analysis "A nation at greater risk." *Curriculum & Teaching Dialogue, 9*(1/2), 221-233.

Engberg, J., & Mihaly, K. (2012, September). Multiple choices options for measuring teaching effectiveness. Retrieved from https://www.rand.org/education/projects/measuring-teacher-effectiveness/multiple-choices.html.

Elementary and Secondary Act (ESEA). (1965). *The Elementary and Secondary Act of 1965.* Retrieved from http://www.socialwelfarehistory.com/programs/education/elementary-and-secondary-education-act-of-1965/.

Finnegan, R. (2013). Linking teacher self-efficacy to teacher evaluations. *Journal of Cross-Disciplinary Perspectives in Education, 6*(1), 18-25. Retrieved from http://jcpe.wmwikis.net/file/view/Finnegan_Linking_Efficacy_to_Evaluations.pdf.

Freedberg, L. (2012, April 2). Publishing teacher effectiveness rankings, pioneered in California, draws more criticism. *Huffpost*. Retrieved from https://www.huffingtonpost.com/2012/04/02/teacher-effectiveness-ran_n_1397536.html.

Gilles, J. F. (2015). *Interpreting teacher evaluation policies: The perspectives of local and state-level policy actors in two U.S. states* (Unpublished doctoral dissertation). University of Minnesota.

Glatthorn, A. A., & Holler, R. L. (1987). Differentiated teacher evaluation. *Educational Leadership*, *44*(7), 56.

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart and Winston, Inc.

Goslin, W. E. (1946). Know your teacher. *Educational Leadership: Teachers as Individuals,3*(6), 260-263.

Guthrie, J. W. (2003). *Encyclopedia of education* (2nd ed.). New York, NY: Gale, Cengage Learning.

Hankamp, G. (1946). Are teachers people? *Educational Leadership: Teachers as Individuals, 3*(6), 250.

Hull, J. (2013). *Trends in teacher evaluation: How states are measuring teacher performance* (pp. 1-40, Rep.). Alexandria, VA: Center for Public Education of the National School Boards Association. Retrieved from http://www.centerforpubliceducation.org/ Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/ Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf.

Hunter, M. (1976). Teacher competency: Problem, theory, and practice. *Theory Into Practice, 15*(2), 162-171.

Hunter, M. (1983). Script taping: An essential supervisory tool. *Educational Leadership*, *41*(3), 43.

Hunter, M. (1987). Beyond rereading Dewey... What's next? A response to Gibboney. *Educational Leadership, 44*(5), 51-53.

Hunter, M. (1990-1991). Lesson design helps achieve the goals of science instruction. *Educational Leadership, 48*(4), 79-81.

Jewell, J. W. (2017). From inspection, supervision, and observation to value-added evaluation: A brief history of U.S. teacher performance evaluations. *Drake Law Review, 65*, 363-419.

Jeynes, W. (2007). American educational history school, society, and the common good. Thousand Oaks, CA: SAGE.

Johnson, J. (2016). Calling on Charlotte Danielson to rethink her rethinking. *Chicago Union Teacher, 79*(8), 44-46. Retrieved from https://www.ctunet.com/media/chicago-union-teacher/downloadable-pdf/521081_CUT_June_reduced.pdf.

Lash, A., Makkonen, R., Tran, L., & Huang, M. (2016). *Analysis of the stability of teacher level growth scores from the student growth percentile model* (REL 2016–104). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/west/pdf/REL_2016104.pdf.

Marsh, C. J., & Willis, G. (2003). *Curriculum: Alternative approaches, ongoing issues* (3rd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice, 13*(6), 522-525.

Martin, W. E., & Bridgmon, K. D. (2012). *Quantitative and Statistical Research Methods* (1st ed.). San Francisco, CA: Jossey-Bass.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: supporting the art and science of teaching*. Alexandria, VA: ASCD.

McGuinn, P. (2014). Presidential policymaking: Race to the top, executive power, and the Obama education agenda. *The Forum, 12*(1), 61-79.

Melchior, W. T. (1950). *Instructional supervision: A guide to modern practice*. Boston, MA: Heath.

Monroe, S., & Cai, L. (2015). Examining the reliability of student growth percentiles using multidimensional IRT. *Educational Measurement: Issues and Practice, 34*(4), 21-30.

Neill, M. (2016). The testing resistance and reform movement. *The Monthly Review,* 8-28. doi:10.14452/MR-067-10-2016-03_2.

