PREDICTING OHIO PRINCIPALS’ INTENTIONS AND PRACTICES TOWARD STATE EVALUATION-BASED PROFESSIONAL GROWTH PLANS USING THE THEORY OF PLANNED BEHAVIOR

Kelly L. Schooler

A Dissertation
Submitted to the Graduate Faculty of The University of Findlay’s College of Education in partial fulfillment of the requirement for the degree of

DOCTOR OF EDUCATION

August 2016

Committee:

Kathleen C. Crates, PhD
Chair, Dissertation Committee

Jon Brasfield, PhD
Committee Member

Jill L. Spanheimer, EdD
Committee Member
ABSTRACT

Even with nation-wide investments, very little is known about how to link teacher evaluation data to teacher development. This study determined the predictability of principals’ intentions and practices toward teachers’ evaluation based professional growth plans using the theory of planned behavior. Ultimately, 170 principals’ Rasch calibrated survey responses were subjected to path analysis. The theory of planned behavior components effectively predicted principals’ intentions and practices toward teacher evaluation based professional growth plans. Subjective norm and attitude were significant predictors of principals’ intentions to conduct evaluation based professional growth plans with teachers; and, principals’ intentions were a significant predictor of principals’ practices toward teachers’ evaluation based professional growth plans. These findings offer insight into mitigating the barrier between evaluation data and teacher development.
DEDICATION

To Dad,

You have always been and will always be my best teacher.

I love you.
ACKNOWLEDGMENTS

This dissertation would not be possible without the support of my family, dissertation committee, and friends. Jillella Schooler Weingart provided the source of inspiration that only springs from a loving daughter eager to witness her mother succeed. Nick Weingart, a hero of a husband, for his strength to push me toward my finish line while in the midst of his battle. Dr. Jill Spanheimer for modeling unconditional love, integrity, generosity, and strength from our first moment. Her guidance gave me the gift of a thoughtful and rewarding motherhood. Her example and encouragement paved the path toward my own, and others, doctoral pursuit. I will forever honor our bond.

I express my genuine and whole-hearted gratitude to my chair, Dr. Kathleen Crates, for blanketing my dissertation process with calm assurance and healing conversations. Dr. Crates had the uncanny ability to reach out to me at my most vulnerable times and intuitively pushed me forward. During the literature review, research design, and statistical analysis Dr. Jon Brasfield’s reliable insights and valuable advice kept me on track to completion. His ability to keenly listen and patiently consult stretched my reach.

Finally, I would like to thank Tammy Adams, Casey Bevilockway, and Dr. Mary liames. These women helped me professionally and personally as I tried to juggle school, family, and completing my dissertation. Thank you Tammy Adams and Casey Bevilockway for your loyalty to me, and our district, in times of plenty and times of crisis. You both epitomize the honor of your profession and it distresses me to think of serving without you by my side. Lastly, I thank Dr. Mary liames for her gift of sisterhood. Your true love and understanding of God’s gift of sisterhood makes our place of service more beautiful.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER I: INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Rationale &amp; Significance of the Study</td>
<td>4</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Theoretical Framework</td>
<td>7</td>
</tr>
<tr>
<td>Research Questions</td>
<td>9</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>9</td>
</tr>
<tr>
<td>Delimitations</td>
<td>12</td>
</tr>
<tr>
<td>Limitations</td>
<td>12</td>
</tr>
<tr>
<td>Researcher Bias</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER II. LITERATURE REVIEW</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation Historical Background</td>
<td>16</td>
</tr>
<tr>
<td>Teacher Evaluation Policy Catalyst</td>
<td>17</td>
</tr>
<tr>
<td>Current Teacher Evaluation Policy</td>
<td>20</td>
</tr>
<tr>
<td>New Implementation Trends</td>
<td>24</td>
</tr>
<tr>
<td>Teacher Evaluation and Professional Development</td>
<td>28</td>
</tr>
<tr>
<td>Teacher Evaluation and School Improvement</td>
<td>36</td>
</tr>
<tr>
<td>Teacher Evaluation and Human Capital Management Systems</td>
<td>41</td>
</tr>
<tr>
<td>Teacher Evaluation Tools and Roles</td>
<td>44</td>
</tr>
<tr>
<td>Teacher Evaluation and Teacher Development Platforms</td>
<td>46</td>
</tr>
<tr>
<td>Teacher Evaluation and Principals’ Roles</td>
<td>52</td>
</tr>
<tr>
<td>Teacher Evaluation and Principals’ Roles</td>
<td>55</td>
</tr>
</tbody>
</table>
Summary .................................................................................................................. 57

CHAPTER III. METHODOLOGY .............................................................................. 63

Research Design ..................................................................................................... 63
Participants ............................................................................................................. 66
Instrumentation & Data Sources .......................................................................... 71
Data Collection Procedures ................................................................................ 80
Research Questions .............................................................................................. 84
Data Analysis ......................................................................................................... 85
Assumptions ........................................................................................................... 91

CHAPTER IV. RESULTS ....................................................................................... 94

Characteristics of the Sample .............................................................................. 95
Instrument Validity and Reliability .................................................................... 96
Research Question 1 ............................................................................................ 110
Research Question 2 ............................................................................................ 114
Research Question 3 ............................................................................................ 117
Summary ............................................................................................................... 121

CHAPTER V. CONCLUSIONS AND RECOMMENDATIONS ................................ 124

Review of the Study ............................................................................................ 124
Discussion ............................................................................................................. 126

   Research Question 1 .......................................................................................... 126
   Research Question 2 .......................................................................................... 131
   Research Questions 3 ........................................................................................ 133
Conclusion ............................................................................................................. 135
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Table 1 New Generation Early Implementation Teacher Evaluation Results</td>
<td>168</td>
</tr>
<tr>
<td>2</td>
<td>Table 3.1 2013 Ohio School District Typology</td>
<td>169</td>
</tr>
<tr>
<td>3</td>
<td>Table 3.2 Cronbach’s Alpha Reliability Test</td>
<td>170</td>
</tr>
<tr>
<td>4</td>
<td>Table 3.3 Final Survey Instrument</td>
<td>171</td>
</tr>
<tr>
<td>5</td>
<td>Table 3.4 Research Questions and Corresponding Statistical Test</td>
<td>173</td>
</tr>
<tr>
<td>6</td>
<td>Table 4.1 2013 Ohio Typology Classification Principals’ Representation</td>
<td>174</td>
</tr>
<tr>
<td>7</td>
<td>Table 4.2 Rasch-Analysis Item Fit Indices</td>
<td>175</td>
</tr>
<tr>
<td>8</td>
<td>Table 4.3 Psychometric Properties of Measurement Scales for TPB</td>
<td>178</td>
</tr>
<tr>
<td>9</td>
<td>Table 4.4 Step Thresholds of Measurement Scales</td>
<td>179</td>
</tr>
<tr>
<td>10</td>
<td>Table 4.5 Means, Standard Deviations, and Correlations of TPB</td>
<td>180</td>
</tr>
<tr>
<td>11</td>
<td>Table 4.6 Means, Standard Deviations, and Correlations of TPB Collapsed Scales.</td>
<td>181</td>
</tr>
<tr>
<td>12</td>
<td>Table 4.7 Multicollinearity Tolerance Test and Variance Inflation Factor Test</td>
<td>182</td>
</tr>
<tr>
<td>13</td>
<td>Table 4.8 Ridge Regression Analysis Report of Eigenvalues of Correlations</td>
<td>183</td>
</tr>
<tr>
<td>14</td>
<td>Table 4.9 Effect of Each TPB Component on OTES Professional Growth Practices</td>
<td>184</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Figure 1.1 OTES Professional Growth Plan Intentions and Practices</td>
<td>185</td>
</tr>
<tr>
<td>2</td>
<td>Figure 4.1 Person Item Map</td>
<td>186</td>
</tr>
<tr>
<td>3</td>
<td>Figure 4.2 Failed Theory of Planned Behavior Based Model 1</td>
<td>187</td>
</tr>
<tr>
<td>4</td>
<td>Figure 4.3 Standardized Result of Second Path Analysis</td>
<td>188</td>
</tr>
<tr>
<td>5</td>
<td>Figure 4.4 Analysis of Variance for Intentions Typology Classifications</td>
<td>189</td>
</tr>
<tr>
<td>6</td>
<td>Figure 4.5 Analysis of Variance for Behaviors Typology Classifications</td>
<td>190</td>
</tr>
</tbody>
</table>
CHAPTER I. INTRODUCTION

Background of the Problem

On September 23, 2011, all fifty state superintendents of education departments received a letter from U.S. Education Secretary Arne Duncan, offering each state the opportunity to request specific requirements of the No Child Left Behind Act of 2001 (NCLB) in exchange for implementing a rigorous education reform. Nationwide, reform efforts in critical areas, for example, transitioning to college, career-ready standards and assessments, developing differentiated recognition systems, accountability and support, and evaluating and supporting teacher and principal effectiveness, were set into motion (U.S. Department of Education, 2013).

