

THE IMPACT OF INTEGRATING STUDENT LEARNING OBJECTIVES IN
MISSOURI TEACHER EVALUATIONS ON STUDENT ACADEMIC
ACHIEVEMENT

© Copyright by

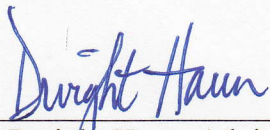
STEVEN GALLIVAN

2019

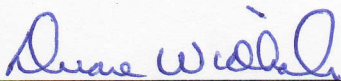
The undersigned, approved by the Department Chair of Graduate Studies in Education, have examined a dissertation entitled:

THE IMPACT OF INTEGRATING STUDENT LEARNING OBJECTIVES IN
MISSOURI TEACHER EVALUATIONS ON STUDENT ACADEMIC
ACHIEVEMENT

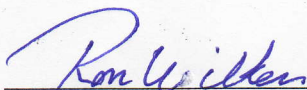
Presented by Steven Gallivan a candidate for the degree of Doctor of Education and hereby certify that in their opinion it is worthy of acceptance.



Dr. Dwight Haun, Advisor/Chair
Teacher Education, Southwest Baptist University



Dr. Duane Widhalm, Committee Member
Graduate Education, Southwest Baptist University



Dr. Ron Wilken, Committee Member
Area Supervisor, Missouri Department of Elementary and Secondary Education

THE IMPACT OF INTEGRATING STUDENT LEARNING OBJECTIVES IN
MISSOURI TEACHER EVALUATIONS ON STUDENT ACADEMIC
ACHIEVEMENT

A Dissertation
Presented to
The Faculty of the Graduate Education Department
Southwest Baptist University

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Education

By
Steve Gallivan, B.S., M.S., M.A.
Dr. Dwight Haun, Dissertation Advisor

May 2019

ACKNOWLEDGMENTS

This dissertation is dedicated to my loving wife, Sara, and our four children, Sloan, Samantha, Shane, and Sterling. Without their constant support and understanding, the completion of this dissertation would not have been possible. I want to thank my parents, Robert and Kathy, for providing encouragement and support of this endeavor and for teaching me the value of work ethic and perseverance during my childhood. I want to also say thank you to all my other family members for the constant reminders and words of encouragements to finish this dissertation.

Professionally, I want to say thank you to my dissertation advisor, Dr. Dwight Haun for his steady encouragement, advice and answering my never-ending e-mail questions regarding all things related to the dissertation. Also, thank you to my other committee members, Dr. Duane Widhalm, and Dr. Ron Wilken. Their comments and support have been invaluable to the completion of this dissertation. I would be remiss if I did not say thank you to the faculty and staff of the Southwest Baptist University Graduate Education Department. Special thanks to Dr. Benny Fong for his comments and help with the statistics portion of the dissertation.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	ii
TABLE OF CONTENTS.....	iii
LIST OF TABLES.....	v
ABSTRACT.....	vi
INTRODUCTION.....	1
Conceptual Framework.....	2
Problem Statement.....	5
Rationale for the Study.....	5
Research Questions.....	6
Null Hypotheses.....	7
Assumptions/ Limitations/Delimitations.....	8
Design Controls.....	11
Definitions of Key Terms.....	11
Summary.....	12
REVIEW OF RELATED LITERATURE.....	13
Introduction.....	13
History of Education Evaluations in the United States.....	13
Use of Student Data in Teacher Evaluation.....	20
Value-Added Models and Student Learning Objectives.....	21
Issues of Using Standardized Tests in VAMs and SLOs in Evaluations.....	24
SLOs’ and VAMs’ Impact on Student Achievement.....	28
Faculty and Administrators’ Perceptions of VAMs and SLOs.....	32
Missouri Assessment Program.....	34
Summary.....	35

RESEARCH DESIGN AND METHODOLOGY	37
Introduction.....	37
Participants.....	38
Research Questions.....	38
Instrumentation.....	40
Research Procedure.....	41
Summary.....	45
ANAYLSIS OF DATA.....	47
Introduction.....	47
Analysis of Data.....	48
Data Presentation.....	50
Summary of Findings.....	60
Summary.....	64
CONCLUSION.....	65
Introduction.....	65
Research Questions and Null Hypotheses.....	66
Professional Implications/Application to Academic Practice.....	70
Recommendations for Future Research.....	73
Conclusion.....	74
REFERENCES.....	75
APPENDIX: Survey Instrument Questions.....	86

LIST OF TABLES

1. Analysis of Variance for First Time Use of SLO (Yes vs. No) in Teacher Evaluation in 2015-2016.....	51
2. Analysis of Variance for First Time Use of SLO (Yes versus No) in Teacher Evaluations in 2016-2017.....	52
3. Analysis of Variance for SLO Use (Yes versus No) in 2016-2017 Teacher Evaluations.....	54
4. Analysis of Variance for Percentage Weight of Teachers' Summative Evaluation Tied to SLO Growth Data in 2016-2017.....	56
5. Analysis of Variance for Teachers That Met SLO Goals in 2016-2017.....	58
6. Analysis of Variance Type of SLOs Used in Teachers Evaluation in 2016-2017..	60

ABSTRACT

Missouri Department of Elementary and Secondary Education (MODESE) requires districts to follow the Missouri Principles of Effective Evaluation, which utilizes student growth data for teacher evaluation purposes. The purpose of this causal-comparative study was to quantify the impact of integrating student learning objectives (SLOs) in Missouri teacher evaluations on student academic achievement as measured by the Missouri Biology and Government end of course (EOC) exams. The study utilized a 18-question demographic survey to identify Missouri high schools that employed SLOs in teacher evaluations for Biology and Government teachers in the 2014-2015, 2015-2016, and 2016-2017 school years. The study used simple mixed design ANOVA test on the 2015, 2016, and 2017 MPI scores from MODESE's Biology and Government EOC tests to investigate the impact of the use of SLO's for teacher evaluation purposes on student academic achievement. The results of the study indicated the use of SLOs in teacher evaluations did not have a statistically significant difference on student achievement on Missouri Biology and Government EOC scores. The study did suggest the type of SLO used, the percentage weight given to SLO data in teacher evaluations, and whether teachers met SLO growth goals did have a medium and large effect size, as measured by partial eta square, on the difference of the variances in the 2016-2017 Missouri Biology and Government EOC scores, even though the differences were not statistically significant. This study adds data to the limited empirical research regarding the use of SLOs in teacher evaluations in Missouri and identified topics of future research regarding the use of student growth data for evaluation purposes.

CHAPTER ONE

INTRODUCTION

On June 29, 2012, Missouri’s Department of Elementary and Secondary Education (MODESE) received a flexibility waiver from some of the requirements of the United States Department of Education (USDOE) federal educational program, No Child Left Behind (MODESE, 2013b). As one of the components of the waiver, MODESE (2013b) was required to develop strategies to address the following three major components: “(1) college and career ready expectations for all students; (2) state developed differentiated recognition, accountability and support; and (3) structures for the support of effective instruction and leadership” (p. 4). Specifically, MODESE had to follow the 2011 USDOE Elementary and Secondary Education Act (ESEA) waiver requirement to develop a rigorous teacher evaluation program to measure teacher effectiveness that required student growth data as a significant contributing factor of the evaluation (Stiggins, 2014). In addition, the waiver required states to make student growth scores from annual tests a considerable percentage of a teacher’s evaluation score (Rich, 2012).

On December 10, 2015, President Barack Obama signed the Every Student Succeeds Act (ESSA) into law (USDOE, n.d.). The ESSA replaced the original requirements of the USDOE requiring state departments to use student growth data as a significant contributing factor in teacher evaluations (USDOE, n.d.). Even though federal requirements on the components of teacher evaluations shifted to state control, MODESE still required districts to follow the Missouri Principles of Effective Evaluation, which utilized student growth data for evaluation purposes (MODESE, 2014;

MODESE, 2016). This study examined the impact of using student growth data/learning objectives for teacher evaluations on student achievement.

The report *The Widget Effect* (Weisberg et al., 2009) suggested the current method of performance-based teacher evaluations (PBTE) was not accurately measuring teacher effectiveness. The report suggested that current PBTE methods only measured teacher performance and did not contribute to the growth of pedagogical practices leading to increased student achievement (Weisberg et al., 2009). Coincidentally during the same year, a 2009 doctoral study maintained that Southwest Missouri educators' perceptions of PBTE were not favorable to the improvement of pedagogy practices (Adams, 2009), which supported the national observation of *The Widget Effect* on a regional, local level. Both the Adams (2009) study and *The Widget Effect* (Weisberg et al., 2009) supported the observation of Tucker and Stronge (2005) that student learning is the primary goal of education and that student achievement is difficult to measure through classroom observation alone. Tucker and Stronge further suggested that a more equitable evaluation would “involve an assessment of the act of teaching as well as the results of teaching” (p. 7).

Conceptual Framework

The current study was based on the constructs of two landmark educational research studies that led to the development of value-added modeling (VAM) and student learning objectives (SLOs) as instruments to measure teacher effect on student academic growth and test scores (Slotnik, Smith, Glass, & Helms, 2004; Wright, Horn, & Sanders, 1997). Wright et al. (1997) suggested the VAM developed by Bill Sanders demonstrated the effect of high- and low-quality teachers on student achievement (Tucker & Stronge,

2005; Wright et al., 1997). The model was developed to meet the theoretical framework of Ralph Tyler's desired purpose that teacher evaluation be based on the analysis of the students' performance from a stated outcome to the actual outcome (Sanders & Horn, 1994). Ralph Tyler suggested that if districts are going to study the success of an educational program, then researchers must know the goals or educational objectives of the program (Tyler, 1949). The measurement of the actual outcomes to the stated goals will provide districts with data to make informed decisions to take significant and valid corrective action (Tyler, 1949). The Tennessee Value-Added Assessment System (TVAAS) model developed by Bill Sanders is a statistical model that focuses on the academic gains of a student from year to year using the scaled scores of the Tennessee Comprehensive Assessment Program (Sanders & Horn, 1994). The TVAAS model has been shown to establish the efficacy of teachers and school districts by utilizing a database that tracks students and their academic outcomes to the assigned teachers (Sanders & Horn, 1998). Sanders and Horn's (1998) research using the TVAAS suggested that residual effects of a poor teacher stay with students their entire educational lifetime, so teacher evaluations need to have a portion of the evaluation tied to student achievement.

The practical use of SLOs in teacher evaluation came from the study of the effect of SLOs in the teacher compensation program in Denver Public Schools on student achievement (Slotnik et al., 2004). The positive results of the Denver study combined with the initial positive results on student achievement when using SLOs in teacher compensation in Austin Public Schools and Charlotte-Mecklenburg Public Schools provided districts an alternative instrument other than VAM for including student growth

measures (SGMs) in teacher evaluation (Schmitt et al., 2009; Slotnik, Smith, Helms, & Qiao, 2013; Slotnik et al., 2004). In addition, prominent educational researchers further developed evaluation models that included evidence of student achievement as measurement of teacher effectiveness (Marzano & Toth, 2013; Popham, 2013). While educational researchers suggest that setting goals in the classroom has a positive effect on student achievement (Hattie, 2009; Marzano, 2009), a review of the research in the use of SLOs in teacher evaluations has suggested mixed results on the impact of SLOs in teacher evaluations to improve student achievement (Lachlan-Hache, 2015).

The use of SLOs as an instrument to measure student growth in education is not new. Ames and Archer (1988) suggested that mastery goal setting within a classroom led students to report having a greater motivation to complete the education work. As such, teachers have used SLOs to guide instruction and self-reflection in an effort to help improve student achievement (MODESE, 2013b). Professional learning communities (PLCs) have used S.M.A.R.T (specific, measurable, attainable, reasonable, time sensitive) goals to help enhance and focus instruction in an effort to improve student achievement, so the education community is very familiar with the concept of SLOs (MODESE, 2013b). Researchers have defined SLOs as an educational practice tool that promotes self-reflection by utilizing the following steps: review standards, develop core concepts, set goals of student achievement on the core concept, monitor progress, and review the outcome (Lachlan-Hache, Cushing, & Biovana, 2012). Even though the use of SLOs in the class is familiar, researchers suggest the implementation of new evaluation instruments that require additional time is hampered due to a myriad of responsibilities that administrators and teachers have that compete for the limited amount of time

(Slotnik, Smith, & Liang, 2013). Research has indicated that as student achievement increases from the implementation of SLOs in teacher evaluations, then administrators and teachers may develop a time management system and a positive belief system towards the use of SLOs in teacher evaluations (Slotnik, Bugler, & Liang, 2015).

Problem Statement

MODESE is requiring districts to use student growth data as a significant and contributing factor in the evaluation of a school district's teachers. Teachers have been using SGMs or SLOs for years in the assessment of pedagogy practices on student classroom achievement (MODESE, 2013b). Essential Principle 4 of Missouri's Essential Principles of Effective Evaluations requires districts to use multiple sources of data, including state standardized testing if available, to measure teacher performance (MODESE, 2013a). The issue is whether the mandated inclusion of student achievement data in teacher evaluation actually has a significant impact on student achievement. Early research studies were conducted with teachers and administrators willing to implement SLO procedures for teacher compensation programs and not mandated for teacher or principal evaluation purposes (Balch & Springer, 2015; Cornetto, Schmitt, Malerba, & Herrera, 2010; Schmitt, 2011; Slotnik, Smith, Helms, et al., 2013; Slotnik et al., 2004). Therefore the problem is the MODESE required school districts to use the inclusion of student growth data in teacher evaluations as a mandate, which is not supported by consistent research data to have a positive effect on student achievement.

Rationale for the Study

The purpose of this causal-comparative study was to quantify the impact of integrating SLOs in Missouri teacher evaluations on student academic achievement as measured by the Missouri Biology and Government end of course (EOC) exams.

Implementation of SLOs in teacher evaluation is a current trend in education; therefore research data on the use of SLOs in non-value-added teacher evaluation models was limited (Lachlan-Hache, 2015; Longchamp, 2017). This study will add further evidence to the body of research regarding SLOs' impact on student achievement through the mandated use of SLOs by MODESE in teacher evaluations, instead of voluntary local decisions to employ SLOs to enhance student achievement. Researchers have suggested that standardized test scores should not be used in teacher evaluations as technical characteristics of developing standardized tests might limit the true measurement of student knowledge, particularly knowledge gained from the teacher's instruction (Braun, 2005; Baker et al., 2010; Hanushek & Rivkin, 2010; Darling-Hammond, Amerin-Beardsley, Haertel, & Rothstein, 2012; Haertel, 2013; Stiggins, 2014). This study could also help educators, researchers, and policy makers develop or refine future mandates to evaluate teachers using standardized assessments, evaluate schools' accountability ratings on standardized test scores, and increase student achievement in Missouri.

Research Questions

This study sought to answer the following questions:

1. What is the difference in Government and Biology EOC scores when schools included student learning objectives in teacher evaluations?
2. What is the difference in Government and Biology EOC scores for schools that assigned SLO growth goal data different percentage weights in the teachers' summative evaluation final score?
3. What is the difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes?