Norušis, M. J. (2004). *SPSS 13.0 advanced statistical procedure companion*. Upper Saddle River, NJ: Prentice Hall.

Orange, C. (2002). *The quick reference guide to educational innovations: Practices, programs, policies, and philosophies*. Thousand Oaks, CA: Corwin Press, Inc.

Pajak, E. (2001). Clinical supervision in a standards-based environment: Opportunities and challenges. *Journal of Teacher Education, 52*(3), 233-243.

PBS News Hour. (2013, February 4). *A brief overview of teacher evaluation controversies*. (2013, February 4). Retrieved from https://www.pbs.org/newshour/education/teacher-evaluation-controversies.

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston, MA: National Board on Educational Testing and Public Policy.

Plano Clark, V. L., & Ivankova, N. V. (2015). *Mixed methods research: A guide to the field*. Los Angeles, CA: SAGE.

Reavis, C. A. (1976). Clinical supervision: A timely approach. *Educational Leadership, 33*(5), 360-363.

Richards, L., & Morse, J. M. (2013). *README FIRST for a user's guide to qualitative methods* (3rd ed.). Thousand Oaks, CA: SAGE.

Roberts, C. M. (2010). *The dissertation journey: a practical and comprehensive guide to planning, writing, and defending your dissertation*. Thousand Oaks, CA: Corwin Press.

Slim, N. (2004). *The influence of governance structure on teacher evaluation practice* (Doctoral dissertation, University of Southern California) (pp. 1-169). Ann Arbor, MI: ProQuest Information and Learning Company.

Starratt, R. J. (n.d.). Supervision of instruction - The history of supervision, roles, and responsibilities of supervisors, issues, trends, and controversies. Retrieved from http://education.stateuniversity.com/pages/2472/Supervision-Instruction.html#ixzz2RZweUdFf.

Steinmayr, R., MeiBner, A., Weidinger, A. F., & Wirthwein, L. (2015, June 4). *Academic achievement*. Oxford Bibliographies. Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml.

Stiggins, R. J., & Duke, D. (1988). The case for commitment to teacher growth research on teacher evaluation. New York, NY: State University of New York Press.

The Glossary of Education Reform. (2015, November 12). *Standardized test*. Retrieved from http://edglossary.org/standardized-test/

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The New Era of Mixed Methods. *Journal of Mixed Methods Research, 1*(3), 4-7. doi:10.1177/2345678906293042.

Tracy, S. J. (1995). How historical concepts of supervision relate to supervisory practices today. *Clearing House*, *68*(5), 320.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our National failure to acknowledge and act on differences in teacher effectiveness* (pp. 31-35). Ann Arbor, MI: Prakken Publication.

Whitehead, M. (1952). Teachers look at supervision. *Educational Leadership, 10*(2), 1011-1106.

Wilson, L. O. (2017). Madeline Hunter lesson plan model or drill that skill – A model of repetition and direct instruction [Web log post]. Retrieved from https://thesecondprinciple.com/teaching-essentials/models-of-teaching/madeline-hunter-lesson-plan-model/.

Wise, A. E., Darling-Hammond, L., Tyson-Bernstein, H., & McLaughlin, M. W. (1984). Teacher evaluation: a study of effective practices. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/reports/R3139.html.

Xu, X., Grant, L., & Ward, T. (2016). Validation of a statewide teacher evaluation system. *NASSP Bulletin, 100*(4), 203-222.

Appendix A

Teacher and Administrator Surveys

Teacher Survey Questionnaire

Teacher Evaluation and Standardized Testing

Part I
The following information will be used to provide an accurate description of the population being surveyed.

A. I am	Male		Female
B. The highest degree I have earned is	Bachelor's	Master's	Doctorate

C. The total years of teaching experience	0	1-5	6-10	11-15	16-20	21-25	26-30	>30
D. The total years of teaching experience within current schooling organization	0	1-5	6-10	11-15	16-20	21-25	26-30	>30

E. I am currently teaching/working in the following role (check those that apply)	Accelerated Algebra 1-2 Algebra 1-2 Geometry 1-2 Honors Geometry 1-2 Algebra 3-4 Honors Algebra 3-4 Pre-Calculus 1-2 Honors Pre-Calculus 1-2 AP Calculus BC 1-2 AP Calculus AB 1-2 AP Statistics 1-2 Statistics and Probability 1-2
F. Would you be willing to participate in an individual interview for the purpose of answering additional questions on this topic?	Yes No If yes please provide contact information:

Part II

Please rate the following statements in terms of how well they describe your feelings, beliefs and opinions as they relate to the topics, issues and areas of your schooling organization. Your responses are strictly confidential. Rate each statement using the following scale:

1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

1. The use of standardized test results is an effective tool for measuring teacher performance.	1	2	3	4	5
2. I feel confident that the use of standardized test results can improve teacher performance in the district.	1	2	3	4	5
3. Student standardized test scores should be a component of the teacher evaluation process.	1	2	3	4	5
4. Student standardized test scores are accurate in assessment of teacher performance.	1	2	3	4	5
5. Student standardized test scores reflect a teacher's knowledge of teaching practices.	1	2	3	4	5
6. The teacher evaluation process includes a discussion on student standardized test results for students.	1	2	3	4	5
7. Student standardized test scores influence a teacher's future teaching performance.	1	2	3	4	5
8. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	1	2	3	4	5
9. Traditional Teacher Evaluation Process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
10. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
11. Student evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
12. Peer Teacher evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
13. Parent Evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
14. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	1	2	3	4	5

15. Students' performance on standardized tests is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
16. Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
17. Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	1	2	3	4	5
18. Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	1	2	3	4	5
19. Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	1	2	3	4	5
20. Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	1	2	3	4	5
21. Results on standardized tests identifies specific areas for professional learning.	1	2	3	4	5
22. Standardized tests help to clarify which learning goals are most important.	1	2	3	4	5
23. Teachers can influence substantially how well their students do on standardized tests.	1	2	3	4	5
24. Standardized tests give me important feedback about how well I am teaching in each curricular area.	1	2	3	4	5
25. Testing creates a lot of tension for teachers and/or students.	1	2	3	4	5
26. I expect my students to perform well on tests.	1	2	3	4	5
27. Standardized testing is helping schools improve.	1	2	3	4	5
28. I am confident that student's standardized test results accurately measure teaching effectiveness.	1	2	3	4	5

Administrator Survey Questionnaire

Teacher Evaluation and Standardized Testing

Part I

The following information will be used to provide an accurate description of the population being surveyed.

A. I am	Male		Female
B. The highest degree I have earned is	Bachelor's	Master's	Doctorate

C. The total years of administrative experience	0	1-5	6-10	11-15	16-20	21-25	26-30	>30
D. The total years of administrative experience within current schooling organization	0	1-5	6-10	11-15	16-20	21-25	26-30	>30

E. I am currently working in the following role (check those that apply)	Assistant Principal High School Principal High School District Level Administrator
F. Would you be willing to participate in an individual interview for the purpose of answering additional questions on this topic? G.	Yes No If yes please provide contact information:

Part II

Please rate the following statements in terms of how well they describe your feelings, beliefs and opinions as they relate to the topics, issues and areas of your schooling organization. Your responses are strictly confidential. Rate each statement using the following scale:

1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

29. The use of standardized test results is an effective tool for measuring teacher performance.	1	2	3	4	5
30. I feel confident that the use of standardized test results can improve teacher performance in the district.	1	2	3	4	5
31. Student standardized test scores should be a component of the teacher evaluation process.	1	2	3	4	5
32. Student standardized test scores are accurate in assessment of teacher performance.	1	2	3	4	5
33. Student standardized test scores reflect a teacher's knowledge of teaching practices.	1	2	3	4	5
34. The teacher evaluation process includes a discussion on student standardized test results for students.	1	2	3	4	5
35. Student standardized test scores influence a teacher's future teaching performance.	1	2	3	4	5
36. Student standardized test scores are a viable source of data that can be used to evaluate teacher performance.	1	2	3	4	5
37. Traditional teacher evaluation process (pre-observation conference, classroom observation, post-observation conference, written report completed by evaluator) is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
38. Teacher self-evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
39. Student evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
40. Peer teacher evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
41. Parent evaluation is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
42. Standardized tests administered to students is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
43. Students' performance on standardized tests is an effective tool that can be used to measure classroom	1	2	3	4	5