As of September 27, 2013, the majority of the states, including Ohio, accepted Duncan’s call to reform teacher and leader evaluation and support systems (U.S. Department of Education, 2013). The U.S. Department of Education’s (DOE) A Blueprint for R.E.S.P.E.C.T (2013) explains the DOE’s vision for recognizing educational success, professional excellence, and collaborative teaching, stating,

We’ve included a focus on evaluation and support throughout our policies and programs because we believe they are the cornerstone of both effective professional development and human capital management. These systems enable schools to help teachers improve their practice by identifying their professional development needs. (p. 13)

The context of the DOE’s claim that new evaluation systems enable schools to improve teaching practices by identifying professional development needs illustrated the foundation of the international prompt for school leaders to create a link between teacher evaluation data and professional development as a means to improve instruction (Robinson & Timperley, 2013; Tuytens & Devos, 2011). The call for school leaders to activate teachers’ professional
development through teacher evaluation rests upon the frail assumption that principals automatically create the connection between teacher evaluation data and professional development (Smylie, 2014). Many principals commonly grapple with the complexities involved in connecting teacher evaluation data and professional development due to the troublesome relationship between evaluation and professional development (Smylie, 2014).

Some principals admit they purposefully oversimplify the connection between teacher evaluation data and professional development because evaluation-based professional development requires uncomfortable professional conversations about teachers’ performance, including strengths and weaknesses, to improve teachers’ practice (Sartain, Stoelinga, & Brown, 2011). Evaluation-based professional development purposefully focuses on identifying teachers’ strengths and weaknesses to promote improved instructional practice (Behrstock-Sherrat & Jacques, 2012).

Myung and Martinez (2013) argued that the current teacher evaluation policy reforms centrally place principals in the role of supporting individual teachers’ professional growth, regardless of the limited guidance available on how it can be done successfully. Facilitating a culture where principals commonly identify individual teacher improvement needs can create a provocative challenge for many school leaders. No one has stated this position as clearly as Kevin Huffman, commissioner of the Tennessee Department of Education,

In many cases, you are sending principals to give feedback to teachers who they have been managing for 10, 15, 20 years, and they have never told that teacher that they need to improve. So now after 15 years of working with this same person, they are about to have to sit down and tell them, ‘Hey, you’re not very good at this particular set of things.’ Even for people who have the best of intentions, the psychology of that is hard. (Myung
Huffman paints an intimate portrait of the typical relationships that exist between principal and teachers and provides a glimpse into the complexities of giving and receiving evaluation-based professional growth recommendations.

Goe, Biggers, and Croft (2012) supported the notion that principals and teachers struggle with using evaluation data and results to identify strengths and weaknesses in instructional practices. Geo et al. (2012) additionally emphasized the critical role principals typically play in interpreting evaluation data and results into professional development recommendations and argued that principals are often not adequately prepared to assume this role. Similar claims by Sartain et al. (2011) reaffirm that many principals are unable to coach teachers toward instructional improvements.

Geo et al. (2012) identified the evaluation process as the primary provision of important evidence used to better align professional growth opportunities, explaining,

The process of evaluation provides the evidence of areas where teachers need help, but that process alone does not change teaching practice. Rather, using evidence for professional growth opportunities and coaching sessions is where it will have an impact on instruction and student outcomes. (p.15)

In making this comment, Geo et al. emphasized the importance of implementing the new evaluation systems with structures for prioritizing evaluation evidence in informing professional development for teaching and learning improvements.

The Ohio Teacher Evaluation System introduced the professional growth plan as the structure through which to prioritize the use of teacher evaluation to inform and guide professional development for teachers. Consequently, Ohio principals’ intentions and practices
toward OTES professional growth plans deserve meaningful examination as teacher evaluation reform progresses.

**Rationale & Significance of the Study**

In 2013, the Ohio Legislature mandated that each district’s board of education, in consultation with teachers employed by the board, adopt a standards-based teacher evaluation policy that aligned with the state framework (Ohio Teacher Evaluation System, [OTES]) for teacher evaluation (Ohio Revised Code [ORC] 3319.112, 2011). The state framework for teacher evaluation in Ohio, OTES, was largely derived from Ohio’s 2011 Race to the Top (RttT) federal grant application and included the following plan elements for redesigning Ohio’s educator evaluation system: (1) annual evaluations of teachers and principals in agreement with state regulations and aligned to the state evaluation model and RttT criteria, (2) evaluator training and credentialing to safeguard the reliability of the model evaluation systems for teachers and principals, and (3) using teacher and principal performance data to plan targeted supports and professional development and to make decisions for retention and dismissal of and tenure and compensation for teachers and principals (Ohio Department of Education, 2011).

As seen by RttT plan elements, the OTES model expected to satisfy the dual need for accountability and professional development. At the onset OTES is presented as an observation tool for administrators to provide teachers with learning supports for enhancing their teaching. Yet simultaneously, OTES aims to directly link teachers’ performance outcomes and employment decisions in the new evaluation system. Policymakers showed trends toward favoring the relationship between evaluation and professional development: Nearly every state identified professional growth as the primary purpose of evaluation reform in their waiver
applications for relief from No Child Left Behind sanctions (Center on Great Teachers and Leaders, 2013).

Ohio has historically placed minimal attention toward teacher and principal evaluation systems; therefore, OTES represented a ‘radical change’ in teacher evaluation (Kowalski & Dolph, 2015). This radical change calls for a major paradigm shift at the district level that embraces and cultivates changes in school cultures and structures that help teachers and principals move toward teacher evaluation that promotes professional growth (Minnici, 2015). The professional development expectations of the new evaluation systems fall squarely on principals’ shoulders and present a huge challenge for them (Cosner, Kimball, Bakowski, Carl, & Jones, 2015). At the most basic level, OTES requires that principals shape teachers’ professional growth opportunities by providing time and financial allocations for professional development based on evaluation results (OTES Training Manual, 2012). Previous researchers emphasized the issue (Hallinger, Heck, & Murphy, 2013; Johnson, 2012) while investigating the importance of principals’ readiness and beliefs in the implementation of new state and federal teacher evaluation systems.

An examination to determine principals’ intentions and practices toward the professional growth plan component of OTES found precedence in this study because the entire OTES model is based on the utilization aim of the professional growth plan to identify improvement areas for teachers, and then to help those teachers find the professional development at that individual level that they need to improve (C. Frey, personal communication, November 3, 2014).

According to Kowalski and Dolph (2015), more than half (52%) of Ohio principals reported that they do not believe the professional growth plan requirement is an effective OTES element. Earlier research findings showed teacher opposition to individual professional growth
plans due to teacher skepticism; that is, teachers remained unconvinced that principals possess the capacity to offer meaningful guidance (Kowalski & Dolph, 2015). Principals’ disposition toward the OTES time requirement paralleled repeated findings that reported principals identify lack of time to be the biggest obstacle of the new evaluation systems (Hill, 2013; Kersten & Israel, 2005; Kowalski & Dolph, 2015). Kowalski and Dolph (2015) state that 96% of Ohio principals participating in their study reported time devoted to OTES as excessive. Ohio principals appear diametrically opposed to OTES professional growth plans when principals specifically identify their required involvement in teacher professional growth plans as a reason for the exacerbated time requirements in the appraisal system (Kowalski & Dolph, 2015).