4. What is the difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes?

Null Hypotheses

In an effort to answer the aforementioned questions the following four null hypothesis were investigated:

1. H_{10} – There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.
2. H_{20} – There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score
3. H_{30} – There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.
4. H_{40} – There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

Assumptions/Limitations/Delimitations

While it was the intent of the researcher to provide a study free from bias and outside forces that would influence the measurement of error on the variable being studied, the researcher was cognizant of the factors that could have influenced the study. The assumptions of the study that provided relevance to the study but were outside the researchers control were as follows (Simon, 2011):

1. Survey participants would answer the questions truthfully.
2. The use of student growth measures, which include SLOs, in teacher evaluations will continue to be important in evaluating teacher performance.

To encourage participants to answer honestly and truthfully, participants' survey answers and identity of schools will not be published to ensure confidentiality and anonymity. The current trend in teacher evaluations is to try and use student learning outcomes as a measure of teacher effectiveness. The report *The Widget Effect* (Weisberg et al., 2009) suggested the current method of PBTEs was not accurately measuring teacher effectiveness. The report suggested that current PBTE methods only measured teacher performance and did not contribute to the growth of pedagogical practices leading to increased student achievement (Weisberg et al., 2009). *The Widget Effect* (Weisberg et al., 2009) supported the observation of Tucker and Stronge (2005) that student learning is the primary goal of education and that student achievement is difficult to measure through classroom observation alone. Tucker and Stronge further suggested that a more equitable evaluation would "involve an assessment of the act of teaching as well as the results of teaching" (p. 7).

The limitations that were outside of the researcher's control that could have influenced the study were as follows:

1. The use of standardized test (Biology and Government EOCs) scores to measure student academic achievement, instead of alternative assessments.
2. The inherent perception of administrators and teachers of the use of SLOs for teacher evaluation.
3. The amount of training received by teachers and administrators on the development of SLOs for teacher evaluation purposes.
4. The rigor level used in the development of teachers' SLO goals on whether teachers met or not met the SLO goals for teacher evaluation purposes.

Researchers have suggested that standardized test scores should not be used in teacher evaluations as technical characteristics of developing standardized tests might limit the true measurement of student knowledge, particularly knowledge gained from the teacher's instruction (Braun, 2005; Baker et al., 2010; Hanushek & Rivkin, 2010; Darling-Hammond et al., 2012; Haertel, 2013; Stiggins, 2014). Standardized achievement tests cannot truly assess the real knowledge of students due to the construction of the test (questions, time limit) and other extrinsic factors unrelated to the subject matter that could affect measurement of student knowledge (Hanushek & Rivkin, 2010). Hanushek and Rivkin (2010) suggested this measurement error would lead to possible inaccurate measurements of teachers' effect on student achievement. No further antecedent public evidence is available to measure student success in the class for teacher evaluation purposes. The survey results are limited to the practicality of the use of standardized tests in measuring student achievement in a subject area. Researchers have presented results

that suggest a positive perception and a willingness of teachers/administrators to implement SLO procedures has led to increase use of SLOs by teachers, which had a corresponding positive effect on student achievement (Schmitt, 2014; Slotnik et al., 2015). The study was limited in that it only measured student achievement on EOC scores without a corresponding measure of the perceptions of a district on the use of SLOs. The study attempted to capture school districts' perception of SLOs by asking school districts what percentage level of the teachers' summative evaluation was associated with meeting SLOs goals. Limited experiential data exists on the difference of how much SLO implementation training administrators/teachers receive has on student achievement (Longchamp, 2017). A 2017 doctoral thesis suggested that teachers understand and see the impact of administrators' leadership on developing an effective school environment for SLO implementation (Longchamp, 2017). Longchamp (2017) suggested administrators that had received training on SLO implementation were better able to develop a culture of collaboration with teachers, which leads to positive attitudes towards the significance of the SLO process on student achievement. Therefore, the study was limited as it did not measure the amount of previous training that administrators received on SLOs, but it did try to measure the difference in student achievement by the type of SLO (unit or yearly growth goal) used by teachers. This could lead to further discussion on the type of training administrators need in the future.

Delimitations of the study that were controlled by the researcher focused on controlling the population as follows:

1. Participants in the survey were high school principals in the State of Missouri.

2. The study only focused on high school Science and History teachers that taught a class required to give the Biology or Government EOC.
3. Teachers were employed by public high schools in the State of Missouri.

Design Controls

This casual-comparative study utilized a demographic survey to identify Missouri high schools that implemented SLOs in teacher evaluations and reached SLO goal attainment for Biology and Government teachers in the 2015, 2016, and 2017 school years. After surveys were returned, schools were categorized by SLO implementation and SLO goal attainment. The 2016 and 2017 Biology and Government EOC scores, as the treatment, were compared to the 2015 Biology and Government EOC scores, as the control, to determine the impact of the SLOs on student achievement.

Definition of Key Terms

Achievement. Specific level met on End of course exams (MODESE, 2017a).

End of course exams. The state of Missouri's standardized tests to measure student knowledge in Biology, Government, Algebra I, Algebra II, English I, and English II and used for accountability purposes (MODESE, n.d.-a).

Student growth measures. Growth measure is the difference of a student's scores over two points in time as measured by a pre- and postassessment to show growth in a student's knowledge of a particular standard or standards (MODESE, 2013b).

Student learning objectives. A method used to quantify a teacher's impact on student achievement by using preassessment data to set student growth goals for a particular class or course and then using a postassessment to measure the percentage of students that met the predetermined growth goals for the class or course (MODESE, 2014).

Summary

The MODESE is requiring Missouri school districts to use SLOs/SGMs in teacher evaluations. The research is limited on the use of SLOs in teacher evaluations and their impact on student achievement. The chapter provided a brief review of the limited research of the use of SLOs in teacher evaluations and their impact on student achievement. Based on the review of the research, the rationale of this study was to quantify the impact of integrating SLOs of Missouri teacher evaluations on student academic achievement as measured by the Missouri Biology and Government EOC exams.

Chapter Two of this paper will provide a review of the literature on the use of SLOs in teacher evaluations. It is organized thematically using literature completed on SLO use in teacher compensation models. The review is used to build the background and conceptual framework to develop this study using SLOs in a teacher evaluation model, not based on a teacher compensation model. Chapter Three will describe the method for selecting schools that used SLOs in the evaluation of Biology and Government teachers. The chapter also outlines the instrumentation and how the data were treated for analysis. Chapter Four of this paper will outline the results of the statistical analysis. Chapter Five will provide the conclusion of the study by providing the professional implications and applications to academic practice of the use of SLOs in teacher evaluations.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Introduction

Chapter Two of this paper will provide a review of the literature on the use of VAMs and SLOs in teacher evaluations. It is organized thematically, starting with the historical summary of the education evaluation in the United States and then proceeds with a review of literature on VAM and SLO use in teacher compensation models. The review is used to build the background and conceptual framework to develop this study using student achievement data and SLOs in a teacher evaluation model not tied to a teacher compensation model.

History of Education Evaluations in the United States

During the 1700s the responsibility of the evaluation of local school districts fell to the local governmental town leaders and clergy (Marzano, Frontier, & Livingston, 2011). The evaluation of teachers was first conducted by committees of local town people and/or clergy and focused on the actions of the teacher and what was being taught, but not on the actual learning of the students (Anderson, 1993; Payne, 2010). The visits were considered “inspectorial visits” (Anderson, 1993, p. 5) and did not aid in the enhancement of the pedagogy skills of the teacher. As such, the feedback given to teachers was wide-ranging from school to school based on the experiences and education of the local leaders and clergy and a lack of standardized teacher evaluation standards (Marzano et al., 2011). This was the prominent method of school evaluation from 1642 to 1875 (Eye, Netzer, & Krey as cited in Payne, 2010).

In the mid-1800s, as the industrial complex increased and cities became larger so did the intricacy of the nation’s school system (Anderson, 1993). As the population grew

in the mid-1800s the need for the development of organized education evaluations was paramount and a standardized system of evaluation by school superintendents was developed by cities and school districts (Anderson, 1993; Payne, 2010; Starratt, n.d.). As the number of schools increased, the public began to realize that teaching was a specialized endeavor (Marzano et al., 2011).

During the 19th century and early 20th century, the work of Frederick Taylor started the period of the scientific view of management towards all sectors of industry (Marzano et al., 2011). The premise of Taylor's work was that all actions within a factory are measured to determine the efficiency of an action in order to select the best action for a desired outcome (Marzano et al., 2011). During this time period, educational theorists Edward Thorndike and Ellwood Cubberley began to espouse the use of scientific measurement to help determine the success of a school district (Marzano et al., 2011). The scientific method of measurement caused Cubberley to use the metaphor of a factory to describe students as the unfinished products that school would mold into a finished product based on the needs of the civilization (Cubberley, 1916; Marzano et al., 2011). To bolster his case of the use of scientific view management in education, Cubberley (1916) provided a comparison between agriculture and education. Cubberley noted that if the scientific view of management worked in providing standardized, measurable results in agriculture yields (with all the uncontrollable factors of production), then it should work in the field of education despite the organic nature of humans and the associated uncertainties of life.

In 1929, William Wetzel advocated the use of student learning measures in addition to analyzing specific pedagogical techniques and behaviors of teachers to

quantify the successful impact of teachers on student learning (Marzano et al., 2011). Wetzal developed the following three recommendations to use student learning as a measurement tool: use aptitude tests, develop clear learning objectives for each class, and incorporate pertinent student learning measures (Marzano et al., 2011). Even though the scientific view of management was becoming the process to evaluate schools and teachers, educational researchers were discussing another educational philosophy.

It should be noted that during this time, John Dewey (1938) was writing that the goal of school was to develop student education through experiences. Dewey was not against the use of learning goals, but the development of an individual's learning goals should be ongoing, not fixed, as the experience of the individual changes (Watras, 2006). This philosophy is based on the idea that experiences would lead to greater experiences, which could lead to an increasingly comprehensive education rather than an education limited to ridged preset instructional goals set by education officials (Dewey, 1916; Watras, 2006).

In response to a complaint from high schools about how colleges were restricting flexibility in the school curriculum, the Progressive Education Association conducted the Eight-Year Study from 1932 to 1940 (Watras, 2006). The study was initially started as an evaluation research project to study the success of students in colleges from high schools that taught the subject required for colleges but not the curriculum required by colleges (Hogan, 2007; Watras, 2006). As funding increased for the project the study goals moved from an evaluation of progressive innovative approaches to an evaluation of educational issues and curriculum (Watras, 2006). During the study Ralph Tyler developed his

rational for determining the success of an educational program, which was based on the growth of student knowledge as the key component (Watras, 2006).

Ralph Tyler suggested that if districts are going to study the success of an educational program, then researchers must know the goals or educational objectives of the program (Tyler, 1949). The measurement of the actual outcomes to the stated goals would provide school districts with data to make informed decisions to take significant and valid corrective action on improving curriculum and teaching (Tyler, 1949). The model of Tyler's evaluation was based on the following steps: (a) faculty developed objectives for a class curriculum; (b) faculty arranged the objectives based on difficulties and specific action; (c) faculty then developed actions that students had to demonstrate; (d) faculty developed activities/situations for students to use desired skills; (e) faculty then developed/selected assessment tools that indicated mastery of each objective; (f) faculty would then assess the accuracy of assessment item and refine as needed; and (g) faculty analyzed data to ascertain the success or failure of a particular program on student growth (Watras, 2006, p.13). Tyler's model of evaluation, as summarized by Watras (2006), was for a teacher to develop specific student objectives that would show success or failure of students' growth in meeting the desired objective (Watras, 2006). Therefore, Tyler's evaluation of student growth was based on the academic concern for a particular teacher with a particular group of students during a particular year (Watras, 2006). This was the start of the use of objective-based evaluation and lead Tyler to be called the pioneer of the objectives-oriented approach to evaluation (Hogan, 2007; Stufflebeam & Shinkfield, 1985).

During the period of the 1930s to 1950s, school administrations were being pulled from the scientific method of evaluation (student data results) to a method that focused on the human relations component of teaching (Pajak & Arrington, 2004; Payne, 2010; Starrat, n.d). The reason for this move in evaluation was the criticism by John Dewey of the difficulty in developing a standardized teacher evaluation across schools because of Dewey's philosophy that education was an art that led to further educational experiences (Glanz, 2008). The premise of Dewey's argument was that determining a teacher's final evaluation score based on scientific method (standardized scoring) was not always applicable because the educational indicators measured may have led a student to consider a different line of thought (that was influenced by the teacher, but not measured by the evaluation), which led to an additional educational experience (Glanz, 2008). Thus, the standardized teacher evaluation process would not accurately give credit to the teacher for causing the student to think about and then consequently develop a new educational experience. The standardized teacher evaluation could not possibly account for all of the student's future additional educational experiences created by the direct and indirect influences of the teacher, which leads to the increase of the student's overall education because of the cumulative effect of experiences building upon each other, thus bolstering Dewey's case that education is an art of experiences (Dewey, 1938; Glanz, 2008). It was during this time teacher evaluations slowly moved from objective-based reasons to measuring the teaching skills and not on actual results of the teaching on student achievement.

During the 1960s, teacher evaluations had moved away from the rigid scientific method (standardized ratings) and focused more on the teacher's actions to control

classroom management and environment, thus evaluation were considered more casual in nature (Aseltine, Faryniarz, & Rigazio-DiGilio, 2006; Payne, 2010). This was the method for teacher evaluation despite calls for greater scrutiny of education from federal government due to the Elementary and Secondary Act, Vocational Education Act, and the space race with Russia due to the 1957 launch of Sputnik (Pajak, 2000; Payne, 2010). During this time, researchers were proposing different evaluation models to help veteran and novice teachers to improve pedagogy skills to meet the demands of the increased scrutiny on education (Anderson, 1993).

In 1973, Morris Cogan developed a “several stage cycle” (Anderson, 1993, p. 6) evaluation idea of clinical supervision to assist school administrators in the process of improving classroom pedagogy skills for all levels of teachers. This process was outlined in the 1973 publication of Morris Cogan’s book entitled *Clinical Supervision* (Payne, 2010). Payne (2010) summarized Cogan’s definition of clinical supervision as “the focus on the improvement of the teacher’s classroom instruction” (p. 22) as a way to use data to understand the pattern of students’ learning in the classroom. The clinical supervision evaluation process was used throughout the 1970s with elements of the process used through the early 2000s (Holland & Garman, 2001). Payne summarized the purpose of the clinical evaluation process “is the development of the professionally responsible teacher who is analytical of his own performance, open to help from other and self-directing” (p. 23). This should lend to the improvement of a teacher’s pedagogy skills based on the targeted professional development from the suggestions of clinical supervision (Acheson & Gall, 1997). Currently, the teacher evaluation process has shifted from clinical supervision models to the inclusion of student academic growth data

in teacher evaluations. This move is best summarized by a quote from Marzano et al. (2011): “If student achievement is not linked to teacher evaluation, teachers have little incentive to develop into experts” (p. 27).