	performance.					
44.	Professional teaching portfolios (collection of reflections, critiques, lesson plans, samples of student work) is an effective tool that can be used to measure classroom performance.	1	2	3	4	5
45.	Teachers trust the use of student's performance on standardized tests as a part of the evaluation process.	1	2	3	4	5
46.	Administrators trust the use of student's performance on standardized tests as a part of the evaluation process.	1	2	3	4	5
47.	Students' results on standardized tests should be an objective of the evaluation process for continuing teachers.	1	2	3	4	5
48.	Students' results on standardized tests should be an objective of the evaluation process for non-continuing teachers.	1	2	3	4	5
49.	Results on standardized tests identifies specific areas for professional learning.	1	2	3	4	5
50.	Standardized tests help to clarify which learning goals are most important.	1	2	3	4	5
51.	Teachers can influence substantially how well their students do on standardized tests.	1	2	3	4	5
52.	Standardized tests give important feedback about how well a teacher is teaching in each curricular area.	1	2	3	4	5
53.	Testing creates a lot of tension for teachers and/or students.	1	2	3	4	5
54.	I expect my students to perform well on tests.	1	2	3	4	5
55.	Standardized testing is helping schools improve.	1	2	3	4	5
56.	I am confident that student's standardized test results accurately measure teaching effectiveness.	1	2	3	4	5

Appendix B

Interview Questions

1. What do you believe is the intended purpose of using student standardized test results as a component of the evaluation tool within your schooling organization?
2. Do you believe that your schooling organization's procedure for utilizing student achievement data as an indicator of performance supports the intended purpose of teacher evaluation? How so?
3. Do you believe that the use of student standardized testing results on a teacher evaluation instrument is a valid measure of teacher competency?
4. Do you believe that your schooling organization's teacher evaluation process results in an accurate measure of a teacher's ability to teach? Why or why not?
5. Describe what you consider to be an effective method of teacher evaluation using standardized testing results?
6. Do you believe that student standardized testing results serve as an indicator of teacher effectiveness? Why or why not?
7. Do you trust the results of student standardized tests as a measure of performance? Why or why not?
8. Do student's standardized testing results serve as a tool that can influence teacher performance in the classroom? How so?
9. Do student's standardized testing results serve as a tool that can influence a teacher's professional growth? How so?

Appendix C

NAU IRB Approval

NORTHERN ARIZONA UNIVERSITY | Office of Regulatory Compliance

Institutional Review Board
Human Research Subjects Protection Program

805 S Beaver St
Building 22, Room 215
PO Box: 4052
Flagstaff AZ 86011
928-523-9551
http://nau.edu/Research/Compliance/Human-Subjects/Welcome

To:	Chad Lanese
From:	NAU IRB Office
Approval Date:	February 5, 2018
Project:	STUDENT STANDARDIZED TEST SCORES AS AN EFFECTIVE MEASURE OF TEACHER PERFORMANCE: TEACHER AND ADMINISTRATOR ATTITUDES
Project Number:	1176778-1
Submission:	New Project
Review Level:	Exempt Review
Action:	EXEMPT
Project Status:	Exempt
Review Category/ies:	**Condition to approval:** Permission from District's Assistant Superintendent must be obtained and submitted to the NAU IRB office for acknowledgment prior to commencement of research. **Exempt Approval 45 CFR 46.101(b)(2):** Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior.

This submission meets the criteria for exemption under 45 CFR 46.101(b). This project has been reviewed and approved by an IRB Chair or designee.

- Northern Arizona University maintains a Federalwide Assurance with the Office for Human Research Protections (FWA #00000357).
- All research procedures should be conducted in full accordance with all applicable sections of the guidance.
- Exempt projects do not have a continuing review requirement.
- This project should be conducted in full accordance with all applicable sections of the guidance and you should notify the IRB immediately of any proposed changes that affect the protocol.
- Amendments to exempt projects that change the nature of the project should be submitted to the Human Research Subjects Protection Program (HRSPP) office for a new determination. See the guidance Exempt Research for more information on changes that affect the determination of exemption. Please contact the HRSPP to consult on whether the proposed changes need further review.
- You should report any unanticipated problems involving risks to the participants or others to the IRB.
- All documents referenced in this submission have been reviewed and approved. Documents are filed with the HRSPP Office. If subjects will be consented, the approved consent(s) are attached to the approval notification from the HRSPP Office.
- Exempt projects are maintained in HRSPP for five (5) years from approval. An updated application is required every five (5) years.