The successful implementation of an evaluation-based professional development system hinges largely on the intentions and practices of school principals toward OTES professional growth plans. Sinnema and Robinson (2007) argued that a need has emerged for research that specifically studies the leadership actions of principals and the use of teacher evaluation systems. Although teacher evaluation policy receives widespread attention, the literature search revealed no previous studies on the role of Ohio principals’ intentions and practices toward evaluation-based professional development in strengthening teacher evaluation systems (Behrstock-Sherrat & Jacques, 2012). Addressing this lack of information about principals’ intentions and practices toward OTES professional growth plans builds on the educational community’s limited knowledge about how principals use teacher evaluation systems.

**Purpose of Study**

Personalizing professional development to the identified needs of a teacher through evaluation data is a frequently mentioned reform effort (Hill & Grossman, 2013). As national reformers focus on improving teacher evaluation systems, more research is needed to identify the
extent to which principals use evaluation data to improve individual performance (Hill & Grossman, 2013). Currently, teacher evaluation research provides limited information regarding the role of evaluation processes in improving the practice of individual teachers and is largely silent on the use of evaluation data to improve professional development (Maslow & Kelly, 2012). The current study seeks to discern whether or not the Ohio Teacher Evaluation System model presented by the Ohio Department of Education effectively meets the intended goal of principals using professional development plans to help teachers find the individual professional development they need to improve professionally.

This study focuses on deepening the understanding of Ohio principals’ intentions and practices toward OTES professional growth plans. It seeks to quantitatively gauge Ohio principals’ intentions and practices toward OTES professional growth plan implementation. The objective of this study is to provide an empirical theoretical base that can estimate principals’ participation in OTES professional growth plans.

This study employed Ajzen’s (1991) theory of planned behavior and used Rasch measurement techniques to estimate principals’ intentions and practices toward OTES professional growth plans. The yield of this theory-based instrument investigated Ohio principals’ attitudes, subjective norms, and perceived behavioral control toward OTES professional growth plans. As a result, the researcher expects that educational practitioners gain an empirically sound tool for estimating principals’ intentions and practices toward evaluation-based professional growth plans in general and in Ohio public school sites specifically.

**Theoretical Framework**

The theoretical framework developed to predict principals’ intentions and practices toward OTES professional growth plans is based on Ajzen’s (1991) theory of planned behavior.
Ajzen’s (1991) theory of planned behavior finds a broad application in the field of general education (Ingram, Cope, Harju, & Wuensch, 2000; Yan 2014). Using Ajzen’s (1991) theory of planned behavior the following Principals’ OTES Professional Growth Plan Intentions and Practices’ model (see Figure 1) emerged to estimate principals’ intentions and practices toward OTES professional growth plans.

Figure 1.1

Principals' OTES Professional Growth Plan Intentions and Practices
The Principals OTES Professional Growth Plan Intentions and Practices model depicts the convergence of factors leading to a principal’s intentions and behaviors toward OTES professional growth plans. The principals’ affective attitude, instrumental attitude, subjective norms, perceived behavior control, self-efficacy, and intentions are mitigating factors in principals’ practices toward OTES professional growth plans.

**Research Questions**

1. Can principals’ intentions toward OTES professional growth plans be predicted by principals’ attitudes, subjective norms, and perceived behavior control toward OTES professional growth plans?

2. Can principals’ OTES professional growth plan practices be predicted based on their intentions and perceived behavior control toward OTES professional growth plans?

3. Do principals’ intentions and practices differ among Ohio’s typology classifications?

**Definition of Terms**

1. **Attitude.** A learned predisposition to respond in a consistently favorable or unfavorable manner with respect to an object or class of objects (Fishbein, 1967); attitudes exert a direct influence on behavior (Kolekofski & Heminger, 2003).

2. **Subjective norm.** A person’s perceived expectations of significant peers with regard to his or her performing the behavior under investigation (Sutton, 1998).

3. **Perceived behavior control.** The extent to which a person feels he or she has control over performing the behavior, or the ease of performing the behavior (Sutton, 1998).

4. **Intention.** How a person plans to perform a given behavior (Fishbein, 1967).

5. **Credentialed evaluator.** OTES-credentialed evaluators must take a three-day in-person training before taking an online evaluator credentialing test developed by the National
Institute for Excellence in Teaching (NIET). All OTES evaluators must be credentialed by NIET (OTES Training Book, 2012).


7. Evaluation process. The evaluation process requires the evaluator to use evidence gathered in a variety of avenues (professional growth or improvement plan, observations, walkthroughs, and conferences) to determine a teacher performance rating (OTES Training Book, 2012. p.13).

8. Improvement plans. Improvement plans are developed for teachers by the evaluator in response to ineffective performance and/or student growth ratings. The improvement plan is intended to pinpoint specific areas for performance improvement and for identifying the support needed to help the teacher improve (OTES Training Book, 2012. p.12).


10. Post-conferences. A post-observation conference provides reflection and feedback on the observed lesson and identifies strategies and resources for the teacher to incorporate into lessons to increase their effectiveness. Following the lesson, the teacher reflects on the lesson and the degree to which the student learning outcomes were met. Professional conversations
between the evaluator and the teacher during the post-conference will provide the teacher with feedback on the observed lesson and may identify additional strategies and resources. The evaluator will make recommendations, which may then become part of the teacher’s professional development plan. In general, the discussion between the evaluator and teacher needs to focus on the degree of the lesson’s success (reinforcement) and areas requiring further support (refinement). Teachers may bring additional evidence supporting the lesson observed to share with the evaluator at the conference. The evaluator may consider this evidence as an indication of student learning or as support for the teacher’s performance rating (OTES Training Book, 2012. p. 13).

11. Professional growth plans. Professional growth plans help teachers focus on areas of professional development that will enable them to improve their practice. The professional growth plan and process includes feedback from the evaluator and the teacher’s self-assessment, as well as the support needed to further the teacher’s continued growth and development (OTES Training Book, 2012. p. 13).


13. Teacher Performance Evaluation Rubric. This rubric is holistically scored, which means that evaluators will assess which level provides the best overall description of the teacher. The scoring process is expected to occur upon completing each 30-minute observation and post-conference. The evaluator is expected to consider evidence gathered during the pre-observation conference, the observation, the post-observation conference, and classroom walkthroughs (if
When completing the performance rubric, please note that evaluators are not expected to gather evidence for all indicators in each observation cycle. Likewise, teachers should not be required to submit additional pieces of evidence to address all indicators. The professionalism section of the rubric may use evidence collected during the pre-observation and post-observation conferences as well as information from the professional growth and/or improvement plan, if applicable (OTES Training Book, p.15).

**Delimitations**

Ajzen’s (1991) theory of planned behavior questionnaire was re-administered three months after participants completed it. In the follow-up questionnaire, participants were contacted and asked to report whether they had performed the intentions of the target behavior administered in the first questionnaire. This study was not longitudinal. The researcher did not administer the second questionnaire.

To date, each district in Ohio must implement one of the four Ohio Department of Education approved teacher evaluation models: (1) OTES model, (2) OTES model with locally-developed student growth measures, (3) a local evaluation system aligned to the OTES model, (4) or a local evaluation system aligned to the OTES model with locally developed student growth measures.

All study participants must be employees of a public district utilizing every component of the OTES model. The researcher will exclude participants employed by districts that independently employ local evaluation systems submitted to the Ohio Department of Education for alignment approval.

**Limitations**
Several limitations exist in this study due to the strong reliance on self-reported data. Self-reported data descriptions of principals’ intentions and practices toward OTES professional growth plans proved difficult to independently verify. Senunyeme (2013) contended that self-reported, or perception data, contain potential sources of bias that must be noted as limitations and include:

(1) Selective memory (remembering or not remembering experiences or events that occurred at some point in the past); (2) Telescoping (recalling events that occurred at one time as if they occurred at another time); (3) Attribution (the act of attributing positive events and outcomes to one’s own agency but attributing negative events and outcomes to external forces); and, (4) Exaggeration (the act of representing outcomes or embellishing events as more significant than is actually suggested from other data). (p. 105)

Most noticeably, the potential effects of principals’ inflated sense of self-efficacy may impact their perception of the actual OTES professional growth plan implementation. The explanation for this noticeable limitation derives from Senunyeme’s (2013) assertion that attribution bias is the tendency to attribute positive events to one’s self and negative events and outcomes to outside sources. For example, a principal completing the survey may inflate the level of his or her implementation performance and may contribute lower levels of professional growth plan implementation to deficiencies in the teacher’s implementation of OTES professional growth plans; thus, the principal attributes a need for implementation enhancements to the teacher and not to him- or herself.