In an article, Hiebert, Morris, Berk, and Jansen (2007) advanced the idea that teacher education programs use student learning as a method to evaluate one’s teaching. Gurl, et al. (2016) summarized Hiebert et al.’s evaluation framework into four processes: (a) developing instruction goals/objectives, (b) evaluating student responses to ascertain their knowledge, (c) determining if the teaching caused the academic achievement, and (d) then proposing modifications, if needed, to teaching to reach desired results. Hiebert and Morris (2012) summarized these four processes imply a movement toward the application of the use of student learning goals to improve the teaching pedagogy quality and not the teacher quality. Hiebert and Morris made the case that developing “artifacts” (p. 93) of knowledge focusing on student learning growth should be encouraged to develop teaching skills by teachers, instead of focusing on traditional acquisition of additional characteristics of effective teachers such as advanced course work and/or advanced degrees. This method puts the focus on the relationship of the teacher and student around the learning objective of the classroom, which is the heart of education (Hiebert & Morris, 2012; Hiebert et al., 2007).

The report *The Widget Effect* (Weisberg et al., 2009) suggested that the current method of PBTE was not accurately measuring teacher effectiveness. The report, based on surveys of 15,000 teachers and 1,300 administrators, challenged that current PBTE methods only measured teacher performance and did not contribute to the growth of pedagogical practices leading to increased student achievement (Weisberg et al., 2009). A

2009 doctoral dissertation investigating Southwest Missouri educators' perceptions of PBTE found these educators did not perceive improvement in pedagogical practices as an outcome of PBTE (Adams, 2009). Adams' (2009) study and *The Widget Effect* (Weisberg et al., 2009) supported the observation of Tucker and Stronge (2005) that student learning is the primary goal of education and that student achievement is difficult to measure through classroom observation alone.

Use of Student Data in Teacher Evaluation

Alkin and Christie (2004) cited Ralph Tyler's (1949) work on the Eight-Year Study for developing the initial theoretical framework of using learning objectives for education accountability purposes. The theoretical framework of Ralph Tyler's work is known as "objectives-oriented or objective-referenced" (Alkin & Christie, 2004, p. 18) evaluation. Ralph Tyler suggested that if districts are going to study the success of an educational program, then researchers must know the goals or educational objectives of the program. The measurement of the actual outcomes to the stated goals will provide districts with data to make informed decisions to take significant and valid corrective action (Tyler, 1949). The objective-oriented evaluation system is based on the premise of developing the primary educational objectives, subgrouping the objectives into different areas, selecting appropriate ways for students to display evidence of mastering the objectives, and then interpreting the results of educational accountability purposes (Alkin & Christie, 2004).

Ralph Tyler is considered the creator of the phrase "educational evaluation" (Alkin & Christie, 2004, p. 18) to summarize the accountability of teachers by measuring the difference between clearly defined learning objectives and the actual results to

determine the educational effectiveness of teachers (Hogan, 2007; Madaus & Stufflebeam, 1989; Sanders & Horn, 1994). Using assessments to measure student knowledge of desired learning objectives assists educators in discerning which educational processes have the desired impact on student achievement (Sanders & Horn, 1994). Stiggins (2014) summarized the decision to use standardized test scores in teacher evaluations is rooted in the notion that teachers would teach to the test. Therefore, the use of an “objective-third party” (Stiggins, 2014, p. 21) analysis of student growth would help to eliminate the concern of teachers manipulating the data due to their use in personnel decisions. Researchers then used this theoretical framework to develop and assess teacher evaluation systems for accountability purposes (Amrein-Beardsley, 2008; Braun, 2005; Briggs, 2012; McLean & Sanders, 1984; Murphy, 2012; Sanders & Horn, 1994, 1998; Webster & Mendro, 1997). Using the argument that effective teaching should lead to increase student achievement, Bergin (2015) stated that it is the expectation of the current emphasis on using student data in evaluations that should allow the education community a more accurate way to gauge the effectiveness of teacher in school districts versus the old observational evaluation system. However, Bergin also stated there is not a fundamental way, nor agreeable way, among researchers of how to effectively use student achievement data in teacher evaluations.

Value-Added Models and Student Learning Objectives

During the recent decade, accountability systems have moved from compliance in nature to a system that values student academic outcomes as a measure of teacher accountability (Murphy, 2012). The main purpose of outcome-based accountability is the evaluation’s emphasis on a nonsubjective measure of student scores placing incentive on

teachers to have higher scores, which leads to increased performance in the classroom (Murphy, 2012). Two landmark educational research studies led to the development of VAMs (Sanders & Horn, 1994) and SLOs as instruments to measure teachers' effect on student scores (Slotnik et al., 2004). Researchers have suggested the value-added model developed by Bill Sanders demonstrated the effect of high- and low-quality teachers on student achievement (Sanders & Rivers, 1996; Tucker & Stronge, 2005; Wright et al., 1997). The VAM was developed based on the theoretical framework of Ralph Tyler that teacher evaluation be based on the analysis of the students' performance from a stated outcome to the actual outcome (Sanders & Horn, 1994). VAMs help to provide further data by showing growth from year to year of a student's academic achievement score (Murphy, 2012). The TVAAS model developed by Bill Sanders is a statistical model that focuses on the academic gains of a student from year to year using the scaled scores of the Tennessee Comprehensive Assessment Program (Sanders & Horn, 1994). The TVAAS model has been shown to establish the efficacy of teachers and school districts by utilizing a database that tracks students and their academic outcomes to the assigned teachers (Sanders & Horn, 1998). Sanders and Horn's (1998) research using the TVAAS suggested that residual effects of a poor teacher stay with a student their entire educational lifetime, so teacher evaluations need to have a portion of the evaluation tied to the student achievement. VAMs using student growth data provide a truer picture of academic growth versus the attainment model, which compares a whole class from year to year (Murphy, 2012). According to Murphy (2012), attainment-based systems do not provide an accurate measure of a single teacher's influence on a class due to the system's inability to account for the effect of past achievement level and the demographic of each

comparable class. Braun (2005) stated the use of VAMs, based on student academic growth from beginning to the end of the year, provide school districts a teacher evaluation model that shows actual increase in student learning, which is the primary goal of pedagogy. Also, through the use of VAMs, school districts can add a quantifiable aspect to the teacher evaluation model that has historically been dominated by subjective and qualitative data (Braun, 2005). A concern of using VAMs in teacher evaluation is the inherent measurement error standardized tests would have while trying to accurately predict teacher effectiveness on student achievement (Hanushek & Rivkin, 2010).

The practical use of SLOs in teacher evaluation came from the study of the effect of SLOs in the teacher compensation program in Denver Public Schools on student achievement (Slotnik et al., 2004). The positive results of the Denver study combined with the initial positive results on student achievement when using SLOs in teacher compensation in Austin Public Schools and Charlotte-Mecklenburg Public Schools provided districts an alternative instrument other than VAMs for including SGMs in teacher evaluation (Schmitt et al., 2009; Slotnik et al., 2004; Slotnik, Smith, Helms et al., 2013). While educational researchers suggest that setting goals in the classroom has a positive effect on student achievement (Hattie, 2009; Marzano, 2009), a review of the research in the use of SLOs in teacher evaluations has suggested mixed results on the impact of SLOs in teacher evaluations to improve student achievement (Lachlan-Hache, 2015).

The use of SLOs as an instrument to measure student growth in education is not new. Ames and Archer (1988) suggested that mastery goal setting within a classroom led students to report having a greater motivation to complete the education work. As such,

teachers have used SLOs to guide instruction and self-reflection in an effort to help improve student achievement (MODESE, 2013b). PLCs have used S.M.A.R.T goals to help enhance and focus instruction in an effort to improve student achievement, so the education community is very familiar with the concept of SLOs (MODESE, 2013b). MODESE has defined SLOs as a method used to quantify a teacher's impact on student achievement by using preassessment data to set student growth goals for a particular class or course and then using a postassessment to measure the percentage of students that met the predetermined growth goals for the class or course (MODESE, 2014). Educational researchers have defined SLOs as an educational practice tool that promotes self-reflection by utilizing the following steps: review standards, develop core concepts, set goals of student achievement on the core concept, monitor progress, and review the outcome (Lachlan-Hache et al., 2012).

Even though the use of SLOs in the class is familiar, researchers suggest the implementation of new evaluation instruments that require additional time is hampered due to a myriad of responsibilities that administrators and teachers have that compete for the limited amount of time (Slotnik, Smith, & Liang, 2013). Research has indicated that as student achievement increases from the implementation of SLOs in teacher evaluations, then administrators and teachers may develop a time management system and a positive belief system towards the use of SLOs in teacher evaluations (Slotnik et al., 2015).

Issues of Using Standardized Tests in VAMs and SLOs in Evaluations

Standardized test scores are a quantitative measurement of student achievement that provides objective evidence to a teacher evaluation that also relies on subjective measurement of student academic actions in the classroom (Braun, 2005). Standardized

achievement tests cannot truly assess the real knowledge of students due to the construction of the test (questions, time limit) and other extrinsic factors unrelated to the subject matter that could affect measurement of student knowledge (Hanushek & Rivkin, 2010). Hanushek and Rivkin (2010) suggested this measurement error would lead to possible inaccurate measurements of teachers' effect on student achievement. The issue of using standardized test scores across years when comparing scores from year to year is that the tests must be scaled through a process called equating, which can introduce uncertainty in the scale score (Braun, 2005). Braun (2005) stated adding to the uncertainty is the procedure used to develop the total score of the test, which is "not a simple sum of the number of correct responses" (p. 13). However, the total reported score is "weighted composite" (Braun, 2005, p. 13) developed from subscale scores acquired from statistical model calculations based off of raw data, then "applied to raw test data" (p. 13).

Classroom assessment expert Rick Stiggins (2014) summarized that standardized assessment should not be used for teacher evaluations for the following four reasons: (a) Tests given are timed, which limits the number of questions to fully assess a student's knowledge; (b) Due to the limited number of questions, certain key instructional standards may be left out of the test, even though the teacher spent time instructing students on the standard, which limits the ability to measure the impact of the "instructional responsibility" (p. 23) of the teacher on all the particular standards for a course; (c) The timed portion and multiple-choice setup of typical standardized tests potentially does not accurately measure the growth of a slow reader due to the inherent constraints to demonstrate knowledge on a timed test, when a different mode of

assessment could demonstrate student growth, thus muting the measured effect of the teacher on learning; and (d) Due to the call for tests to only cover grade spans, a teacher's measured effect could be dampened because of the "ceiling effect" (p. 22) where an advanced student would not have the evidence of student growth because they showed mastery of grade span even though they are maybe a grade or two higher because of the impact of the teacher. Stiggins (2014) suggested the four abovementioned technical characteristics of standardized tests causes the content of the test to be limited, which would not accurately measure a teacher's impact on student knowledge. Braun (2005) also stated his concern as to whether the test covers all the content standards for high level classes due to the constraints of making a test that can be finished in a prescribed time allotment. This causes a concern as to whether the test is a valid measure of what a student truly has learned in the classroom (Braun, 2005). In addition to the technical characteristics of standardized tests, a review of the research suggests that standardized tests cannot accurately account for teacher impact due to the effect of the multitude of factors outside of school on student achievement (Gurl et al., 2016; Stiggins, 2014).

Based the review of the research, Gurl et al. (2016) stated that many teacher preparations programs still "perpetuate technical approaches" (p. 21) and do not fully examine the reasons/cause for academic shortcomings of disadvantaged students. Gurl et al. further outlined that teacher preparation programs are still predominantly based on Euro-centric, middle class pedagogy practices even though classroom are becoming more diverse by race, culture, and socioeconomic status. Therefore, Gurl et al. summarized the use of standardized and high-stakes assessment tools struggle to have a place in teacher

assessment due to the variety of factors (diverse student socioeconomic backgrounds) that influence highly effective teaching

Darling-Hammond et al.'s (2012) review of research suggested that student achievement is affected by more factors than just a single teacher. Therefore, the use of a standardized test as the sole measurement of a teacher's impact on learning would be inadequate due to a review of research that indicates other factors that have an effect on student achievement (Darling-Hammond et al., 2012; Mathis, 2012). Gurl et al. (2016) further argued the point by suggesting that a teacher at a lower economic status school should not be held accountable by a standardized test, whose technical characteristics favor a different socioeconomic class because the teacher was trained with research-based teaching methods for a different socioeconomic class. Teachers must be trained and provided resources to always understand the "complexity" (Gurl et al., 2016, p. 23) of effective teaching by managing the intertwined challenges of "constructing knowledge" (Gurl et al., 2016, p. 23) in their rooms. The measurement tool of teacher performance does not have a relevant meaning unless the tool facilitates the teacher's effort to build an emotional rapport with students (Gurl et al., 2016). In addition, the measurement tool needs to add benefit to the "sociopolitical and cultural understanding of pedagogy" (Gurl et al., 2016, p. 22) to help teachers improve their teaching of the diverse student population. In addition, a 2017 Phi Delta Kappan (PDK) survey of 1,588 American adults indicated that 42% of surveyed adults believed that school quality should be measured by standardized tests.

SLOs' and VAMs' Impact on Student Achievement

The first landmark study to show a positive relationship between SLO and student achievement was the Slotnick et al. (2004) research in Denver, Colorado, which suggested SLOs had a statistically significant positive effect on middle school mathematics. The study addressed the initial research questions of the impact of the use of student growth data in teacher compensation programs and the impact of teacher-developed student learning objectives on student achievement (Slotnick et al., 2004). Slotnick et al. mentioned one of the limitations of the study was the issue that pilot schools voluntarily voted to participate in the study. While the results suggested a positive impact on student achievement, at certain building levels, school districts would have to create teacher buy-in on the effectiveness of high-quality student learning objectives on student achievement (Slotnick et al., 2004). Slotnick et al. also suggested the increased emphasis on student learning objectives forces districts and teachers to spend more time looking at teaching strategies that increase student achievement due to the high stakes that are tied to student growth achievement.

Research conducted by Schmitt et al. (2009) in Austin [Texas] Public Schools had results similar to what Slotnick et al. (2004) had in Denver. Schmitt et al. conducted research on the teacher compensation program in Austin Public Schools and found a positive association between SLOs and student performance. The study indicated that teachers who met the goals of the SLOs had students that performed at a higher level on the Texas Assessment of Knowledge and Skills on Math versus those students whose teacher did not meet the SLOs (Schmitt et al., 2009). The results of the study indicated

the increase in performance was noted only for the middle school level and not the elementary and high school.