Appendix D

NAU Informed Consent

NORTHERN ARIZONA UNIVERSITY

Online Survey Consent

Project Number: 1176778-1
Approval Date: February 5, 2018
This stamp must be on all consenting documents

APPROVED

You are being invited to participate in a research study titled *Student Standardized Test Scores As An Effective Measure of Teacher Performance: Teacher and Administrator Attitudes*. This study is being done by Chad Lanese from Northern Arizona University.

The purpose of this research study is to describe the attitudes of high school mathematics teachers and administrators towards the use of student standardized test results on performance evaluations. If you agree to take part in this study, you will be asked to complete an online survey/questionnaire. This survey/questionnaire will ask about topics related to years of experience, current position, standardized tests and experiences with teacher evaluation. It will take you approximately 15 minutes to complete.

You may not directly benefit from this research; however, I hope that your participation in the study may assist in better defining educator attitudes towards how standardized test results may be used in performance evaluations.

I believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach of confidentiality is always possible. To the best of my ability your answers in this study will remain confidential. I will minimize any risks by limiting identifiable information and maintaining anonymity by not collecting email addresses, names or addresses. All data collected in the study will be maintained and disposed of by the researcher and utilized for the sole purpose of this study.

Your participation in this study is completely voluntary and you can withdraw at any time. You are free to skip any question that you choose. If you choose not to participate it not affect your relationship with Northern Arizona University or result in any other penalty or less of benefits to which you are otherwise entitled.

If you have questions about this project or if you have a research-related problem, you may contact the researcher, Chad Lanese at cjl34@nau.edu or 602.909.3702. If you have any questions concerning your rights as a research subject, you may contact Northern Arizona University IRB Office at irb@nau.edu or (928) 523-9551.

By submitting this survey, I affirm that I am over 18 years of age and agree that the information may be used in the research project described above. Your participation is most sincerely appreciated!

Consent Version: 02/01/2018

NORTHERN ARIZONA UNIVERSITY

Project Number: 1176778-1
Approval Date: February 5, 2018
This stamp must be on all consenting documents

Human Subject Informed Consent

Title of Study: Student Standardized Test Scores As An Effective Measure of Teacher Performance: Teacher and Administrator

Principal Investigator: Chad Lanese

This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate. Please consider the information carefully. Feel free to discuss the study with your friends and family and to ask questions before making your decision whether or not to participate.

Why is this study being done?

The purpose of this study is to examine both high school math teachers and administrator attitudes regarding whether or not student standardized test results are seen as effective measures of instructional performance. In order to collect more information on this topic it is necessary to conduct a research study.

How many subjects will participate and how long will the study take?

High School mathematics teachers and administrators serve as specific groups that will participate in this study. Sample size within the questionnaire survey will consist of approximately 35-40 high school mathematics teachers and 10-15 administrators. Interviewees will be comprised of approximately ten teachers and five administrator participants.

What will happen if I take part in this study?

If you choose to participate in this study you will be contacted by the investigator in order to arrange a convenient time to participate in a one on one interview. The interview questions will be provided in advance so as to provide any needed clarification so that questions can be answered during the interview. The interview will last approximately one hour and you will be asked if the interview can be recorded for the purpose of transcription and qualitative analysis.

Will there be any cost to you to take part in this study?

As a result of participation in this study there will be no financial costs to participants. However, when conducting one on one interviews a time commitment of approximately one hour will be necessary.

Will you be paid to take part in this study?

You will not be paid for your participation in this research study.

Can I stop being in the study?

Participation in this study is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled and the subject may discontinue

participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

Your participation is voluntary. You may refuse to participate in this study. If you decide to take part in the study, you may leave the study at any time. No matter what decision you make, there will be no penalty to you and you will not lose any of your usual benefits. Your decision will not affect your future relationship with Northern Arizona University. If you are a student or employee at Northern Arizona University, your decision will not affect your grades or employment status.

What are the risks and/or discomforts you might experience if you take part in this study?
No known risks are associated with this study.

Are there any benefits for you (or for others) if you choose to take part in this research study?
You may not directly benefit from this research; however, I hope that your participation in the study may assist in better defining educator attitudes towards how standardized test results may be used in performance evaluations.

What other choices do I have if I do not take part in the study?
You may choose not to participate in this study without penalty or loss of benefits to which you are otherwise entitled.