Ohio’s Teacher Evaluation System comprises 50% student growth measures and 50% performance on a standards-based rubric. Combined, these two equal parts produce the final
evaluation rating. Teachers who do not meet acceptable scores on student growth measures are required to participate in an improve plan. This study solely focuses on principals’ intentions and practices toward professional growth plans, not improvement plans.

**Researcher Bias**

The intended topic—principals’ intentions and practices toward OTES professional growth plans—is wrapped heavily in the researcher’s opinions, assumptions, expectations, and personal experiences. The researcher is a practicing principal directly involved in teacher evaluation.

The researcher holds a skeptical view that a high number of principals possess the will to implement OTES professional growth plans as robustly as described in the OTES training manuals. The researcher believes that many principals superficially engage in the teacher evaluation process for a variety of reasons, including self-preservation, conflict avoidance, a lack of support at the district level, and the inability or unwillingness to prioritize the link between teacher evaluation data and professional development for teachers.

The researcher believes that many principals intuitively know they have blind spots or hidden weaknesses directly related to evaluation-based professional development, and that many principals allow themselves to ignore the obvious misalignment of teacher evaluation data and professional development as compensation for the hyper-demands of the principal’s job. For example, the daily demands on the principal are often subtle in that teachers expect him or her to shield them from district and state policies. Welch (1978) graphically illustrated the complex relationship between principals and teachers. Welch (1978) draws a clear connection between the principal and the sin-eater, a person who willingly consumes the sins of the dead, and proposes that successful principals do not shun such a role but embrace eating the sins of those
they serve. These sins can translate into the principal willingly swallowing anything from
evaluation oversights such as district curriculum directives being ignored to contract
expectations being deficiently weighted against the district.

The researcher expects it will be difficult to solicit honest feedback from principals about
their genuine intentions and practices toward OTES professional growth plans. The researcher
expects some principals to rationalize how they are making meaningful connections to satisfy the
professional growth plan component of OTES. In most cases, the researcher believes principals
are genuinely doing the best they can to push a teacher to grow while simultaneously
maintaining a necessary balance of peace. However, the researcher suspects that if the teachers
they served were asked to report on their individual professional growth plan experiences, a
significantly lower level of OTES professional growth plan principal behaviors might emerge.
The researcher equally anticipates that those same teachers’ appreciation for their principals’
evaluation skills and professional development provisions arise when discussions of outside
evaluators are introduced. The researcher believes teachers are skeptical of outside evaluators
because they intuitively know these creatures do not eat sins. Ultimately, the researcher
considers the complexity of the relationships between principals and teachers to distort the
teacher-evaluation process.
CHAPTER II. LITERATURE REVIEW

Research into teacher evaluation theory has seen renewed focus in the wake of studies indicating that many teachers are not evaluated systematically, and that professional development is rarely linked to explicit needs identified during the evaluation process (Looney, 2011). This chapter presents a summary of the literature related to the implementation of Ohio’s Teacher Evaluation System (OTES) and processes for aligning teacher evaluation data and professional development. It will begin with a brief discussion of the historical context of teacher evaluation and the policy catalyst for a new generation of teacher evaluation systems with broad implementation trends across the nation. Next, the literature related to exposing the nature of the relationship among teacher evaluation and professional development, systemic school improvement, and human capital resource management will be examined. Third, the chapter will present an exploration of the literature related to the challenges found in the use of observation tools to improve teachers and teaching quality. The fourth examination of literature will explore the influence of new generation teacher evaluation implementation on the role of the principal and online professional development tools. Finally, this chapter will conclude by exploring the major agreements within the extant literature on teacher evaluation.

Research into literature on implementing new teacher evaluation systems and aligning evaluation data to teachers’ professional development needs was conducted using One Search and Ohio Link. One Search parameters were set for peer-reviewed sources found predominantly in American academic journals with a major emphasis on recent literature and minor emphasis on seminal literature. The following key search terms were employed: Ohio Teacher Evaluation System, teacher evaluation, teacher evaluation and professional development, teacher evaluation models, teacher observation, teacher observation evaluation, teacher observation protocol,
teacher evaluation reform, teacher evaluation and student achievement, teacher perceptions of evaluation, principal perceptions of evaluation, effectiveness of teacher evaluation, teacher professional growth plans, and professional development and standards-based observations. After completing an annotated bibliography for each key term, the research was organized into four subtopics: a) births and trends of new teacher evaluation systems, b) teacher evaluation and professional development, systemic school improvement and human capital resource management, c) teacher observation challenges, and d) influences of new teacher evaluation systems on the principal’s role, online professional development tools, and evolutionary evaluation.

**Teacher Evaluation Historical Background**

In the 1700s, education was not a formal field of study or professional discipline; thus, early towns throughout the United States turned to prevailing power structures, such as the church clergy, to hire teachers, offer judgments about their job performance, and supervise instruction (Marzano, Frontier, & Livingston, 2011). During this time, the community categorized teachers as public servants, and teachers’ supervisors or supervisory committees had almost unlimited power to establish the criteria for quality performance (Burke & Krey, 2005). The feedback that teachers received varied radically because concepts of educational pedagogy did not exist (Marzano et al., 2011). Historians commonly accept that feedback for teachers during this time ranged from performance on mundane custodial tasks, like cleanliness of the environment, to specific feedback regarding religious instruction in schools (Burke & Krey, 2005).

During the middle 1800s, a view of specialized teaching surfaced, and value was placed on teachers receiving complex feedback to advance teacher efficacy developed (Marzano, et al.,
According to Marzano et al. (2011), the rise of pedagogical awareness paved way for our modern-day comprehensive approach to developing teacher expertise. In turn, recognizing the importance of enhancing teacher expertise propagated the development of efforts to improve educational opportunities.

In 1965, President Lyndon Johnson introduced the Elementary and Secondary Education Act (ESEA) of 1965, which was designed to improve educational opportunities for poor children (Mondragon & Stapleton, 2005). Federal funding did not serve as a general aid package to all schools, and does not today. Equity based allocation formulae are directed to assist local education agencies (LEAs) with the most significant proportions of children living in poverty. The aid distribution design still flows first through the state education agencies (SEAs), allowing the federal government to leverage initiatives with funding. Since the passage of ESEA in 1965, reauthorization has appeared periodically throughout the years. Significant relationship threads appear in the ESEA’s reauthorizations of 1994, Improving America’s Schools Act (IASA), President George W. Bush’s 2001 No Child Left Behind (NCLB) reauthorization, President Obama’s ESEA Flexibility Waiver Program, and today’s Ohio’s Teacher Evaluation System’s (No Child Left Behind-Overview, 2014) policy.

According to The New America Foundation (New America Foundation, 2014), the ESEA 1994 reauthorization and IASA have supported challenging standards and aligning assessments and accountability elements for states and local school districts receiving federal funding under the law. The 1994 reauthorization required Title 1 states to develop challenging content and performance standards by the 1997-1998 school year. Additionally, aligned assessments, and an accountability system based on student performance against those standards, were expected by the 2000-2001 school year (U.S. Department of Education, 2001). The origin of the Ohio
Teacher Evaluation System can be traced to 1994 IASA Reauthorization. A significant foreshadowing, extracted from “High Standards for All Students: A Report from the National Assessment of Title I on Progress and Challenges Since the 1994 Reauthorization” (2001), fostered the notion that the Ohio Teacher Evaluation System’s arrival was near: “Parents, policymakers, and educators agree that every child needs a competent and well-qualified teacher in the classroom. Research shows the quality of teaching is the most important in-school factor in improving student achievement” (p. 37).

The 2001 NCLB Act, using the original ESEA’s federal funding distribution leverage design, received acclaim for its requirement that states and schools disaggregate standardized student achievement results by race, poverty, disabilities, and English-speaking proficiency in an effort to systematically target achievement gaps. The NCLB Act undeniably contributed to the framework of OTES and influenced the passage of Ohio Senate Bill 2 in 2004. According to Metz and Wraight (2011) Senate Bill 2 mandated an Educator Standards Board charged with developing Ohio Standards for the Teaching Profession and the Ohio Standards for Professional Development. The Ohio Department of Education adopted Ohio Standards for the Teacher Profession and aligned to the Ohio Standards for Professional Development in 2005, as proposed by the Educator Standards Board (OTES Training Book, 2012).