Slotnik, Smith, Helms, et al. (2013) found similar results in a study conducted in North Carolina. Slotnik, Smith, Helms, et al. measured 5 years of student data based on a control group of students that received instruction from teachers not enrolled in the teacher compensation program and the experimental group of students that received instruction from the teachers enrolled in the teacher compensation program. Slotnick, Smith, Helms, et al. stated SLOs had a positive, statistically significant effect on student achievement in middle schools and elementary during the first year of SLO implementation. This study used a quasi-experimented design to try and provide data on the effect of the use of SLOs in a teacher compensation program on student achievement. In addition, the research suggested high-quality SLOs led to gains in student achievement versus low-quality SLOs (Slotnik, Smith, Helms, et al., 2013). The study noted the quality, depth of descriptions, and clear goals of the teacher's SLOs had a positive effect on student achievement in elementary and middle school and growth rates were higher at the schools enrolled in the teacher compensation program (Slotnik, Smith, Helms, et al., 2013). Slotnik, Smith, Helms, et al. suggested that SLOs can play an integral part in student achievement when used in teacher evaluations because the increased attention to detail and quality in the SLO development is key to the success of SLOs on student achievement.

Early research studies suggested that SLOs have a statistically significant positive impact on student achievement; however, later research has indicated either no effect or lack of a statistically significant effect on student achievement. Cornetto et al.'s (2010)

study of Austin Public Schools indicated no relationship between SLOs and student achievement during the second year as measured by the state standardized test. In a follow-up study, Schmitt (2011) found students performed better on the Texas standardized test when their teacher met at least one SLO, although the difference was not statistically significant. Balch and Springer's (2015) study of Austin Public Schools initially found a relationship of SLO use to improved student scores for math. However, when scores were statistically adjusted to account for other factors the relationship was zero. Balch and Springer suggested it is "impossible to separate the effects of the teacher-level performance incentives from the other simultaneously implemented components of the comprehensive reform" (p. 124).

Goldhaber and Walch's 2012 study of the Denver ProComp system suggested that scores increased for the district during the first year of Denver ProComp, but the student growth gains were seen for teachers no matter if they were enrolled in ProComp or not. However, Goldhaber and Walch stated even though the gains were inconsistent during the 10-year study, data indicated teachers who voluntarily enrolled in the program were effective based on test scores. A 3-year research study of Denver's ProComp teacher system, which utilized SLOs, based on Slotnik et al.'s 2004 research, suggested a positive effect in mathematics and negative effect in reading and writing, but not at a statistically significant level to imply that SLOs had a practical effect on student achievement (Briggs, Diaz-Bilello, Maul, Turner, & Bibilos, 2014).

VAMs have used statistical analysis to provide evidence of a teacher's effectiveness by measuring the impact on student achievement in the current year and in subsequent educational years, as well (Sanders & Rivers, 1996). VAM research also

suggests that as the effectiveness of teachers improves, students at lower achievement levels improve first (Sanders & Rivers, 1996). In addition, research suggests teachers with the highest level of effectiveness have an impact on high, middle, and low achievement levels (Sanders & Rivers, 1996). This suggests a linear relationship of teacher effectiveness on the impact at different student achievement levels (Sanders & Horn, 1998; Sanders & Rivers, 1996). Since Wright et al.'s (1997) study, educational researchers have suggested that VAMs cannot accurately account for all the influences on students that affect student outcomes due to the statistical nature of the VAM (Amrein-Beardsley, 2008; Berliner, 2014; Marzano & Toth, 2013; Stiggins, 2014). Two doctoral research studies also support the analysis of the statistical problems of VAM. A doctoral work in using Stronge's teacher performance evaluation system in Arizona (Monroe, 2014) found no statistical significance difference between the teacher summative score using a VAM and student growth on Arizona state assessment. Subsequently, Hu's (2015) doctoral work found that a weak association was present between the VAM estimates and the attainment of SLOs. The study noted the relationship was weak because of the fluctuations in students' achievement data from year to year and grade to grade, which had a direct effect on the VAM estimate. In addition, the study indicated that quality of SLO had a weak relationship with the teachers' VAM estimate and that quality of SLO had a weak relationship with SLO attainment (Hu, 2015). Results of VAMs should not be viewed as a truly accurate account of teacher effectiveness because of the nonrandomized nature of placing teachers and students in a particular classroom (Braun, 2005). In addition, the statistical nature of VAMs does not account for all outside

influences on student achievement such as intrinsic and extrinsic motivation factors (Braun, 2005).

Faculty and Administrators' Perceptions of VAMs and SLOs

Researchers have presented results that suggest a positive perception and a willingness of teachers/administrators to implement SLO procedures has led to increase use of SLOs by teachers, which has had a corresponding positive effect on student achievement (Schmitt, 2014; Slotnik et al., 2015). The majority of educators and administrators have a desire to see students succeed and understand the importance of using some form of student achievement data in the evaluation process (Zesiger, 2015). In a doctoral study, Zesiger (2015) surveyed teachers and administrators in Pennsylvania, asking participants to indicate the importance of including student achievement data in teacher evaluations and to outline any concerns that they had with the use of student data (Zesiger, 2015). The results indicated that respondents felt strongly about the importance of including student achievement data in evaluation, but had issues with how student achievement data were being calculated and included to measure teacher effectiveness (Zesiger, 2015). The Zesiger study supported the results of a different doctoral study regarding the use of student achievement data in evaluation. A 2017 doctoral thesis suggested that teachers understand and see the impact of administrators' leadership on developing an effective school environment for SLO implementation (Longchamp, 2017). Longchamp (2017) suggested administrators that had received training on SLO implementation were better able to develop a culture of collaboration with teachers, which leads to positive attitudes towards the significance of the SLO process on student achievement. In a survey of teachers, there was consensus that SLOs could be an

effective tool to measure student learning growth and evaluate teachers (Pollins, 2014). Marzano et al. (2011) made the observation that teachers have no “incentive to develop into experts” (Lessons From History section, para.1) unless student academic performance is tied to teacher performance evaluations. A survey of administrators and teachers found staff felt pressured because of the myriad of responsibilities that compete for the limited amount of time (Slotnik, Smith, & Liang, 2013). However, survey research conducted in Maryland indicated if the research shows that if student achievement increases from the implementation of SLOs in teacher evaluations, then administrators and teachers may develop a time management system and a positive belief system towards the use of SLOs in teacher evaluations (Slotnik et al., 2015). Kraft, Brunner, Daugherty, and Schwegman’s (2018) research suggested adoption of high-stakes teacher evaluations using student test data and VAMs lowered the number of new teacher licenses by 15%. With an estimated enrollment in public school set to rise by 1,000,000 students over the next 10 years, any new education evaluation reforms need to carefully measure possible consequences against the impact on the increase/decrease of new entrants in the teacher profession (Kraft et al., 2018). This is especially true for schools and subject areas for which it has traditionally been hard to locate and hire quality staff members (Kraft et al., 2018). Another cause of concern in the use of VAMs for evaluation is how to manage the effect of an assignment of a teacher to a challenging group of students versus a teacher that was assigned an easier group of students (Hanushek & Rivkin, 2010). The issue with precise and equitable measurements in VAMs is a legitimate concern when employing VAMs for the determination of job performance and subsequent renewal/nonrenewal (Hanushek & Rivkin, 2010).

Missouri Assessment Program

The Missouri Assessment Program (MAP) was created to evaluate Missouri's school districts' ability to teach students by measuring students' acquisition of knowledge in relation to the framework of knowledge in the Missouri Learning Standards (MODESE, n.d.-b). The current program has been shaped by the landmark education legislation mandates starting in 1993 (MODESE, n.d.-b). Legislation from the 1993 Outstanding Schools Act mandated that Missouri develop a yearly academic evaluation system to test the knowledge of Missouri students of the material in the Sho-Me Standards, which were the education standards used prior to the current Missouri Learning Standards (MODESE, n.d.-b). The first state tests were known as MAP tests and tested students on a grade knowledge for Communication Arts in Grades 3, 7, 10; Mathematics in Grades 4, 8, 10; and Science in Grades 3, 7, 10 (MODESE, n.d.-b). This was the assessment program used until the 2001 No Child Left Behind legislation required states to test students yearly by grade level for Grades 3-8 and once in high school using proficiency levels (MODESE, n.d.-b). Starting in 2008-2009, MODESE implemented EOCs, instead of the high school grade level tests (MODESE, n.d.-b). During that year, EOCs were given for the following courses: Biology I, Algebra I, and English II (MODESE, n.d.-b). The subsequent year, MODESE implemented EOCs for Government, American History, and English I (MODESE, n.d.-b). Student scores on EOCs are published by one of the four achievement levels: Advanced, Proficient, Basic, and Below Basic (MODESE, 2017a). MODESE then assigns a point value for each of the achievement levels to determine a district's MPI which is the MAP Performance Index (MODESE, 2017a). To calculate the MPI, MODESE (2017a) assigns each of the student score levels the following point value: 1 (*Below Basic*), 3 (*Basic*), 4 (*Proficient*) and 5

(*Advanced*). The precipitous drop in score from Basic to Below Basic is to encourage school districts to not have or to move students from the lowest level of Below Basic (MODESE, 2017a). Then the total number of students that scored Advanced is multiplied by 5, total number of students that score Proficient is multiplied by 4, total number of students that scored Basic is multiplied by 3, and the total number of students that scored Below Basic is multiplied by 1 (MODESE, 2017a). Then all four categories are summed and then divided by the total number of students that took the EOC test (MODESE, 2017a). The quotient is then multiplied by 100 to get the MPI (MODESE, 2017a). For example, if a district had 100 total students take an EOC and 25 students each scored in one of the four achievement levels, the MPI for the school would be $325 [(25 * 5) + (25 * 4) + (25 * 3) + (25 * 1)] / 100 * 100 = 325$]. MODESE utilizes the MPI to assess year-to-year improvements on the EOC and to ascertain if districts are meeting the following achievements levels targets set by MODESE for a school district's yearly Annual Performance Report (APR): 2020 Target, On Track, Approaching or lowest level of Floor (MODESE, 2017a).

Summary

This chapter provided a historical review of the use of learning objectives in a teacher's evaluation. From the initial beginnings of just evaluating teacher actions, to the use of incorporating student learning outcomes, then moving the focus of teacher effectiveness to only the actions of effective teaching and now, the current trend, of moving back to incorporating teacher pedagogy skills with student learning data to measure teaching effectiveness, this chapter outlined how the phases of teacher supervision and evaluation have cycled throughout the history of education in the United

States. In addition, a review of the limited research on the use of SLOs in teacher evaluations was provided. The literature that was reviewed was research on SLO use in teacher compensation models. The review was used to build the background and conceptual framework to develop this study on the use of SLOs in a teacher evaluation model. Currently, limited research can be found to review the use of SLOs in a teacher evaluation model that is not associated with teacher compensation. Chapter Three will describe the method for selecting schools that will use SLOs in the evaluation of Biology and Government teachers. The chapter also outlines the instrumentation and how the data will be treated for analysis. Chapter Four of this paper will outline the results of the statistical analysis. Chapter Five will provide the conclusion of the study by providing the professional implications and applications to academic practice of the use of SLOs in teacher evaluations.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

Introduction

The purpose of this causal comparative study was to determine if the MODESE requirement to include student achievement data in teachers' evaluations had a statistical impact on student achievement on the Government and Biology EOC. The researcher had the choice of using SGMs or SLOs to measure student achievement data. The researcher chose SLOs as the measurement tool for student achievement data because MODESE had developed an informational guide titled *Student Learning Objectives Handbook* for districts to use as a resource to use SLOs in teacher evaluations (MODESE, 2013b). As such, the study utilized a 18-question demographic survey to identify Missouri high schools that employed SLOs in teacher evaluations for Biology and Government teachers in the 2014-2015, 2015-2016 and 2016-2017 school years. In addition, principals were asked if they used Biology and Government EOC scores in the teachers' student learning objectives for teacher evaluation purposes and what percentage of teachers met SLO goals for the Biology and Government classes. School administrators were also asked what type (unit growth goal or yearly growth goal) of SLO was used and what percentage of the teacher's summative evaluation final score was based on meeting the student growth goals of the SLO. After surveys were returned, schools were categorized by which year the SLO implementation started, the percentage of teachers that attained SLO growth goals, type of SLO used, and the percentage weight given to the SLO growth data in the teachers' summative evaluation. Using a simple mixed design analysis of variance (ANOVA) statistical test, the researcher analyzed the difference in districts' 2015, 2016, and 2017 district Biology and Government EOC scores based on the use of SLO, type of

SLO used, the percentage of teachers that attained SLO growth goals, and the percentage weight given to the SLO growth data in the teachers' summative evaluation to determine the effect of the SLOs on student achievement.

Participants

A 18-question demographic survey was sent to 545 high school principals in the state of Missouri. As building administrators, principals would have knowledge to answer the demographic questions related to teachers' SLOs and teacher evaluations. The study covered approximately 66,000 Biology I test scores and 60,000 Government test scores per year in the state of Missouri.

In accordance with the guidelines of the Research Review Board (RRB) of Southwest Baptist University regarding the protection of human participants, a request for review was submitted for approval to survey approximately 545 participants for this study. After receiving RRB approval, participant recruitment and collection of data commenced. Participants' consent was gathered by the acceptance to complete the online survey. Complete anonymity for the administration and teachers was provided as the survey did not ask for individual names of teachers. The final tabulation of survey responses and corresponding state test scores was on a ProQuest database and flash drive of the researcher to ensure confidentiality.

Research Questions

The researcher measured the difference in standardized state testing scores of the EOC exams in Government and Biology among the test years of 2015, 2016, and 2017 for school districts that employed SLOs in Biology and Government classes to provide answers to the following research questions:

1. What is the difference in Government and Biology EOC scores when schools included student learning objectives in teacher evaluations?

The researcher measured the difference in 2016 and 2017 Biology and Government EOC scores among school districts that assigned SLO goal data different percentage weights in summative teacher evaluations.

2. What is the difference in Government and Biology EOC scores for schools that assigned SLO growth goal data different percentage weights in the teachers' summative evaluation final score?

The researcher looked at the difference in state test scores for districts whose teachers met their SLOs versus the districts in which teachers did not meet the SLOs to answer the following question:

3. What is the difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes?

The researcher looked at the difference in 2016 and 2017 Biology and Government EOC state test scores for districts whose teachers met used unit goals in their SLOs versus the districts in which teachers used yearly goals in their SLOs to answer the following question:

4. What is the difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes?

The results from the aforementioned questions were used to investigate the following null hypotheses:

1. H_{10} – There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.
2. H_{20} – There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score
3. H_{30} – There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.
4. H_{40} – There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

Instrumentation

The study used a 18-question demographic survey (located in Appendix) that was e-mailed to 544 high schools in Missouri. The survey was developed using ProQuest survey software. The survey was built using logic functions of if/then answers to help ease the participants' experience with the survey. For example, if a participant marked No to the question regarding SLO use for Biology teachers, then the survey jumped to the questions regarding SLO use for Government teachers and did not ask any more questions regarding SLO use for Biology teachers. The survey consisted of four fill-in-the-blank questions, 13 Yes or No multiple-choice questions, and two multiple-choice questions with the following choices: less than 50%, 50% to 74%, 75% to 89%, 90 to

99%, and 100%. The survey was piloted in September 2018 among 18 administrators in Southwest Missouri to determine if the questions were clear and concise, as determined by administrators' comments about the questions. The 18 administrators were selected by the sample of convenience method as the researcher had developed professional, working relationships with the administrators selected for the pilot survey. Out of the administrators selected only 13 returned the survey after two reminder e-mails. A Cronbach's alpha was computed using IBM's Statistical Package for Social Sciences software (SPSS) and had a result of .764. After a review of the responses of the administrators and the Cronbach's alpha, the survey was e-mailed to the remaining 531 participants during the last week of September 2018. A second e-mail reminder was sent the first week of October, with a third reminder e-mail sent the second week of October. The response rate for the survey was 20.9% (114 returned/544 surveys e-mailed) with a full survey completion rate of 19.5% (106 full completion / 544 surveys).