Will my study-related information be kept confidential?
Efforts will be made to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law.

Also, your records may be reviewed by the following groups:

- Office for Human Research Protections or other federal, state, or international regulatory agencies
- Northern Arizona University Institutional Review Board

Who can you call if you have any questions?
If you have any questions about taking part in this study or if you feel you may have suffered a research related injury, you can call the Principal Investigator at: 602-909-3702

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact the Human Subjects Research Protection Program at 928-523-9551 or online at http://nau.edu/Research/Compliance/Human-Research/Welcome/.

If you are injured as a result of participating in this study or for questions about a study-related injury, you may contact **Chad Lanese at cjl34@nau.edu or via phone at 602.909.3702**.

Consent Version: 02/01/2018

NORTHERN ARIZONA UNIVERSITY

Project Number: 1176778-1
Approval Date: February 5, 2018
This stamp must be on all consenting documents

An Institutional Review Board responsible for human subjects research at Northern Arizona University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

AGREEMENT TO PARTICIPATE

I have read (or someone has read to me) this form, and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Subject Name:_____

Subject Signature:_____ Date:_____

AGREEMENT TO BE AUDIORECORDED

Subject Signature:_____ Date:_____

Signature of Investigator/Individual Obtaining Consent:

To the best of my ability, I have explained and discussed the full contents of the study including all of the information contained in this consent form. All questions of the research subject and those of his/her parent or legal guardian have been accurately answered.

Investigator/Person Obtaining Consent:_____

Signature:_____ Date: _____

Consent Version: 02/01/2018

Appendix E

District Approval

January 8, 2018

Dr. Jason Reynolds
Assistant Superintendent – Secondary
Paradise Valley Unified School District
15002 N. 32nd Street, Phoenix, AZ 85032

RE: Permission to Conduct Research Study

Dear Dr. Reynolds:

I am writing to request permission to conduct a research study within the Paradise Valley Unified School District. I am currently enrolled in the Doctoral Program at Northern Arizona University and am in the process of writing my dissertation. The study is entitled *STUDENT STANDARDIZED TEST SCORES AS AN EFFECTIVE MEASURE OF TEACHER PERFORMANCE: TEACHER AND ADMINISTRATOR ATTITUDES.*

I hope that the district administration will allow me to consult with high school principals in order to identify mathematics teachers to anonymously complete a digital questionnaire sent via email (copy enclosed). The study would also include surveying administrators who supervise and/or train mathematics teachers. Additionally, I would like to conduct one on one interviews with approximately ten teachers and five administrators in order to develop greater insight into their attitudes about the use of student standardized test results as a measure of teacher performance.

If approval is granted, participants will complete the survey digitally. The survey process should take no longer than 20 minutes and individual interviews will last no longer than an hour. The survey and interview results will be pooled for the dissertation data project and individual results of this study will remain absolutely confidential and anonymous. Informed consent will be provided to all participants. Should this study be published, only aggregate results will be documented. No costs will be incurred by the district, schools or the individual participants.

Your approval to conduct this study will be greatly appreciated. I would be happy to answer any questions or concerns that you may have at this time. You may contact me at my email address: cjl34@nau.edu. If you agree, kindly sign below and I can pick up this form at your convenience. Additionally, kindly submit a signed letter of permission on your institution's letterhead acknowledging your consent and permission for me to conduct these surveys/interviews at your institution.

Sincerely,

Chad J. Lanese – Northern Arizona University

cc: Dr. Richard Wiggall, Dissertation Chair, NAU

Approved by:

Jason W. Reynolds, Ed.D - Assistant Superintendent — Print your name and title here

Jason W. Reynolds — Signature

2.6.18 — Date

Paradise Valley
Unified School District

Governing Board
District Administrative Center
15002 North 32nd Street
Phoenix AZ 85032
602.449.2298

February 8, 2018

RE: Permission to Conduct Research Study

Dear Mr. Lanese:

Please be aware that after review of your proposed research plan, the decision has been made to approve implementation of this study in the Paradise Valley Unified School District. This letter serves as permission allowing you to do so. You are now able to conduct the aspects of the study as outlined in your request for permission. We wish you the best of luck with your research.