In 2009, Ohio’s Education Reform Plan gave the directive to the Educator Standards Board to recommend a model evaluation system for teachers and principals to the State Board of Education (Metz & Wraight, 2011). Ohio’s 2009 Education Reform Plan coincided well with the federal education policy of the American Reinvestment and Recovery Act (ARRA) of 2009. Since ARRA, federal dollars started to roll out to states under a competitive grant program entitled Race to the Top (RttT). Smarick (2009) described the RttT federal grant application
guidelines as *extraordinarily prescriptive*. Hence, much of Ohio’s RttT application emphasized teacher evaluation reform (Ohio Department of Education, 2010) and accountability. As of May 2014, slightly under half of Ohio’s districts were participating in RttT (Ohio Department of Education, 2014). Unfortunately, those districts in Ohio opting out of the RttT competitive grant missed the funding benefit. Later, those same districts were exposed to the same accountability assurances that were established in Ohio’s approved application to receive President Obama’s 2011 ESEA Flexibility Waiver.

The ESEA Flexibility Waiver Program borrowed many ideas from RttT and expanded the application to offer relief from NCLB to states in exchange for aggressive reform (Riley, 2012). Ohio’s ESEA Flexibility Waiver resulted in House Bill 153, signed into law June 2011, and dramatically affected teacher evaluations in Ohio. Since OTES, legislatures have been continually updating the policy. The most recent policy change occurred when House Bill 362 passed on June 3, 2014 and changed the frequency of evaluations for teachers receiving “skilled” and “accomplished” ratings beginning in school year 2014-2015. The new policy also allows districts to choose between a new alternative teacher evaluation structure and the original structure (Ohio Department of Education, 2014). The Ohio Department of Education website provides links to a variety of new policy updates and useful sources that discuss Ohio’s progress with implementing new teacher evaluations (http://education.ohio.gov).

**Teacher Evaluation Policy Catalysts**

State and federal mandates are not alone in influencing teacher evaluation systems, as outside interest groups are entering discussions with their own opinions, ideas, and agendas. Although politicizing teacher quality breaches the walls of education, the interests of some and celebrity circles, political entities, non-governmental groups (Agranat, 2014), and philanthropists
are peaked by the notion of evaluating education systems and teachers. For example, teacher
evaluation reform development currently shares en vogue status among school reform celebrities
like Mark Zuckerberg and Oprah Winfrey (Agranat, 2014). These school reform celebrities join
philanthropist Bill Gates and Michelle Rhee, former chancellor of Washington D.C. Schools
turned non-profit education activist, and fund pet educational projects (Agranat, 2014).

Ravitch (2014), an educational policy analyst and former assistant secretary of education,
explained that teacher evaluation reached policy agenda status:

A few years ago, Arne Duncan, Bill Gates, David Coleman, and a merry band of policy
wonks had a grand plan. The non-governmental groups like Achieve, the National
Governors Association, the Council of Chief State School Officers, and Coleman’s own
Student Achievement Partners would write the Common Core standards (paid for by the
Gates Foundation); Duncan would require states to agree to adopt them as a condition of
eligibility for a share of the billions of Race to the Top funds at a time when states were
broke; the Feds would spend $370 million to develop tests for the standards; and within a
few short years the U.S. would have a seamless system of standards and assessments that
could be used to evaluate students, teachers, and schools. (Para. 1)

Hence, Bill Gates, Arne Duncan, and Michelle Rhee leveraged the necessary public influence to
command attention to the problem of teacher quality to ensure that teacher evaluation reform
would join the nation’s political agenda (Agranat, 2014). More significantly, Gates largely
influenced how that problem became perceived and defined during the initial stages of the
policy’s formulation.
Gates made a famous declaration regarding the policy formulation for teacher evaluation in a 2011 opinion piece published by the Wall Street Journal, when he called for a business model and the use of measurement to evaluate teachers (Morris, 2013). Bill Gates (2011) wrote,

At Microsoft, we believed in giving our employees the best chance to succeed, and then we insisted on success. We measured excellence, rewarded those who achieved it, and were candid with those who did not. Teachers don’t work in anything like this kind of environment. (Para. 3)

This Microsoft model, referred to as stacked ranking, categorized employees as top performers, good performers, average performers, below average performers, or poor performers. Morris (2013) assessed the design drafting of the OTES goals ultimately developed from the hundreds of millions of philanthropic Gate’s dollars used to persuade state and federal policymakers that the Microsoft stacked ranking model would also worked for public school teacher evaluation reform.

The Gates Foundation and stacked ranking policy formulation influences flowed directly into OTES reform policy through the 2009 eligibility requirements of President Obama’s RttT large grants. Shortly after voluntary adoption of teacher evaluation systems based on measured student achievement results through RttT grants, President Obama invited all states to apply for No Child Left Behind (NCLB) relief waivers. The 2011 NCLB Waivers, sometimes called ESEA Waivers, required assurances from the applying state education agencies to mobilize political support and formal enactment of teacher evaluations systems based on student achievement measures. Koppich and Esch (2012) examined this federal government-foundation nexus over 25 years of teacher effectiveness policy efforts and reported:
The philanthropic community has assumed a role not seen since the activities of the Ford Foundation in school finance reform in the 1970s. Interestingly, some of the same concerns that roiled the teaching policy agenda in the 1980s and 1990s are the focus of contemporary policy thrusts. In both of these periods, American economic competitiveness, or more accurately, fear of losing it, frame the teaching reform debate. In both periods, better and different teacher preparation, improved evaluation, and different forms of pay lie at the heart of teaching policy reform. It is worth noting, however, the most recent spate of policy efforts has had a sharper edge than earlier ones. The call for ‘teacher professionalism’ has fallen out of favor and in its place is the ‘get rid of the bad ones’ fervor. (p. 91)

Philanthropist Bill Gates, activist Michelle Rhee, and celebrity reformers highlight the generalized need for change agreements with national teacher unions, as they legitimize the need for teacher evaluation reform in the national media. However, the real thrust for the legislation implementation of OTES came from right, left, and center political think tanks that united to develop model legislation under the name of the American Legislative Exchange Council (ALEC; Agranat, 2014). The ALEC is among the largest producers of educational reform legislation. ALEC is hallmarked for creating legislation to include value-added modeling (VAM) in teacher evaluation. According to Agranat (2014), ALEC’s 2010 version of the Great Teacher and Leaders Act, combined with President Obama’s 2011 NCLB Waivers, created a flurry teacher evaluation legislation throughout the nation. ALEC’s Great Teachers and Leaders Act formulated the original 50% student growth measures and 50% teacher performance rating language found in the Ohio Revised Code that we use today to implement OTES.

Current Teacher Evaluation Policy
As an amalgamation of the aforementioned influences and discussions, Ohio’s House Bill 153 (Ohio Revised Code [ORC] 3319.112, 2011) codified Ohio’s teacher evaluation systems, requiring all districts to implement new teacher and principal evaluation policies that aligned with state-developed frameworks by July 1, 2013. The Ohio Teacher Evaluation System (OTES) is the current standards-based model framework designed by the Ohio Department of Education. The OTES’s policy design fulfilled the legislative requirements (Ohio Department of Education [ODE], 2011) resulting from House Bill 153, and became law in June 2011.

The OTES model’s framework relies on two key components consisting of 50% teacher performance ratings and an equally weighted 50% student growth measure factors. The teacher performance category is founded within the seven Ohio Standards for the Teaching Profession (2005). The Ohio Teacher Evaluation System (OTES) provides a rubric that delineates levels of performance ratings for each teacher using these two key components. The teacher performance level categories of “accomplished,” “skilled,” “developing,” and “ineffective” comprise the rating scale. These ratings are determined through two formal observations of at least thirty minutes each and walk-through observation data collected by evaluators. Each teacher receives a written report of the observations and walk-through observation feedback.