Research Procedure

This causal comparative study utilized a 18-question demographic survey to investigate how Missouri high schools employed student learning objectives in teacher evaluations for Biology and Government teachers in the 2015, 2016, and 2017 school years. The survey was sent to all high school principals in the state of Missouri during the last week of September 2018. A reminder e-mail was sent the first week of October 2018 and then a third and final mass e-mail reminder was sent the second week of October 2018. The analysis of the data began in mid-October 2018 once the study had a minimum level of 20% response rate. At a minimum 20% response rate, the study had a representative sample of the population of 114 participants.

After a minimum of 20% of the surveys was returned, the answers to the demographic survey were downloaded from ProQuest into a Microsoft Excel spreadsheet. Then each participants survey response was coded with 1 (*Yes*) and 2 (*No*) for each Yes/No question. The two questions regarding percentages met were coded with a 5 (100%), 4 (90% to 99%), 3 (75% to 89%), 2 (50% to 74%) and 1 (< 50%). The only exception was the answer to the percent of evaluation based on SLOs, which was entered as the numerical answer. The responses were checked twice for accuracy before saving the information.

Then the 2015, 2016, and 2017 Biology and Government EOC scores from the districts that returned the survey were collected from the Missouri Comprehensive Data System and placed in the Microsoft Excel spreadsheet. Student scores on EOCs were published by one of the four achievement levels: Advanced, Proficient, Basic, and Below Basic (MODESE, 2017a). MODESE then assigned a point value for each of the achievement levels to determine a district's MPI (MODESE, 2017a). To calculate the MPI, DESE assigned each of the student score levels the following point value: Advanced-5, Proficient-4, Basic-3, and Below Basic-1 (MODESE, 2017a). Then the total number of students that scored Advance was multiplied by 5, total number of students that score Proficient was multiplied by 4, total number of students that scored Basic was multiplied by 3, and the total number of students that scored Below Basic was multiplied by 1 (MODESE, 2017a). Then all four categories were summed and then divided by the total number of students that took the EOC test (MODESE, 2017a). The quotient was then multiplied by 100 to get the MPI (MODESE, 2017a). The MPI score was collected for the test results because this score indicated student results for all categories of Below

Basic, Basic, Proficient, and Advanced. The categories were determined by MODESE to measure the Adequate Yearly Progress of school districts in Missouri. Utilizing the MPI score instead of just focusing on the percent of students in top two categories showed a comprehensive view of student achievement for the whole class. The researcher imputed any missing data for districts that did not have a complete MPI score due to missing achievement public data due to the MODESE reporting procedure of not allowing a district to have an achievement area to show a 0%. MODESE would mark the achievement data category with an asterisk on the public reporting data if an achievement category had a 0%. The researcher imputed the missing data for a district by looking at the total number of reported students that took a test and subtracting the publicly reported students for the other achievement data category that was reported and then estimating which category the missing data would be associated with by looking at the scale score. These data were placed into a Microsoft spreadsheet.

The study used scores from the Biology I and Government EOC exams that had been validated by MODESE. The reasoning for using Biology and Government EOC exams was these subject tests were scaled and equated to previous years tests. New Math and ELA EOC tests forms were introduced during the school years selected for the study, so a comparison of Math and ELA test scores among the selected years could not be used in the study because the new tests were not scaled and equated to previous years' tests. For the 2015, 2016, and 2017 school year Government and Biology EOC tests, MODESE used operational test forms that were considered homogenous in subject areas by evidence of statistical attributes (MODESE, 2017b). The Government and Biology EOC tests were scaled year-to-year by using postequation using a raw scale to scale score

(RSS) tables developed in 2008 and 2009 (MODESE, 2017b). It should be noted that due to a Biology EOC form difference in 2016, MODESE recalibrated the RSS tables using a combination of 2014, 2015, and 2016 test data and rescored the 2016 test data and used the same RSS data for 2017 test, so the tests would be scaled year to year (MODESE, 2017b). MODESE used Winstep software to determine a scale score by evaluating students' answers to selected response and performance event EOC questions (MODESE, 2017b). EOC performance events and selected responses questions were adjusted to a scale standard by using the "Masters partial credit model" (MODESE, 2017b, p. 69) and the Rasch scaling model, respectively.

Data were analyzed using the SPSS statistical software program. The research questions and whether to accept or reject the null hypotheses were answered by completing a quantitative ANOVA test of the MPI scores. The simple mixed design ANOVA test was used to measure the variance between subjects, the variance within the subjects, and then the effect of the interaction of between and within-subjects (Gamst, Meyers, & Guarino, 2008). The individual differences of subjects can be measured by using the variance of between subjects in a within-subjects design (Gamst et al, 2008). The change of a score based on previous exposure to a treatment or item related to the treatment is called a carryover effect (Gamst et al., 2008). Carryover effect is what a researcher wants to measure when using a within-subjects design (Gamst et al., 2008). Gamst et al., (2008) stated that as long as a pretest does not change the posttest results and/or a natural event did not happen during the same treatment period and the study could not measure the control condition, then the pretest /posttest design may be used versus no study at all (Gamst et al., 2008).

To measure the impact of SLOs in teacher evaluations, the study analyzed the difference in 2015 and 2016 Biology and Government EOC scores to determine the impact of the SLOs on student achievement using a simple mixed design ANOVA. The study also analyzed the difference in 2016 and 2017 Biology and Government EOC scores to determine the impact of the SLO on student achievement using a simple mixed design ANOVA. The simple mixed design ANOVA was used to determine the impact of teachers that met their SLOs on student achievement by comparing the difference in 2016 and 2017 test scores of districts whose teachers met the SLO and those districts whose teachers did not meet the SLO goal. Also, the simple mixed design ANOVA was calculated to determine the statistical difference between 2016 and 2017 EOC scores for districts that gave SLO growth data different percentage weights in a teacher's final summative evaluation. The researcher also conducted an ANOVA to measure the variance in EOC scores between districts that used unit or yearly growth goals in their SLOs.

Summary

Chapter Three presented the framework for the purpose of this simple mixed design study to examine the impact of the use of SLOs in teachers' evaluation on student academic achievement. In addition, the chapter outlined the procedures for the quantitative analysis of the four guiding research questions and the null hypotheses. In this chapter the justification of the study design was given as well as the detailed explanation of the survey and the piloting of the survey. Chapter Four contains the survey results, quantitative analysis, and the presentation of the data results of this study. The results are to be explained and presented in the context of each of the four guiding

research questions and null hypotheses. Chapter Five will provide the conclusion of the study by providing the professional implications and applications to academic practice of the use of SLOs in teacher evaluations.

CHAPTER FOUR

ANALYSIS OF DATA

Introduction

The purpose of this causal comparative study was to examine the impact of the use of SLOs in teachers' evaluation on student academic achievement. In addition, the study researched whether the type of SLO had a statistical impact on student academic achievement. The study used the 2015, 2016, and 2017 MAP Index scores from MODESE's Biology and Government EOC tests to answer the following research questions and null hypotheses:

Research Question 1: What is the difference in Government and Biology EOC scores when schools included student learning objectives in teacher evaluations?

Null Hypothesis 1 (H_{10}): There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.

Research Question 2: What is the difference in Government and Biology EOC scores for schools that assigned meeting SLO goals data different percentage weights in the teachers' summative evaluation final score?

Null Hypothesis 2 (H_{20}): There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score

Research Question 3: What is the difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes?

Null Hypothesis 3 (H3₀): There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.

Research Question 4: What is the difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes?

Null Hypothesis 4 (H4₀): There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

The research questions and whether to accept or reject the null hypothesis were answered by completing an ANOVA test of the MPI scores. The simple mixed design ANOVA test was used to measure the variance between subjects, the variance within the subjects, and then the effect of the interaction of between and within subjects (Gamst et al., 2008). The rest of this chapter will discuss the demographics of samples, the response rate to the survey, analysis of survey responses, and the results of the simple mixed design ANOVA test for each of the research questions and null hypotheses.

Analysis of Data

A 18-question demographic survey was e-mailed three separate times over a 3-week period to 544 high school principals in the state of Missouri. High school principals were selected because as building administrators, principals had sufficient knowledge to answer questions related to teachers' SLOs and teacher evaluations. Of the 544 surveys that were e-mailed, 203 participants viewed the survey, 165 participants started the survey, and 106 answered every survey question for a 19.4 % completion rate.

In response to the question regarding SLO use for Biology teachers, 117 participants answered with 90 responses (76.9%) answering “Yes” to using SLOs in 2014-2015, 2015-2016, and 2016-2017. Out of the 90 “Yes” responses, 89 participants provided what type of SLO was used in 2016-2017 with 58 responses (65.2%) using unit growth goal, 26 responses (29.2%) using yearly growth goal, and five responses (5.6%) using other types. In response to the question regarding the percentage of Biology teachers that met predetermined SLO goals, 55 participants answered with 54.6% reporting 100% met, 18.2% reporting 90% to 99% met, 12.7% reporting 75% to 89% met, 10.9% reporting 50% to 74% met, and 3.6% reporting less than 50% met. The survey results regarding what year SLOs were used (denoted as “Yes” response) in Biology teacher evaluations were as follows: 26.51% (22 out of 83 responses) in 2014-2015, 48.19% (40 out of 83 responses) in 2015-2016, and 65.17% (58 out of 89 responses) in 2016-2017.

In response to the question regarding SLO use for Government teachers, 108 participants answered with 83 responses (76.85%) answering “Yes” to using SLOs in 2014-2015, 2015-2016, and 2016-2017. When asked the type of SLO was used in 2016-2017, 83 responded with 51 responses (61.5%) using unit growth goal, 29 responses (34.9%) using yearly growth goal, and three responses (3.6%) using other types. In response to the question regarding the percentage of Government teachers that met predetermined SLO goals, 52 participants answered with 61.5% reporting 100% met, 9.6% reporting 90% to 99% met, 19.3% reporting 75% to 89% met, 7.7% reporting 50% to 74% met, and 1.9% reporting less than 50% met. The survey results regarding if SLOs were used (denoted as “Yes” response) in Government teacher evaluations in the school

years of 2014-2015, 2015-2016, and 2016-2017 were as follows: 22.9% (19 out of 83), 43.4% (36 out of 83), and 62.7% (52 out of 83), respectively.

The quantitative analysis of the four research questions were answered by completing six separate ANOVA tests of the Biology and Government EOC MPI scores. The results of the ANOVA tests were used to accept or reject the four null hypotheses. For ease of reading, the data analysis for each of the four research questions and four null hypotheses will be presented as four separate subsections under the section of data presentation.

Data Presentation

Research Question 1. Three simple mixed design ANOVA tests were conducted on the variance of Biology and Government EOC scores between schools that used SLOs in teachers' evaluations to answer the following research question and accept or reject the following null hypothesis:

Research Question 1: What is the difference in Government and Biology EOC scores when schools included student learning objectives in teacher evaluations?

Null Hypothesis 1 (H_{10}): There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.

A simple mixed design ANOVA test was conducted to measure the variance difference between Biology and Government EOC scores of schools that used, for the first time, SLOs in teacher evaluations in 2015-2016 versus schools that did not use SLOs in teacher evaluations. The researcher established $p < .05$ as the significance level. Based on the repeated measure ANOVA results, there was no statistically significant difference in the Biology EOC scores between the schools that used SLOs in teacher evaluations in

2015-2016 and the schools that did not use the SLOs in teacher evaluations in 2015-2016 since $F(1, 78) = .044$, $p = .835$. There was a statistically significant difference in each group's score from 2014-2015 Biology EOC to 2015-16 Biology EOC since $p = .006$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC scores between the schools that used SLOs in teacher evaluation in 2015-2016 and the schools that did not use the SLOs in teacher evaluation in 2015-2016 since $F(1, 73) = .300$, $p = .585$. Also, there was no statistically significant difference in each group's score from 2014-2015 Government EOC to 2015-2016 Government EOC since $p = .324$. The results are in Table 1.

Table 1

ANOVA for First Time Use of SLO (Yes versus No) in Teacher Evaluations in 2015-2016

Source		Sum of squares	df	Mean of squares	F	p	η_p^2
Biology EOC	Between Groups	49.002	1	49.002	.044	.835	.001
	Error	87668.043	78	1123.949			
	Within Groups	1815.943	1	1815.943	7.828	.006*	
	Error	18093.359	78	231.966			
Government EOC	Between Groups	335.937	1	335.937	.300	.585	.004
	Error	81668.303	73	1118.744			
	Within Groups	462.547	1	462.547	.984	.324	
	Error	34315.426	73	470.074			

* $p < .05$.

A simple mixed design ANOVA test was conducted to measure the variance difference between Biology and Government EOC scores of schools that used, for the first time, SLOs in teacher evaluations in 2016-2017 versus schools that did not use SLOs in teacher evaluations. The researcher established $p < .05$ as the significance level. Based on the repeated measure ANOVA results, there was no statistically significant difference

in the Biology EOC scores between the schools that used SLOs in teacher evaluations in 2016-2017 and the schools that did not use the SLOs in teacher evaluations in 2016-2017 since $F(1, 62) = 1.066, p = .306$. Also, there was no statistically significant difference in each group's score from 2015-16 Biology EOC to 2016-2017 Biology EOC since $p = .484$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC scores between the schools that used SLOs in teacher evaluations in 2016-2017 and the schools that did not use the SLOs in teacher evaluations in 2016-2017 since $F(1, 58) = 2.433, p = .124$. Also, there was no statistically significant difference in each group's score from 2015-2016 Government EOC to 2016-2017 Government EOC since $p = .376$. The results are in Table 2.

Table 2

ANOVA for First Time Use of SLO (Yes versus No) in Teacher Evaluations in 2016-2017

Source		Sum of squares	df	Mean of squares	F	p	η_p^2
Biology EOC	Between Groups	1413.961	1	1413.961	1.066	.306	.017
	Error	82213.952	62	1326.031			
	Within Groups	193.886	1	193.886	.496	.484	
	Error	24223.846	62	390.707			
Government EOC	Between Groups	2844.336	1	2844.336	2.433	.124	.040
	Error	67818.489	58	1169.284			
	Within Groups	365.406	1	365.406	.794	.376	
	Error	26682.085	58	460.036			

* $p < .05$.