Sincerely,

James P. Lee, Ed.D.
Superintendent

dc

Appendix F

Principal Letter

January 8, 2018

Dear Principal:
I am writing to request information from you about the identification of mathematics teachers at your site. I have recently gained permission from Dr. Reynolds to conduct a research study within the Paradise Valley Unified School District. I am currently enrolled in the Doctoral Program at Northern Arizona University and am in the process of writing my dissertation. The study is entitled *STUDENT STANDARDIZED TEST SCORES AS AN EFFECTIVE MEASURE OF TEACHER PERFORMANCE: TEACHER AND ADMINISTRATOR ATTITUDES.*

My hope is to consult with high school principals in order to identify mathematics teachers to confidentially complete a digital questionnaire sent via email. The purpose of the survey is to develop greater insight into their attitudes about the use of student standardized test results as a measure of teacher performance. Your assistance in the identification of specific mathematics teachers would prove most helpful. Below are a list of mathematics courses that are offered by each high school in the Paradise Valley Unified School District. Please list the teachers at your site that are responsible for instructing these courses. Please include teachers in ALL courses they are responsible for teaching for the 2017-18 school year, even if that means that are listed twice under different course titles.

Course Title	Instructor Name	Email
Accelerated Algebra 1-2		
Algebra 1-2		
Geometry 1-2		
Honors Geometry 1-2		
Algebra 3-4		
Honors Algebra 3-4		
Pre-Calculus 1-2		
Honors Pre-Calculus 1-2		
AP Calculus BC 1-2		
AP Calculus AB 1-2		
AP Statistics 1-2		
Statistics and Probability 1-2		

I very much appreciate your support and would be happy to answer any questions or concerns that you may have at this time. You may contact me at my email address: cjl34@nau.edu.

Sincerely,
Chad J. Lanese – Northern Arizona University
cc: Dr. Richard Wiggall, Dissertation Chair, NAU

Appendix G

Teacher Cover Letter

Dear Colleague. You are being invited to participate in a research study titled Student Standardized Test Scores As An Effective Measure of Teacher Performance: Teacher and Administrator Attitudes. This study is conducted by Chad Lanese from Northern Arizona University.

The purpose of this research study is to describe the attitudes of high school mathematics teachers and administrators towards the use of student standardized test results on performance evaluations. If you agree to take part in this study, you will be asked to complete an online survey/questionnaire. This survey/questionnaire will ask about topics related to years of experience, current position, standardized tests and experiences with teacher evaluation. It will take you approximately 15 minutes to complete.

You may not directly benefit from this research; however, I hope that your participation in the study may assist in better defining educator attitudes towards how standardized test results may be used in performance evaluations.

I believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach of confidentiality is always possible. To the best of my ability your answers in this study will remain confidential. I will minimize any risks by limiting identifiable information and maintaining anonymity by not collecting names or addresses. All data collected in the study will be maintained and disposed of by the researcher and utilized for the sole purpose of this study.

Your participation in this study is completely voluntary and you can withdraw at any time. You are free to skip any question that you choose. If you choose not to participate it will not affect your relationship with Northern Arizona University or result in any other penalty or less of benefits to which you are otherwise entitled.

If you have questions about this project or if you have a research-related problem, you may contact the researcher, Chad Lanese at cjl34@nau.edu or 602.909.3702. If you have any questions concerning your rights as a research subject, you may contact Northern Arizona University IRB Office at irb@nau.edu or (928) 523-9551.

By submitting this survey, I affirm that I am over 18 years of age and agree that the information may be used in the research project described above. Your participation is most sincerely appreciated!

Biographical Information

Chad Jeffrey Lanese was born in Cleveland, Ohio but moved to the state of Arizona at a very young age. He is the youngest child of Connie and Jeffrey Lanese and has one older brother, Troy. Chad earned his Bachelor's degree in Elementary Education from the University of Arizona in 1998 and began his career as a first grade bilingual teacher. He has taught first grade, third grade, fourth grade, and seventh/eighth grade American History, Government and Economics. Chad earned his Master's degree in Bilingual Education from Arizona State University in 2002 and his principal certification from Northern Arizona University in 2005. Chad has served in education for 21 years and the past 15 have been in the capacity of assistant principal and principal at the elementary and high school levels. He is currently the principal at Pinnacle High School in Paradise Valley Unified School District. Chad has two children, Katelyn and Ethan, with his wife Kelly. Chad is currently completing his doctoral degree in Educational Leadership from Northern Arizona University. Chad is eager to make a positive impact on students, staff, and stakeholders as an educational leader.