The student growth measure categories fall into the descriptive labels of “below,” “expected,” or “above,” and directly fulfill the requirement that teacher evaluation frameworks use multiple measures that include student growth on standardized tests and standards-based teacher performance (ODE, 2012). The teacher’s performance rating and the results of student growth measures produce a summative evaluation rating. In addition, House Bill 153 (ORC 3319.112, 2011) requires professional development for all developing and ineffective teachers to
facilitate professional growth. The policy also requires the allocation of financial resources to support each teacher’s professional growth and development.

According to the OTES Training Workbook (2012), OTES resulted from a collaborative design effort by Ohio teachers, school leaders, higher education representatives, and delegates from Ohio’s teacher unions, in collaboration with national teacher evaluation experts. The passage of HB 153 served as the state’s impetus to move OTES to the local board level policy adoption tables. HB 153 did not require districts to specifically adopt the OTES model (ORC 3319.112, 2011). Instead, HB 153 required districts to design or revise their teacher evaluation systems comply with the new sections of the Ohio Revised Code. The Ohio Department of Education provided criteria for school districts requiring stakeholder meetings when creating their own local teacher evaluation systems (ODE, 2012). The Ohio Department of Education (ODE) presented OTES as a fully legislative-compliant model that could be adopted by local boards of education, in consultation with teachers employed by the board.

The push to increase teacher effectiveness by transforming teacher evaluation systems throughout the nation occurred rapidly, and many of the states’ implementation timelines were too aggressive to realistically and authentically engage stakeholders in teacher evaluation reform (Behrstock-Sherratt, Rizzolo, Laine, & Friedman, 2013). Pennington (2014) affirmed that ninety-two percent of the districts in Ohio opted to use the state’s teacher evaluation reform model rubric and did not elect to engage teachers and administrators in revising or re-creating teacher evaluation rubrics and/or observation tools. This ambitious push to reform teacher evaluation, combined with growing evidence that no single best approach to measuring teacher effectiveness existed, led to many challenges (Behrstock-Sherratt et al., 2013) in implementing the new generation of teacher evaluation systems.
The OTES’s implementation presented school principals with challenges and conflicts already inherent in teacher evaluation, and then were arguably intensified with teacher evaluation reform efforts (Conley & Glasman, 2008). According to Porter, Lawler, and Hackman (1975), conflict and fear are natural pressures when professionals give and receive valid feedback about performance and performance expectations. Recent OTES policy demonstrated the perspective of Strike and Soltis (2009) concerning consequentialist and non-consequentialist ethical theories related to the dual frameworks of student growth measure rating usage in personnel decisions and the teacher performance observation-based rubric. According to the original OTES, half of a teacher’s overall rating was based upon standardized tests or student growth measures. The student growth measure component of OTES policy was clearly correlated to Strike and Soltis’ results-oriented consequentialist theories; hence, school leaders could utilize test results as an influential method of ousting ineffective teachers from the profession. A clear picture of Strike and Soltis’ consequentialist framework flows throughout the portion of OTES policy that combines the student growth measures and teacher performance observation-based rubric to create a final summative rating that is ultimately designed to influence major personnel decisions such as retention, tenure, and dismissal.

On the other-hand, OTES policy also establishes provisions emphasizing professional-development support systems. Strike and Soltis’ (2009) path to non-consequentialist perspectives relied upon observation-based teacher performance rubrics as a means to support the development of teachers’ professional skills. Strike and Soltis’ non-consequentialist perspective clearly relates to a recently released position paper by the National Council of Teachers of English (NCTE) on teacher evaluation. The NCTE (2012) detailed an explanation of professional-development-based accountability including that the overall professional
development of teachers at every stage be aligned with high standards of practice established by the teaching profession and the ability to exit teachers who show no interest in the hyper demands of the profession. New OTES policy stands upon two equal and opposing ethical theories combining reliance on student growth measured accountability, or consequentialism, and the equal reliance on observation-based professional development provisions of support, or non-consequentialism. These opposite and competing accountability and support frameworks can be difficult for evaluators to ethically balance. In addition, the two opposing frameworks feed into the research, supporting the ongoing historic record of fear and distrust between principals and teachers (Murphy, Hallinger, & Heck, 2013).

To measure and assess OTES policies, the ODE holds contracted partnerships with outside entities. It’s partnership with Ohio Education Research Center (OERC) provides for collaboration among universities and research organizations that connect education policy for Ohio schools and provide OTES implementation data to ODE. Franco, Zigler, and Lindsey (2013) released an OERC initial investigation of OTES and offered the following four generic common themes among the 2012-2013 first year district implementers: a) positive responses to evaluation systems, b) lack of training and preparation, c) issues with student growth measures, and d) feelings of stress.

Ohio’s Department of Education has also partnered with the American Institute for Research (AIR) and RANDA Solutions to assist it in crafting a larger and more significant design to measure and assess OTES policy implementation (Pennington, 2014). RANDA Solutions, a software firm specializing in educational markets, created Ohio’s electronic Teacher and Principal Evaluation System, or eTPES. The online tool met ODE’s requirement that all districts upload summative evaluation data to the state by the May 1st mandated deadline of each
The collaborative support of AIR and RANDA’s eTPES provided ODE with the ability to conduct systematic statewide teacher evaluation audits (Pennington, 2014). ODE (2012) discussed future audit intentions during evaluator trainings. Pennington (2014) further expected ODE to search for districts with extremely high teacher evaluation data, or at the lowest levels of the evaluation system.

**Teacher Evaluation New Implementation Trends**

Doherty and Jacobs (2013) contended that policymakers and reformers inferred that because states and districts adopted new teacher evaluations with a stronger emphasis on student achievement gains, extremities in teacher evaluation results at the highest levels would predictably cease to exist. This assumption has already been proven incorrect by the release of early findings using new generation teacher evaluation systems. Signs of the extremities Pennington (2014) claimed states and districts should be targeting showed. According to Aldis (2014), early results from states like Florida, Delaware, and New York, represented in Table 1, suggested the new generation teacher evaluation systems failed to accurately identify which teachers are making the biggest difference for students and which were struggling.

**Table 1**

New Generation Early Implementation Teacher Evaluation Results from Three States

<table>
<thead>
<tr>
<th>State</th>
<th>Percentage of Teachers Rated Effective or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delaware</td>
<td>99%</td>
</tr>
<tr>
<td>Florida</td>
<td>97%</td>
</tr>
<tr>
<td>New York</td>
<td>95%</td>
</tr>
</tbody>
</table>
Aggressive state involvement in local district teacher evaluations is a recent phenomenon that has hindered many states’ ability to effectively dissolve local education agencies’ historical autonomy of teacher evaluation systems and forge new symbiotic systems (Geo, Holdheide & Miller, 2011). States vary significantly in their approaches to seamlessly integrating their influence on local education agencies to produce the federal teacher evaluation reform movement’s intended results (McGuinn, 2012). According to McGuinn (2012), state and local education agencies mirrored many of the same tensions between principals and teachers in relation to their fears and distrust specifically linked to the implementation of new teacher evaluation systems. McGuinn contended the tight focus of state education agencies on compliance and accountability resulted local education agencies inability to be candid about the issues they face in implementing new teacher evaluations and stifled local districts’ requests for assistance.

This restructuring of state education agencies’ roles also caused tensions between states’ desires to give local education agencies the flexibility to select and adapt evaluation instruments; and, the states’ low capacity threshold to secure implementation supports for individual districts for a variety of evaluation instruments (McGuinn, 2012). Geo, et al. (2011) presented evidence pertinent to this tension:

For example, several states codified the weight (percentage) of student achievement in the teacher evaluation system (e.g., Tennessee specified 50 percent, Rhode Island specified 51 percent, and Colorado specified 50 percent). Such state legislation was intended to drive changes in evaluation systems and provide better information about teachers’ contributions to student learning growth. However, the legislation often did not address the other logistical and procedural aspects of teacher evaluation. States should
play a critical role in interpreting such legislation and be prepared to help districts address
specific challenges, unintended consequences, and implementation considerations at the
district level. (p. 5)

Although the lack of available research-based methods and models of comprehensive teacher
evaluation thwarted states’ abilities to offer assistance to districts, the federal policy continued to
push for reforms that force states to define their roles and intensity of involvement within the
new teacher evaluation implementation (Geo et al., 2011).