Since the data suggested no statistical difference in the first year that SLOs were used in teacher evaluations, the researcher ran a simple mixed design ANOVA test to

measure the variance difference between Biology and Government EOC scores of schools that used SLOs in 2016-2017 teacher evaluations, no matter if started in 2014-2015, 2015-2016, or 2016-2017, versus schools that did not use SLOs in teacher evaluations. The researcher established $p < .05$ as the significance level. Based on the repeated measure ANOVA results, there was no statistically significant difference in the Biology EOC scores between schools that used SLOs in teacher evaluation in 2016-2017, no matter when the school first used SLO in teacher evaluations, and the schools that did not use the SLOs in teacher evaluation in 2016-17 since $F(1, 98) = .213, p = .645$. It should be noted there was a statistically significant difference in each group's score from 2015-2016 Biology EOC to 2016-17 Biology EOC since $p = .008$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC between schools that used SLOs in teacher evaluation in 2016-2017, no matter when the school first used SLO in teacher evaluations, and the schools that did not use the SLOs in teacher evaluation in 2016-2017 since $F(1, 90) = .017, p = .897$. It should be noted there was a statistically significant difference in each group's score from 2015-2016 Government EOC to 2016-2017 Government EOC since $p = .016$. The results are in Table 3.

Table 3

ANOVA for SLO Use (Yes versus No) in 2016-2017 Teacher Evaluations

Source		Sum of squares	df	Mean of squares	F	p	η_p^2
Biology EOC	Between Groups	298.468	1	298.468	.213	.645	.002
	Error	137303.819	98	1401.059			
	Within Groups	2632.204	1	2632.204	7.269	.008*	
	Error	35375.077	98	360.97			
Government EOC	Between Groups	20.223	1	20.223	.017	.897	.000
	Error	108407.707	90	1204.530			
	Within Groups	2136.571	1	2136.571	5.979	.016*	
	Error	34315.426	73	470.074			

* $p < .05$.

The data of the three ANOVA tests suggested the null hypothesis of H_1 (No difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations) could not be rejected, therefore the null hypothesis was accepted.

Research Question 2. A simple mixed design ANOVA tests was conducted on the variance of Biology and Government EOC scores between schools that assigned SLO data different percentage weights in the teachers' summative evaluation final score to answer the following research question and accept or reject the following null hypothesis:

Research Question 2: What is the difference in Government and Biology EOC scores for schools that assigned meeting SLO growth goals different percentage weights in the teachers' summative evaluation final score?

Null Hypothesis 2 (H_{20}): There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score

The schools that used SLOs in teacher evaluations were grouped by the percentage weight given to SLO growth goals in a teachers' summative evaluation final score using the following five categories: 0%, 1% to 10%, 11% to 19%, 20%-29%, and 30% and greater. The ANOVA test was conducted to determine the difference in variance in Biology and Government EOC scores among the five categories. The researcher established $p < .05$ as the significance level. Based on the repeated measure ANOVA results, there was no statistically significant difference in the 2016-2017 Biology EOC scores between schools that assigned SLO data different percentage weights in the teachers' summative evaluation final score since $F(4, 36) = 1.916, p = .129$. However, a partial eta square effect size of $\eta_p^2 = .176$ was calculated, suggesting that the percent weight given to SLO had a large effect (Watson, 2018) on the variance of the 2016-2017 Biology EOC scores. Also, there was no statistically significant difference in each group's score from 2015-2016 Biology EOC to 2016-2017 Biology EOC since $p = .139$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC scores between schools that assigned meeting SLO goals data different percentage weights in the teachers' summative evaluation final score since $F(4, 36) = .870, p = .491$. However, a partial eta square effect size of $\eta_p^2 = .088$ was calculated, suggesting that the percent weight given to SLO had a medium effect (Watson, 2018) on the variance of the 2016-2017 Government EOC scores. Also, there was no statistically significant difference in each group's score from 2015-2016 Government EOC to 2016-2017 Government EOC since $p = .233$. The results are in Table 4.

Table 4

ANOVA for Percentage Weight of Teachers' Summative Evaluation Tied to SLO Growth Data for 2016-2017

Source		Sum of square	df	Mean of square	F	p	η_p^2
Biology EOC	Between Groups	10168.174	4	2542.043	1.916	.129	.176
	Error	47757.965	36	1326.610			
	Within Groups	928.053	1	928.053	2.293	.139	
	Error	14567.580	36	404.655			
Government EOC	Between Groups	4201.447	4	1050.362	.870	.491	.088
	Error	43475.675	36	1207.658			
	Within Groups	240.116	1	240.116	1.471	.233	
	Error	5877.42	36	163.262			

* $p < .05$.

The data of the ANOVA test suggested the null hypothesis of H_{20} (There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score) could not be rejected, therefore the null hypothesis was accepted.

Research Question 3. A mixed design ANOVA tests was conducted on the variance of Biology and Government EOC scores for schools whose teachers met student learning goals in Biology and Government classes to answer the following research question and accept or reject the following null hypothesis:

Research Question 3: What is the difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes?

Null Hypothesis 3 (H3₀): There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.

The schools that used SLOs in teacher evaluations were grouped by percentage of teachers who met SLO goals into the following five categories: 100% met, 90% to 99% met, 75% to 89% met, 50% to 74% met, and less than 50% met. The ANOVA test was conducted to determine the difference in variance in 2015-2016 and 2016-2017 Biology and Government EOC scores among schools based on the percentage of their teachers that met SLO goals by grouping the schools in one of the five aforementioned categories. The researcher established $p < .05$ as the significance level.

Based on the repeated measure ANOVA results, there was no statistically significant difference in the 2016-2017 Biology EOC scores among schools that had different percentage of teachers who met predetermined SLO goals since $F(4, 45) = .780$, $p = .544$. However, a partial eta square effect size of $\eta_p^2 = .065$ was calculated, suggesting that the percentage of teachers meeting SLO goals had a medium effect (Watson, 2018) on the variance of the 2016-2017 Biology EOC scores. Also, there was no statistically significant difference in each group's score from 2015-2016 Biology EOC to 2016-2017 Biology EOC since $p = .085$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC scores among schools that had different percentage of teachers who met predetermined SLO goals since $F(4, 41) = 1.715$, $p = .165$. However, a partial eta square effect size of $\eta_p^2 = .143$ was calculated, suggesting that the percent weight given to SLO had a large effect (Watson, 2018) on the variance of the 2016-2017 Government EOC scores. Also,

there was no statistically significant difference in each group's score from 2015-2016 Government EOC to 2016-2017 Government EOC since $p = .071$. The results are in Table 5.

Table 5

ANOVA for Teachers That Met SLO Goals in 2016-2017

Source		Sum of squares	df	Mean of squares	F	p	η_p^2
Biology EOC	Between Groups	4224.676	4	1056.169	.780	.544	.065
	Error	60941.877	45	1354.264			
	Within Groups	1227.278	1	1227.278	3.109	.085	
	Error	17763.962	45	394.755			
Government EOC	Between Groups	7355.569	4	1838.892	1.715	.165	.143
	Error	43969.398	41	1072.424			
	Within Groups	633.016	1	633.016	3.439	.071	
	Error	7545.789	41	184.044			

* $p < .05$.

The data of the ANOVA test suggested the null hypothesis of H_{30} (There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.) could not be rejected, therefore the null hypothesis was accepted.

Research Question 4. A mixed design ANOVA tests was conducted on the variance of Biology and Government EOC scores for schools whose teachers met student learning goals in Biology and Government classes to answer the following research question and accept or reject the following null hypothesis:

Research Question 4: What is the difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes?

Null Hypothesis 4 (H4₀): There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

The schools that used SLOs in teacher evaluations were grouped by type of SLO used into the following three categories: unit SLO Goal, yearly SLO goal, and other. The ANOVA test was conducted to determine the difference in variance in 2015-2016 and 2016-2017 Biology and Government EOC scores among schools based on the type of SLO goal used in the teacher evaluation by grouping the schools in one of the three aforementioned categories. The researcher established $p < .05$ as the significance level.

Based on the repeated measure ANOVA results, there was no statistically significant difference in the Biology EOC scores among schools that used different types of SLOs in teacher evaluation in 2016-2017 since $F(2, 49) = 1.449, p = .245$. However, a partial eta square effect size of $\eta_p^2 = .056$ was calculated, suggesting the type of SLO used in teacher evaluation had a medium effect (Watson, 2018) on the variance of the 2016-2017 Biology EOC scores. Also, there was no statistically significant difference in each group's score from 2015-2016 Biology EOC to 2016-2017 Biology EOC since $p = .213$. In addition, based on the repeated measure ANOVA results, there was no statistically significant difference in the Government EOC scores among schools that used different types of SLOs in teacher evaluation in 2016-2017 since $F(4, 43) = .965, p = .389$. However, a partial eta square effect size of $\eta_p^2 = .043$ was calculated, suggesting the type of SLO used in teacher evaluation had a small effect (Watson, 2018) on the variance of the 2016-2017 Government EOC scores. Also, there was no statistically

significant difference in each group's score from 2015-2016 Government EOC to 2016-2017 Government EOC since $p = .141$. The results are in Table 6.

Table 6

ANOVA Type of SLOs Used in Teachers Evaluation in 2016-2017

Source		Sum of squares	df	Mean of squares	F	p	η_p^2
Biology EOC	Between Groups	3647.324	2	1823.662	1.449	.245	.056
	Error	61661.839	49	1258.405			
	Within Groups	597.739	1	597.739	1.589	.213	
	Error	18433.413	49	376.192			
Government EOC	Between Groups	2203.623	4	1101.812	.965	.389	.043
	Error	49121.344	43	1142.357			
	Within Groups	401.433	1	401.433	2.251	.141	
	Error	7666.788	43	178.297			

* $p < .05$.

The data of the ANOVA test suggested the null hypothesis H_{40} (There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes) could not be rejected, therefore the null hypothesis was accepted.

Summary of Findings

This section summarizes the results of the ANOVA tests for the four research questions and four null hypotheses of the study. The results will be presented by research question and the associated null hypothesis.

Research Question 1. What is the difference in Government and Biology EOC scores when schools included student learning objectives in teacher evaluations?

Upon reviewing the results of three ANOVA tests, the researcher determined there were no statistically significant differences in Biology and Government EOC scores

between schools that used SLOs in teacher evaluations and schools that did not use SLOs in teacher evaluations.

Null Hypothesis 1 (H₁₀). There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.

There were no statistically significant differences in Biology and Government EOC scores between schools (participating in this study) that used SLOs in teacher evaluations and schools that did not use SLOs in teacher evaluations. The data of the three ANOVA tests suggested the null hypothesis H₁₀ could not be rejected, therefore the results failed to reject the null hypothesis.

Research Question 2. What is the difference in Government and Biology EOC scores for schools that assigned meeting SLO goals data different percentage weights in the teachers' summative evaluation final score?

Based on the repeated measure ANOVA results, the researcher determined there was not a statistically significant difference in the 2016-17 Biology and Government EOC scores between schools (participating in the study) that assigned SLO data different percentage weights in the teachers' summative evaluation final score. However, the researcher determined that based on a partial eta square effect size of $\eta_p^2 = .176$, the data suggested the percent weight given to SLOs had a large effect on the variance of the 2016-2017 Biology EOC scores. The researcher also determined, based on a partial eta square effect size of $\eta_p^2 = .088$, the data suggested the percent weight given to SLOs had a medium effect on the variance of the 2016-2017 Government EOC scores.

Null Hypothesis 2 (H₂₀). There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score.

There was not a statistically significant difference in the 2016-17 Biology and Government EOC scores between schools (participating in the study) that assigned SLO data different percentage weights in the teachers' summative evaluation final score. The data of the ANOVA test suggested the null hypothesis of H₂₀ could not be rejected, therefore the results failed to reject the null hypothesis.

Research Question 3. What is the difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes?

Based on the repeated measure ANOVA results, the researcher determined there was not a statistically significant difference in the 2016-2017 Biology and Government EOC scores based on the different percentage of teachers who met predetermined SLO goals among schools that participated in the study. The researcher also determined, based on a partial eta square effect size of $\eta_p^2 = .065$, the data suggested the percentage of teachers meeting SLO goals had a medium effect on the variance of the 2016-2017 Biology EOC scores. This was evident as the schools with less than 74% of teachers that met SLO goals had a lower mean score ($\mu=354$) than schools with 75% or higher meeting SLO goals ($\mu=372$). The researcher determined, based on a partial eta square effect size of $\eta_p^2 = .143$, the data suggested that percentage of teachers meeting SLO goals had a large effect on the variance of the 2016-2017 Government EOC scores. This was evident

as the schools with less than 74% of teachers that met SLO goals had a lower mean score ($\mu=359$) than schools with 75% or higher meeting SLO goals ($\mu=381$).

Null Hypothesis 3 (H3₀). There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.

There was no statistically significant difference in Government and Biology EOC scores among schools participating in this study based on whether the teacher met SLO learning goals for Biology and Government classes. The data of the ANOVA test suggested the null hypothesis of H3₀ could not be rejected, therefore the results failed to reject the null hypothesis.

Research Question 4. What is the difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes?

Based on the repeated measure ANOVA results, the researcher determined there was no statistically significant difference in the Biology and Government EOC scores among schools that used different type of SLOs in teacher evaluations in 2016-2017. However, the researcher determined, based on a partial eta square effect size of $\eta_p^2 = .065$, the data suggested the type of SLO (with schools using yearly growth goals recording a higher mean than schools with unit growth goals) used in teacher evaluations had a medium effect on the variance of the 2016-2017 Biology EOC scores.

Null Hypothesis 4 (H4₀). There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

There was no significant statically difference in Government and Biology EOC scores among schools participating in the study due to type of SLO (yearly or unit goal) used in Biology and Government classes. The data of the ANOVA suggested the null hypothesis H_4 could not be rejected, therefore the results failed to reject the null hypothesis.

Summary

Chapter Four contained the survey results, quantitative analysis, and the presentation of the data results of this study. The results were explained and presented in the context of each of the four guiding research questions and null hypotheses. Chapter Five will present the researcher's conclusion, applications to current educational practices, and recommendations for future academic research surveys.

CHAPTER FIVE

CONCLUSION

Introduction

On December 10, 2015, the ESSA was signed into law by President Barack Obama (USDOE, n.d.). The ESSA replaced the original requirements of the USDOE requiring state departments to use student growth data as a significant contributing factor in teacher evaluations (USDOE, n.d.). Though federal education requirements on the components of teacher evaluations shifted to state control, MODESE still required districts to follow the Missouri Principles of Effective Evaluation, which utilized student growth data for evaluation purposes (MODESE 2016). The purpose of this causal comparative study was to determine if the MODESE requirement to include student achievement data in teachers' evaluations had a statistical impact on student achievement on the Government and Biology EOC. Implementation of SLOs in teacher evaluation is a current trend in education; therefore, the research on the use of SLOs in non-value-added teacher evaluation models was limited (Lachlan-Hache, 2015; Longchamp, 2017). Researchers have defined SLOs as an educational practice tool that promotes self-reflection by utilizing the following steps: review standards, develop core concepts, set goals of student achievement on the core concept, monitor progress, and review the outcome (Lachlan-Hache et al., 2012). This study added further evidence to the body of research regarding SLOs' impact on student achievement through the mandated use of student growth measures by MODESE in teacher evaluations, instead of voluntary local decisions to employ student growth measures to enhance student achievement. The study utilized a 18-question demographic survey to identify Missouri high schools that employed SLOs in teacher evaluations for Biology and Government teachers in the 2014-

2015, 2015-2016, and 2016-2017 school years. The study used an ANOVA test on the 2015, 2016, and 2017 MPI scores from MODESE's Biology and Government EOC tests to investigate the study's four research questions and decide whether to accept or reject the associated four null hypotheses.

Research Questions and Null Hypotheses

Research Question #1 asked if there was a difference in Government and Biology EOC scores when schools included student learning objectives in teachers' evaluations. Three ANOVA tests were conducted to measure the variances in Government and Biology EOC scores between schools that used SLOs for teacher evaluation purposes in 2014-2015, 2015-2016, and 2016-2017 versus schools that did not use SLOs for teacher evaluation purposes. The ANOVA tests indicated there were no statistically significant differences in Biology and Government EOC scores between schools that used SLOs in teacher evaluations and schools that did not use SLOs in teacher evaluations. It is interesting to recognize that even though no statistical difference was noted, the schools that used SLOs had a higher mean score for 2016 and 2017 Biology EOCs (Mean= 379.8 and 371.6) versus schools that did not use SLOs in 2016 and 2017 (Mean= 376.3 and 370.1) for Biology EOCs in 2016 and 2017. The ANOVA results were $p = .645$ and $p = .897$ for the difference in Biology and Government EOC scores, respectively, between schools that used SLOs in teachers' evaluations versus schools that did not use SLOs in teacher evaluations. Since $p > .05$, the researcher failed to reject the following H_0 null hypothesis: There will be no difference in Government and Biology EOC scores due to the inclusion of SLOs in teacher evaluations.

Research Question 2 asked if there was a difference in Government and Biology EOC scores for schools that assigned meeting SLO goals data different percentage weights in the teachers' summative evaluation final score. Based on the repeated measures ANOVA results to answer Research Question #2, the researcher determined there was not a statistically significant difference in the 2016-2017 Biology and Government EOC scores between schools that assigned SLO data different percentage weights in the teachers' summative evaluation final score, since $F(4, 36) = 1.916, p = .129$ and $F(4, 36) = .870, p = .491$, respectively. However, the researcher determined that based on a partial eta square effect size of $\eta_p^2 = .176$, the data suggested that the percent weight given to SLOs had a large effect on the variance of the 2016-2017 Biology EOC scores. This could help explain the observation that the schools (μ_1) that did not give SLO data a percent weight (0% group) in teachers' summative evaluations had the lowest MAP Index mean score for Biology EOCs as compared to the other four categories (1% to 10%, 11% to 19%, 20% to 29%, and 30% and greater) for 2015-2016 ($\mu_1=374$ vs. $\mu_{2,3,4,5}=383$) and 2016-2017 ($\mu_1=365$ vs. $\mu_{2,3,4,5}=375$). The researcher also determined, based on a partial eta square effect size of $\eta_p^2 = .088$, the data suggested the percent weight given to SLOs had a medium effect on the variance of the 2016-2017 Government EOC scores. This could help explain the observation that the schools (μ_1) that did not give SLO data a percent weight (0% group) in teachers' summative evaluation had the lowest MAP Index mean score for Government EOCs as compared to the other four categories (1% to 10%, 11% to 19%, 20% to 29%, and 30% and greater) for 2015-2016 ($\mu_1=357$ vs. $\mu_{2,3,4,5}=369$) and 2016-17 ($\mu_1=357$ vs. $\mu_{2,3,4,5}=377$). In addition, the average of the three percentage categories starting with 10% and higher had a higher MAP Index

score mean than the two lowest categories for Government and Biology EOCs ($\mu_{3,4,5}=376$ vs. $\mu_{1,2}=369$ and $\mu_{3,4,5}=376$ vs. $\mu_{1,2}=370$, respectively). This observational data combined with suggested medium and large effect of the percentage weight given to the SLO data in teachers' evaluations on the variance of Biology and Government EOC scores might suggest that schools consider using a percentage weight of 10% or greater when using SLO data in teachers' summative evaluations. Since $p > .05$, the researcher failed to reject the following H_0 null hypothesis: There will be no difference in Government and Biology EOC scores due to different percentage weights given to SLO data in the teachers' summative evaluation final score.

Research Question 3 asked if there was a difference in Government and Biology EOC scores for schools whose teachers met student learning goals in Biology and Government classes. Analyzing the repeated measure ANOVA results from Research Question 3, the researcher determined there was not a statistically significant difference in the 2016-2017 Biology and Government EOC scores based on the different percentage of teachers who met predetermined SLO goals among schools, since $F(4, 45) = .780, p = .544$ and $F(4, 41) = 1.715, p = .165$, respectively. The researcher also determined, based on a partial eta square effect size of $\eta_p^2 = .065$, the data suggested that the percentage of teachers meeting SLO goals had a medium effect on the variance of the 2016-2017 Biology EOC scores. This was evident as the schools with less than 74% of teachers that met SLO goals had a lower mean score ($\mu = 354$) than schools with 75% or higher meeting SLO goals ($\mu = 372$). The researcher determined, based on a partial eta square effect size of $\eta_p^2 = .143$, the data suggested that percentage of teachers meeting SLO goals had a large effect on the variance of the 2016-2017 Government EOC scores. This was

evident as the schools with less than 74% of teachers that met SLO goals had a lower mean score ($\mu = 355$) than schools with 75% or higher meeting SLO goals ($\mu = 375$). This observational data combined with suggested medium and large effect of the percentage of teachers that met SLO goals on the variance of Biology and Government EOC scores might suggest that schools consider developing high-quality professional development and give importance to the use of SLOs for teaching and evaluation purposes. Since $p > .05$, the researcher failed to reject the following H_{30} null hypothesis: There will be no difference in Government and Biology EOC scores based on whether the teacher met SLO learning goal for Biology and Government classes.

Research Question 4 asked if there was a difference in Government and Biology EOC scores for schools based on type/method of student learning objective/student growth measure (yearly goal or unit goal) used in Biology and Government classes. Using the repeated measure ANOVA results, the researcher determined there was no statistically significant difference in the Biology EOC scores, $F(2, 49) = 1.449, p = .245$, and Government EOC scores, $F(4, 43) = .965, p = .389$. However, a partial eta square effect size of $\eta_p^2 = .043$ was calculated, suggesting the type of SLO used in teacher evaluation (yearly goal $\mu = 372$ and unit goal $\mu = 374$) had a small effect on the variance of the 2016-2017 Government EOC scores among schools that used different type of SLOs in teacher evaluation in 2016-2017. In addition, the researcher determined, based on a partial eta square effect size of $\eta_p^2 = .065$, the data suggested the type of SLO (with schools using yearly growth goals having a higher $\mu = 385$ than schools with unit growth goals $\mu = 368$) used in teacher evaluation had a medium effect on the variance of the 2016-2017 Biology EOC scores. Since $p > .05$, the researcher failed to reject the

following H₀ null hypothesis: There will be no difference in Government and Biology EOC scores due to type of SLO (yearly or unit goal) used in Biology and Government classes.

Professional Implications/Application to Academic Practice

The results of the study suggested that SLOs' use in teacher evaluations did not have a statistically significant impact on Missouri Biology and Government EOC scores. The results of the study concur with earlier research studies, which suggested that even though SLOs had an effect, the differences were not statistically significant (Balch & Springer, 2015; Briggs et al., 2014; Schmitt, 2011). Even though a statistically significant difference was not found, the results of the study indicate that meeting SLO goals, percentage weight given to SLO data in teacher evaluations, and the type of SLO goal had an effect on the variance of Biology and Government EOC scores. The suggested effect size could have implications for administrators and policy makers to consider regarding the use of SLOs in teacher evaluations.

The study suggests that schools that had 75% of teachers meet their SLO goals for their Biology and Government classes had a higher mean score on the Biology and Government EOC scores than schools that had less than 75% meet their SLO goal. The study indicated the schools with less than 74% of teachers meeting SLO goals had a lower mean score ($\mu = 354$) than schools with 75% or higher meeting SLO goals ($\mu = 372$) on the 2016-2017 Biology EOC test. The difference in scores, even though not statistically significant, did have a partial eta square effect size of $\eta_p^2 = .065$, which is considered a medium effect on the variance of the 2016-2017 Biology EOC scores. The study also indicated the schools with less than 74% of teachers meeting SLO goals had a

lower mean score ($\mu = 355$) than schools with 75% or higher meeting SLO goals ($\mu = 375$) on the 2016-2017 Government EOC. The difference in scores, even though not statistically significant, did have a partial eta square effect size of $\eta_p^2 = .143$, which is considered a large effect on the variance of the 2016-2017 Government EOC scores. The effect size results support the results of early research studies (Schmitt, 2011; Schmitt et al., 2009) indicating teachers who meet SLO goals have students that score higher on standardized tests, although the differences were not statistically significant. This observational data combined with suggested medium and large effect of the percentage of teachers that met SLO goals on the variance of Biology and Government EOC scores might suggest schools consider developing high-quality professional development and give importance to the use of SLOs for teaching and evaluation purposes. This recommendation is in line with previous research results suggesting a positive perception and a willingness of teachers/administrators to implement SLO procedures has led to increased use of SLOs by teachers, which had a corresponding positive effect on student achievement (Schmitt, 2014; Slotnik et al., 2015).

The findings of the study indicate that percent weight given to SLO data in teacher evaluations had an effect on the variance in Biology and Government EOC scores. The effect size on Biology EOC was a partial eta square effect size of $\eta_p^2 = .176$ and the effect size on Government EOC was partial eta square effect size of $\eta_p^2 = .088$, which are considered as large and medium effects on 2016-2017 Biology and Government EOC scores, respectively. In addition, the average of the three categories of 10% and higher had a higher MAP Index score mean than the two lowest categories for Government and Biology EOCs ($\mu_{3,4,5}=376$ vs. $\mu_{1,2}=369$ and $\mu_{3,4,5}=376$ vs. $\mu_{1,2}=370$,

respectively). This observational data combined with suggested medium and large effects of the percentage weight given to the SLO data in teachers' evaluation on the variance of Biology and Government EOC scores might suggest that policy makers and school administrators consider using a percentage weight of 10% or greater when using SLO data in teachers' summative evaluation. However, this must be done with caution and in collaboration with teachers as Kraft et al.'s (2018) research suggested adoption of high stakes teacher evaluations, using student test data and VAMs, lowered the number of new teacher licenses by 15%. With an estimated enrollment in public school set to rise by 1,000,000 students over the next 10 years, any new education evaluation reforms need to carefully measure possible consequences against the impact on the increase/decrease of new entrants in the teacher profession (Kraft et al., 2018). This is especially true for schools and subject areas in which locating and hiring quality staff members has been challenging (Kraft et al., 2018). Thus, the focus of the professional development on use of SLOs in teacher evaluations needs to be a collaborative effort with teachers and administrators to develop an SLO to be a true measure of student education and teacher effectiveness, which is student growth. This supports the observation that the majority of educators and administrators have a desire to see students succeed and understand the importance of using some form of student achievement data in the evaluation process (Zesiger, 2015). In addition, Longchamp (2017) suggested administrators that had received training on SLO implementation were better able to develop a culture of collaboration with teachers, which lead to positive attitudes towards the significance of the SLO process on student achievement.

The findings of the study indicate that the type of SLO used in teacher evaluations had an effect on the variance in Biology and Government EOC scores. The effect size on Biology EOC was a partial eta square effect size of $\eta_p^2 = .065$ and the effect size on Government EOC was partial eta square effect size of $\eta_p^2 = .043$, which are considered as medium and small effects on 2016-2017 Biology and Government EOC scores, respectively. In addition, the study indicated schools using yearly growth goals had a higher $\mu = 385$, than schools with unit growth goals $\mu = 368$ on 2016-2017 Biology EOC and schools using yearly goals $\mu = 372$ and unit goals $\mu = 374$ on 2016-2017 Government EOC. The implications of the results suggest school administrators might need to match the type of SLO used to the recommended length of the class. Biology is a yearlong class that has more standards to test, so an SLO based on yearly goals may be more appropriate given the class has more units. Whereas the Government class is typically a semester long class, so a unit SLO goal may be more appropriate. It is vital that administrators develop SLOs that are high quality as the research suggests high-quality SLOs led to gains in student achievement versus low-quality SLOs (Slotnik, Smith, Helms, et al., 2013). Slotnik, Smith, Helms, et al., suggested SLOs can play an integral part in student achievement when used in teacher evaluations because the increased attention to detail and quality in the SLO development is key to the success of SLOs on student achievement.

Recommendations for Future Research

As state education departments and districts look to use SLOs in teacher evaluations, policy makers and administrators will need to consider the different nuances of SLOs to accurately and effectively measure teacher effectiveness using student achievement scores. Areas of interest that need further research to measure impact of

student achievement in teacher evaluations based on this study include the following: (a) increase the number of study participants to increase the power of the effect sizes found in the study; (b) study the percentage weight that student achievement is given in teacher evaluations; (c) study the type of SLOs used in teacher evaluations; (d) study the perception of administrators and teachers on the use of SLOs and SGMs in teacher evaluation; (e) measure the results of English II and Algebra EOCs; and (f) determine the effect of prior professional development training of administrators and teachers on the use of SLOs in teacher evaluations.

Conclusion

The results of the study indicated the use of SLOs in teacher evaluations did not have a statistically significant difference on student achievement on Missouri Biology and Government EOC scores. The study did suggest the type of SLO used, the percentage weight given to SLO data in teacher evaluations, and whether teachers met SLO goals did have an effect on the difference on the variances of the 2016-2017 Missouri Biology and Government EOC scores, even though the differences were not statistically significant. This study adds data to the limited empirical research regarding the use of SLOs in teacher evaluations in Missouri and identified topics of future research regarding the use of student growth data for evaluation purposes.