Ultimately, this federal policy’s lean on state agencies mirrored the policy lean from state
agencies to local education agencies. McGuinn (2012) contended the issues surrounding distrust
and fear extend between teachers and principals, local education agencies and state education
agencies, and further echoed between the U.S. Department of Education and state education
agencies. For example, McGuinn (2012) offered candid reports of state education agency
leaders participating in top executive-level networks who were honest about the challenges states
faced in reinventing state roles in teacher evaluation implementation. Later, state agency leaders
experienced unwanted attention, intervention, heightened scrutiny, and reporting requirements
from the federal government to assist the states with the new expected federal role
transformation (McGuinn, 2012). McGuinn further contended that local education agencies,
state education agencies, and the U.S. Department of Education needed to acknowledge that the
new teacher evaluation implementation inevitable involves mistakes. McGuinn (2012)
advocated that these mistakes deserved shared attention at all levels of the education delivery
chain without fear of punitive interventions from higher authorities.

Geo et al. (2011) provided a body of research that defines high authority roles and state
responsibilities in teacher evaluation systems. In their “State-Level Evaluation System,” the
state provides a strict interpretation of legislation package and all requirements of each of the components. The identified strength in this system includes increased system reliability and standardized implementation and data collection. However, the state-level system diminishes local flexibility, ownership, and accord because it fails to consider local contexts and issues.

According to the Geo et al.’s (2011) model, states significantly dictate certain aspects of local district models but do allow local decisions in other parts of the evaluation system. Areas of strength in this system include data collection, implementation, and reliability in the required components of the evaluation system. The state supervision and management provide challenges for states and understanding teacher evaluation results proved extremely difficult for these states (Geo et al.’s, 2011).

In Geo et al.’s (2011) “District Evaluation System with Required Parameters” model, states only provide some guidance to districts and set the parameters for districts models. In addition, states also provide some type of screening or approval processes to ensure compliance with district models. States also audit district systems to ensure operations are within the state-defined parameters. Because local-level agencies are able to prioritize certain aspects of the model to fit the specific context of their local system, local school districts are more apt to claim ownership of the evaluation system and therefore work in accord with the state. However, with this model, reliability is vulnerable across school districts, and it is difficult to compare progress and the result of data aggregation (Geo et al., 2011).

Practical state implementation models following the aforementioned Geo et al. (2011) can be found in Delaware, New York, and Ohio’s models. For example, Delaware simulates a state-level evaluation system, New York’s system mirrors the elective state model evaluation system, and Ohio’s evaluation system for teachers and principals is akin to the district evaluation
according to Doherty and Jacobs (2013), Ohio and 38 other states retain model-design discretion, with 15 of those states requiring state review and authorization of local education agencies’ evaluation systems. Specifically, in Ohio, every local system employs the student growth measure component as the determining element of a teacher’s effectiveness rating. For example, if a teacher is rated skilled in all aspects of practice in his or her classroom observations but rated below in student growth, then the teacher’s overall rating would be ineffective.

In October 2013, the National Council on Teacher Quality (NCTQ) released a report focused on the progress made toward the adoption of teacher evaluation models and policies by all 50 states and the District of Columbia Public Schools (DCPS). The October 2013 NCTQ report specifically looked for how states used information gained from new teacher evaluation systems to discuss teacher policy and classroom practice with their teachers (Doherty & Jacobs, 2013). According to Doherty and Jacobs (2013),

In 2013, it is clear that states are developing multiple measures to assess teacher performance. Almost every state (44 and DCPS) now requires that classroom observations be incorporated into teacher evaluations. In 24 of those states and DCPS, multiple annual observations are required as part of each evaluation for at least new, if not all teachers. Twenty-one states and DCPS provide specific guidelines for when classroom observations should take place during the year, and 14 states and DCPS require that at least some classroom observations are unannounced. Twenty-one states and DCPS are explicit that teachers receive feedback on classroom observations. (p. 14)

Doherty and Jacobs (2013) reported that Ohio’s new teacher evaluation model policy only met three out of eleven areas for connecting teacher evaluation data to inform teachers
about their practice. For example, Ohio showed evidence of connecting teacher evaluation data to inform policy on improvement plans, layoffs, and preparation program accountability (Doherty & Jacobs, 2013). Conversely, Doherty and Jacobs revealed that Ohio did not clearly link teacher evaluation data to professional development, public reporting to aggregated teacher ratings, compensation, dismissal, licensure advancement, licensure reciprocity, and student teacher placements. Doherty and Jacobs biggest policy recommendations for Ohio included the following:

- Base tenure decisions on evidence of classroom effectiveness, rather than the number of years in a classroom;
- Ensure that districts use teacher evaluation results to determine professional development needs and activities;
- Make aggregate school-level data about teacher performance publicly available to shine a light on how equitable teachers are distributed across and within school districts;
- Develop compensation structures that recognize teachers for their effectiveness;
- Specify that classroom ineffectiveness is grounds for dismissal so that districts do not feel they lack the legal basis for terminating consistently poor performers;
- Require evidence of teacher effectiveness to be a factor in determining whether teachers renew or advance their licenses;
- Make evidence of teacher effectiveness the basis for granting licenses to out-of-state candidates, especially for those who come from states that make student growth a significant factor in evaluations; and
• Place student teachers with cooperating teachers and seek evidence that they are effectively teaching student.

Ohio’s teachers and principals are on the front line of this barrage of recommendations concerning new teacher evaluation implementation. A recent study by Franco et al. (2013) investigated 12 Ohio districts’ implementation plans and found that Ohio’s teachers generally offered positive feedback characterized by a united approach. Teachers also indicated that a useful goal-setting process provided opportunities for discussion about teaching and student growth, and the system seemed designed to focus educators on specific areas of professional development. The teachers interviewed focused their comments on the use of Ohio’s standards-based rubric and reported that they felt the rubric accurately represented the work of teachers and administrators. Furthermore, Ohio teachers who were interviewed expressed acceptance of growth measures as an important component of evaluation. Teachers also expressed that because the principals’ evaluation systems similarly use student growth measures, a sense that “we are in this together” (Franco et al., 2013) surfaced.

Franco et al. (2013) uncovered trends of Ohio’s teachers’ concern about principals not being as visible and available to them due to the increased time demands of the observation cycles. The researchers also revealed that many educators expressed great concern centered around the public perceptions associated with the color-codes and labels of each of the rating categories and noted relief that the labels and categories were adjusted to reflect skilled as pale green and developing as a deep yellow (Franco et al., 2013). An additional theme of educator concerns manifested pertaining to the documentation and observation processes that compelled teachers to focus on and overemphasize their individual teaching behaviors rather than focusing on students and their learning. The final concerns expressed by teachers addressed fears a) that
the school culture might shift from collaboration to competition, b) the knowledge gap would surface in the utilization of student growth measures, and c) a void surrounding how teachers without student growth measures might be evaluated with student achievement data.

Ohio’s teachers’ assertions that teacher reform is necessary and teachers’ positivity about goal-setting, which show as trends in Franco et al. (2013) study, were corroborated in another study by Donaldson (2012). Donaldson conducted 92 in-person interviews in a northeastern urban district in Ohio operating in its second year of implementation. Although Donaldson’s findings of teacher perceptions toward new evaluation systems were generally positive, teachers reported that their instructional practices remained unaffected by the new teacher evaluation systems. Interestingly, Donaldson contended that teachers rated lower by the new evaluation program reported higher rates of negative views of the program but were most likely to report the new evaluation program changed the way they taught and planned.

Kane, Kerr, and Pianta (2014) presented similar evidence in the Measurement of Effective Teaching (MET) Project, which extended data exploration of the relationship among value-added results, evaluation, and teacher type. Interestingly, these authors found that teacher types who rated their school leaders lower named isolated agnostics (delineated as having few interactions with colleagues and doubting expectations for students) produced higher versus lower value added results (Kane et al., 2014). Explaining the correlation between isolated agnostics and high value-added results with lower ratings of school leadership may or may not reside in the assumption that sustained improvement pressure from a supervisor causes both dislike and improved performance (Kane et al., 2014).

A discussion of the MET (2012) investigation of value-added relationships between lower school leadership ratings and high value-added results addressed a larger matter: school
leaders’ responsibility to engage in difficult conversations, provide honest feedback on performance, and create pathways for underinvested teachers to work collaboratively for improvement, even when faced with resistance or lower job satisfaction (Pianta, 2011).