REFERENCES

- Acheson, K. A., & Gall, M. D. (1997). *Techniques in the clinical supervision of teachers* (4th ed.). White Plains, NY: Longman.
- Adams, A. A. (2009). *A study of the attitudes and opinions of southwest Missouri educators regarding the value and outcome of the performance based teacher evaluation process* (Doctoral dissertation). Available from ProQuest Dissertation & Theses Global. (UMI No. 3390229)
- Alkin, M., & Christie, C. (2004). An evaluation theory tree. In M. Aklin (Ed.), *Evaluation roots, tracing theorists' views and influences* (pp. 12-21). Thousand Oaks, CA: Corwin.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260-267.
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.
- Anderson, R. H. (1993). Clinical supervision: Its history and current context. In R. H. Anderson & K. J. Snyder (Eds.), *Clinical supervision: Coaching for higher performance* (pp. 5-18). Lancaster, PA: Technomic.
- Aseltine, J. M., Faryniarz, J. O., & Rigazio-DiGilio, A. J. (2006). *Supervision for learning: A performance-based approach to teacher development and school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn,.,.,
Shepard, L.A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Balch, R., & Springer, M. (2015). Performance pay, test scores and student learning objectives. *Economics of Education Review*, 44(2015), 114-125.
- Bergin, C. (2015). *Using student achievement data to evaluate teachers* [White paper]. Retrieved from University of Missouri College of Education website:
https://nee.missouri.edu/documents/NEE_White_PaperStudent_Achievement_in_TeacherEvaluation2015_5_20.pdf
- Berliner, D. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, 116(1), 1-31.
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.
- Briggs, D.C. (2012). Making value-added inferences from large-scale assessments. In M. Simon, K. Ercikan, & M. Rousseau (Eds.), *Improving large scale assessment in education: Theory, issues and practice* (pp. 186-206). London, England: Routledge.
- Briggs, D.C., Diaz-Bilello, E., Maul, A., Turner, M., & Bibilos, C. (2014). *Denver ProComp evaluation report: 2010-2012*. Boulder, CO: Colorado Assessment Design Research and Evaluation (CADRE) Center and the National Center for the Improvement of Educational Assessment.
- Cornetto, K. M., Schmitt L. N. T., Malerba, C., & Herrera, A. (2010). *AISD REACH Year*

- 2 evaluation report II, 2008–2009* (DRE Publication No. 08.97). Austin, TX: Austin Independent School District. Retrieved from http://www.austinisd.org/sites/default/files/dre-reports/08.97_AISD_Reach_Year2_Evaluation_ReportII_2008_2009.pdf
- Cubberley, E. (1916). *Public school administration*. Boston, MA: Houghton Mifflin.
- Darling-Hammond, L., Amerin-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Dewey, J. (1916). *Democracy and education*. New York, NY: Macmillan.
- Dewey, J. (1938). *Experience and education*. New York, NY: Simon & Schuster.
- Gamst, G., Meyers, L., & Guarino, A.J. (2008) *Analysis of variance designs: A conceptual and computational approach with SPSS and SAS*. Cambridge, UK: Cambridge University Press.
- Glanz, J. (2008, March). *John Dewey's critique of scientific dogmatism in education with implications for current supervisory practice within a standards-based environment*. Paper presented at Annual meeting of the American Education Research Association Supervision and Instructional Leadership, New York, NY. Retrieved from <http://files.eric.ed.gov/fulltext/ED500776.pdf>
- Goldhaber, D., & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, (31), 1067-1083.
- Gurl, T.J., Caraballo, L., Grey, L., Gunn, J.H., Gerwin, D., & Bembenutty, H. (2016). *Policy, professionalization, privatization, and performance assessment:*

- Affordances and constraints for teacher education programs*. Zug, Switzerland: Springer International.
- Haertel, E. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Princeton, NJ: Educational Testing Service.
- Hanushek, E.A., & Rivkin, S.G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267-271.
- Hattie, J.A.C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon, England: Routledge.
- Hiebert, J., & Morris, A. (2012). Teaching, rather than teacher, as a path toward improving classroom instruction. *Journal of Teacher Education*, 63(2), 92-102.
- Hiebert, J., Morris, A., Berk D., & Jansen A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, 58(1), 47-61.
- Hogan, R. (2007). The historical development of program evaluation: Exploring the past and present. *Online Journal of Workforce Education and Development*. II(4), 1-14. Retrieved from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1056&context=ojwed>
- Holland, P. E., & Garman, N. (2001). Toward a resolution of the crisis of legitimacy in the field of supervision. *Journal of Curriculum and Supervision*, 16(2), 95-111.
- Hu, J. (2015). *Teacher evaluation based on an aspect of classroom practice and on student achievement: A relational analysis between student learning objectives and value-added modeling* (Doctoral dissertation). Available from ProQuest Dissertation & Theses Global. (UMI No. 3689454)

- Kraft, M.A., Brunner, E.J., Dougherty, S.M., & Schwegman, D. (2018). Teacher accountability reforms and the supply of new teachers. Manuscript in preparation. Retrieved from <https://scholar.harvard.edu/mkraft/publications/teacher-accountability-reforms-and-supply-new-teachers>
- Lachlan-Hache, L. (2015). *The art and science and science of student learning objectives: A research synthesis*. Retrieved from <http://www.air.org/sites/default/files/downloads/report/Art-and-Science-of-Student-Learning-Objectives-April-2015.pdf>
- Lachlan-Hache, L., Cushing, E., & Biovana, L. (2012). *Student learning objectives as measures of educator effectiveness: The basics*. Washington, DC: America Institute of Research.
- Longchamp, J.C., (2017). *The effect of student learning objectives on teachers and teaching as part of the teacher evaluation process: A grounded theory study*. (Doctoral Dissertaion). Retrieved from <http://scholarworks.uvm.edu/graddis/734>
- Madaus, G.F., & Stufflebeam, D.L. (1989). *Educational evaluation: Classic works of Ralph W. Tyler*. Boston, MA: Kluwer Academic.
- Marzano, R. (2009). *Designing & teaching learning goals & objectives*. Bloomington, IN: Marzano Research.
- Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Retrieved from <http://www.ascd.org/publications/books/110019/chapters/A-Brief-History-of-Supervision-and-Evaluation>.

- Marzano, R.J., & Toth, M.D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mathis, W. (2012). *Research-based options for educational policy making: Teacher evaluation*. Boulder, CO: National Education Policy Center. Retrieved from http://www.greatlakescenter.org/docs/Policy_Briefs/Research-Based-Options/01-Mathis_TEval.pdf
- McLean, R.A., & Sanders, W.L. (1984). *Objective component of teacher evaluation: A feasibility study* (Working Paper No. 199). Knoxville, TN: University of Tennessee, College of Business Administration.
- Missouri Department of Elementary and Secondary Education. (n.d.-a). *End of course*. Retrieved from <https://dese.mo.gov/college-career-readiness/assessment/end-course>
- Missouri Department of Elementary and Secondary Education. (n.d.-b). *Guide to the Missouri Assessment Program*. Retrieved from <https://dese.mo.gov/college-career-readiness/assessment/guide-missouri-assessment-program>
- Missouri Department of Elementary and Secondary Education. (2013a). *Overview of essential principles of effective evaluation*. Jefferson City, MO: Author.
- Missouri Department of Elementary and Secondary Education. (2013b). *Use of student growth measures in educator evaluation*. Jefferson City, MO: Author.
- Missouri Department of Elementary and Secondary Education. (2014). *Student learning objectives handbook*. Jefferson City, MO: Author.

- Missouri Department of Elementary and Secondary Education. (2016). *Administrative memo: EQ-16-001*. Jefferson City, MO: Author. Retrieved from <https://dese.mo.gov/sites/default/files/am/documents/EQ-16-001.pdf>
- Missouri Department of Elementary and Secondary Education. (2017a). *Comprehensive guide to the Missouri School Improvement Program*. Jefferson City, MO: Author. Retrieved from https://dese.mo.gov/sites/default/files/MSIP_5_2017_Comprehensive_Guide_0.pdf
- Missouri Department of Elementary and Secondary Education. (2017b). *End of Course Assessments Technical Report, 2016-17*. Retrieved from <https://dese.mo.gov/sites/default/files/asmt-eoc-tech-report-1617.pdf>
- Monroe, C.A. (2014). *The relationship between new teacher evaluation system ratings for teachers and their students' academic growth* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3621127)
- Murphy, D. (2012). *Where is the value in value-added modeling?* [White paper]. London, England: Pearson Education.
- Pajak, E. (2000). *Approaches to clinical supervision: Alternatives for improving instruction* (2nd ed.). Norwood, MA: Christopher-Gordon.
- Pajak, E., & Arrington, A. (2004). Empowering a profession: Rethinking the roles of administrative evaluation and instructional supervision in improving teacher quality. *Yearbook of the National Society for the Study of Education*, 103(1), 228–252.

- Payne, E.T. (2010). *Implementing walkthroughs: One school's journey* (Doctoral dissertation). Retrieved from https://vtechworks.lib.vt.edu/bitstream/handle/10919/30049/Payne_ET_D_2010.pdf?sequence=2
- Phi Delta Kappan. (2017). *PDK poll of the public's attitudes toward the public schools*. Retrieved from <http://pdkpoll.org/>
- Pollins, T. A. (2014). *Student learning objectives: A Rhode Island case study* (Doctoral dissertation). Available from ProQuest Dissertation & Theses Global. (UMI No. 3665193)
- Popham, W.J. (2013). *Evaluating America's teachers: Mission possible?* Thousand Oaks, CA: Corwin.
- Rich, M. (2012, July 6). "No Child" law whittled down by White House. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Sanders, W.L., & Horn, S.P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3), 299-311.
- Sanders, W.L. & Horn, S.P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.
- Sanders, W.L., & Rivers, J.C. (1996). Cumulative and residual effects of teachers on Future student academic achievement. Manuscript in preparation. Retrieved from http://bulldogcia.com/Documents/Articles/sanders_rivers.pdf

- Schmitt, L. (2011, Fall). AISD REACH program update, 2010–2011: Texas assessment of knowledge and skills: Growth and student learning objectives (DRE Publication No. 10.84). *ASID Research Update*, 1-7. Retrieved from http://www.austinisd.org/sites/default/files/dre-reports/rb/10.84_AISD_Reach_TAKS_and_SLOs_2010-2011.pdf
- Schmitt, L. N. T. (2014). *AISD REACH program: Summary of findings from 2007–2008 through 2012–2013* (DRE Publication No. 12.96). Austin, TX: Austin Independent School District. Retrieved from http://www.austinisd.org/sites/default/files/dre-reports/DRE_12.96_AISD_REACH_Program_Summary_of_Findings_2007_2008_Through_2012_2013_0.pdf
- Schmitt, L. N. T., Cornetto, K. M., Malerba, C., Ware, A., Bush-Richards, A., & Imes, A. (2009). *Strategic compensation initiative REACH pilot: 2007–2008 evaluation report* (DRE Publication No. 07.86). Austin, TX: Austin Independent School District. Retrieved from https://www.austinisd.org/sites/default/files/dre-reports/07.86_Strategic_Compensation_Initiative_REACH_Pilot_2007-08_Evaluation_Report.pdf
- Simon, M. K. (2011). Assumptions, limitations and delimitations. *Dissertation and scholarly research: Recipes for success*. College Grove, OR: Dissertation Services, LLC. Retrieved from <http://dissertationrecipes.com/wp-content/uploads/2011/04/AssumptionslimitationsdelimitationsX.pdf>

- Slotnik, W., Bugler, D., & Liang, G. (2015). *Change in practice in Maryland: Student learning objectives and teacher and principal evaluation*. Boston, MA: Community Training and Assistance Center.
- Slotnik, W.J., Smith, M., Glass, R., & Helms, B.J. (2004). *Catalyst for change: Pay for performance in Denver final report*. Boston, MA: Community Training and Assistance Center.
- Slotnik, W.J., Smith, M., Helms, B.J., & Qiao, Z. (2013). *It's more than money: Teacher Incentive Fund—Leadership for educators' advanced performance, Charlotte-Mecklenburg Schools*. Boston, MA: Community Training and Assistance Center.
- Slotnik, W.J., Smith, M., & Liang, G. (2013). *Focus on Rhode Island: Student learning objectives and evaluation*. Boston, MA: Community Training and Assistance Center.
- 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>
- Stiggins, R.J. (2014). *Defensible teacher evaluation: Student growth through classroom assessment*. Thousand Oaks, CA: Corwin.
- Stufflebeam, D.L., & Shinkfield, A.J. (1985). *Systematic evaluation: A self-instructional guide to theory and practice*. Boston, MA: Kluwer-Nijhoff.
- Tucker, P., & Stronge, J.H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tyler, R. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.

- U. S. Department of Education. (n.d.). *Every Student Succeeds Act (ESSA)*. Retrieved from <http://www.ed.gov/essa>
- Watras, J. (2006). The eight-year study: From evaluative research to a demonstration project, 1930-1940. *Education Policy Analysis Archives, 14*(21), 1-20. Retrieved from <http://epaa.asu.edu/ojs/article/view/92/218>.
- Watson, P. (2018). *Rules of thumb for magnitude of effect size*. Retrieved from University of Cambridge, MRC Cognition and Brain Science Unit website: <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize/>
- Webster, W., & Mendro, R. (1997). The Dallas Value-Added Accountability System. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press.
- Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunk, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness* (2nd ed.). New York, NY: The New Teacher Project.
- Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.
- Zesiger, J. W. (2015). *Transforming educator instructional practice in Pennsylvania secondary schools through the addition of teacher-specific student achievement data as a component of teacher evaluation* (Doctoral Dissertation). Available from ProQuest Dissertation & Theses Global. (UMI No. 3729496)

Appendix

Survey Instrument Questions

The definition of Student Learning Objective used in this survey and research is as follows: A method used to quantify a teacher's impact on student achievement by using pre-assessment data to set student growth goals for a particular class or course and then using a postassessment to measure the percentage of students that met the predetermined growth goals for the class or course (MODESE, 2014)

1. What is the name of your high school?
2. How many teachers teach high school Biology classes that take the Biology EOC?
3. Did your school's high school Science teacher(s) have an SLO for Biology I class for **any** of the following years: 2014-15, 2015-16, 2016-17?
4. Was the SLO used in the Biology class(es) based on yearly student growth goals or a unit growth goal?
5. Did your school district include the results of the high school Science teacher's SLO in Biology I for teacher evaluation purposes for the 2016-2017 school year?
6. What percent of the high school Science teacher(s) met the predetermined SLO growth goal for Biology in the 2016-2017 school year? (choices: less than 50%, 50% to 74%, 75% to 89%, 90 to 99%, 100%)
7. Did your school district include student test performance from the Missouri Biology EOC in the 2016-2017 SLOs?

8. Did your school district include the results of the high school Science teacher's SLO in Biology I for teacher evaluations in the 2015-2016 school year?
9. Did your school district include the results of the high school Science teacher's SLO in Biology I for teacher evaluation in the 2014-2015 school year?
10. How many teachers teach high school Government classes that take the Missouri Government End of Course (EOC) exam?
11. Did your school's high school Government teacher(s) have an SLO for Government class for **any** of the following years: 2014-15, 2015-16, 2016-17?
12. Was the SLO used in the Government class based on yearly student growth goals or unit growth goal?
13. Did your school district include the results of the high school Social Science teacher's SLO in Government for teacher evaluation purposes for the 2016-2017 school year?
14. What percent of the high school Social Science teacher(s) met the predetermined SLO growth goal for Government in the 2016-2017 school year? (choices: less than 50%, 50% to 74%, 75% to 89%, 90 to 99%, 100%)
15. Did your school district include student test performance from the Missouri Government EOC on the 2016-2017 SLOs?
16. Did your school district include the results of the high school Social Science teacher's SLO in Government for teacher evaluation purposes in the 2015-2016 school year?

17. Did your school district include the results of the high school Social Science teacher's SLO in Government for teacher evaluation purposes in the 2014-2015 school year?
18. What percentage of your district's Biology and Government teachers' summative evaluation final score was based on meeting the student growth goals of the SLO?