**Teacher Evaluation and Professional Development**

Goe, Biggers, and Croft (2012) proposed that with purposeful planning states, districts, schools, and school leaders’ can use teacher evaluation to balance the opposing goals of accountability and support for improvement through professional development. Evaluators must balance applying trusted supports to help and sustained pressure to improve while simultaneously acknowledging the potential unreliability of test-based accountability evaluation systems. According to Conley and Glasman (2008), evaluators who fail to strike a critical balance between support and pressure tensions create environments in which teachers’ fears of poor summative evaluation results render them unable to be forthright about their professional development needs and in which evaluators avoid offering detailed feedback to teachers. The result of this environment creates an evaluation system with little concern for professional growth and a focus on evaluation as a rhetorical exercise or political hoop jumping, ultimately marginalizing teacher learning (Conley & Glasman, 2008).

Blank (2009) and Desimone’s (2009) model frameworks of the effect of professional development student achievement conceptualized the transfer of teacher learning to student learning. Blank’s logic model and Desimone’s conceptual framework both emphasize similar characteristics that produce positive outcomes from professional development and insight into teacher learning and student learning. Still, there is a lack of quality studies on professional development programs and their impact on student achievement in the research (Kane et al., 2014).
Blank’s (2009) findings lend support to the claim that the education field has too few well-designed studies on the connection between teachers’ professional development and the subsequent increase in student learning. For example, previous research findings from Yoon, Duncan, Lee, Scarloss, and Shapley (2007) found that only nine of 1,300 studies examined the effects of professional development on student achievement and met research qualifications for scientific evidence. Blank (2009) also found that only 16 of 74 studies met inclusion criteria for a meta-analysis. Moreover, a wide variety of analytical approaches to studying the relationship between teachers’ professional development and student achievement have created specific challenges for researchers to summarize studies across wide-ranging techniques (Kane et al., 2014).

While it is true that most researchers acknowledge the extensive challenges in studying and summarizing the effect of professional development on student achievement, both Yoon et al. (2007) and Blank’s (2009) meta-analysis found validity in the links between teachers’ professional development and student achievement (Kane et al., 2014.) Yoon et al. (2009) supported the claim that teachers’ professional development improved student achievement by quoting an actual value (i.e., average of 49 hours) that describes substantial professional development. Yoon et al. measured 20 effect sizes and found that the average increase across science, math, and English language arts showed a standard deviation of 0.54. These results of Yoon et al.’s study contradicted Blank’s meta-analysis, specifically in regard to the remarkably high effect size. However, Blank’s (2009) meta-analysis of 16 studies with average effect sizes of smaller results proved consistent with Yoon et al.’s study.

The works of Yoon et al. (2007) and Blanks (2009) clearly demonstrate a valid link between professional development effects on teachers and student achievement. However, their
studies did not completely support the link among professional development and evaluation outcomes and student data related to achievement. Similarly, Kane et al. (2014) explained the critical concern that identifying quality empirical evidence of teacher professional development and student achievement stems from the focus of most studies examining professional development in the context of workshops. Kane et al. further contended that the lack of quality empirical evidence of teachers’ professional development and student achievement, versus studying the effects on classroom observations, teacher skills, or reporting pre- and post-intervention scores of student achievement, limits researchers’ ability to summarize effect sizes across studies.

The current literature on linking teacher professional development and teacher evaluation bloomed with recent examples of highly structured observation protocols employed to assess teacher performance, thus forming the basis of professional development recommendations (Kane et al., 2014). Unfortunately, relatively few studies on observation-based or evaluation-based professional development have been conducted. The MET (2012) study investigated the effect sizes for links between teacher practices and student learning using observation rubrics. According to the MET (2012) study, the largest effect size for the relationship between teacher practices and student learning produced an approximately 4% point difference in student learning and can be attributed to teacher practices. The MET (2012) study investigated the relationship between teacher practices and student learning using similar observation rubrics; however, the study design did not employ randomized controlled trials and did not provide observation-based feedback or professional development to teachers (Kane et al., 2014).

Experimental data from another study involved the MyTeachingPartner program. This study suggested teachers could improve their observation scores through formal training on the
classroom assessment scoring system (CLASS) observation protocol domains and coaching
sessions that focus on very short videos of the teacher teaching (Allen, Pianta, Gregory, Mikami,
significant changes that led to improved student achievement. The MET (2012) and Allen et al.
(2011) studies found that the results of studies on professional development outcomes and
student achievement fluctuated widely. Marshall (2013) substantiated Allen et al.’s claim that
fluctuating student achievement outcomes resulted from the implementation of new evaluation
variances. Marshall further contended schools must radically improve the synergy of teacher
supervision, curriculum planning, interim assessments, and teacher evaluation to improve teacher
learning and student learning. Even then, the research only ambiguously indicated that the
processes of observing and evaluating teachers actually improved effective teachers (Marshall,
2013).

King (2014) supported Kane et al.’s (2014) contention that researchers should
acknowledge issues related to demonstrating the link between teachers’ professional
development and student outcomes. Wayne, Yonn, Zhu, Cronen and Garet (2008) argued that it
is “best to initially study the impact of PD intervention on teacher knowledge only [because] to
have an impact on student achievement of a detectable magnitude, the impact on teacher
knowledge must be quite substantial” (p. 476). Coggshall, Rasmussen, Colton, Milton, and
Jacques (2012) advocated that although federal and state policies and guidance documents refer
to teacher learning as professional development rather than job-embedded professional
development, the policies were clearly created to capture teacher evaluation systems for
continuous improvement of teaching effectiveness demonstrated by outcome student
achievement results.
Coggshall et al. (2012) maintained that there is little evidence about how the new
generation teacher evaluation systems work to improve teacher practices because they are still so
new. Evidence regarding three trends in extant studies focused on using teacher evaluations to
increase teacher effectiveness surfaced in the literature review. First, teacher evaluation
promotes teacher learning through a shared understanding of effective teaching (Coggshall et al.,
2012; Geo et al., 2011; Goe, Biggers & Croft, 2012; Sartain, Steoelinga & Brown, 2011).
Secondly, evaluation acts as a catalyst for teacher learning through evidence-based feedback
(Coggshall et al., 2012; Goe et al., 2012; Hill & Herlihy, 2011; Marshall, 2013). Third,
evaluation promotes teacher learning through assessing professional development and promoting
collaboration (Coggshall et al., 2012; Geo et al., 2011; Goe, Biggers & Croft, 2012; Hill &
Herlihy, 2011). Aligning these high levels of teacher learning opportunities to teacher
evaluations depends on the implementation drive and learning conditions across schools,
districts, and states (Coggshall et al., 2012).

Goe, Biggers, and Croft, (2012) developed a research and policy brief designed to help
states and districts strategically use the results of teacher evaluation processes to assist with
professional growth decisions and opportunities. Geo et al. (2012) proposed that evidence
collected for accountability equally possessed useful qualities to determine professional growth
and development opportunities. Geo et al. developed an informal framework of six components
necessary for alignment of teacher evaluation and professional development systems. The
framework emphasized the central work of ensuring teachers, and therefore students, and the
benefits of an evaluation system specifically designed to align evaluation and professional
development.
Geo et al.’s (2012) six linked components—teaching standards, multiple measures of teacher performance, high quality training on standards, trained individuals to evaluate and provide feedback, professional growth opportunities, and professional learning standards—demonstrated critical elements in the alignment of teacher evaluations to professional learning outcomes for teachers. This alignment required daunting, complex, and transformational work between states, districts, and schools. A closer inspection of this assumption by Minnici (2014) suggested that insufficient thought and time allocations have been targeted toward shifts in school culture, school leader and teacher behaviors, and the new knowledge and skill acquisitions required to do this intensive work.

The largest proportion of the studies supported the six-element premises of the Geo et al. (2012) informal framework as essential in any teacher evaluation reform implementation (Archibald, Coggshall, Croft, & Goe, 2011; Bill & Melinda Gates Foundation, 2010; Council of Chief State School Officers, 2011; Kane & Staiger, 2012). Implicit in the notion of the development of the six components for aligning evaluation and professional development rests the basic assumption that the urgency states and districts across the nation face for teacher evaluation reform has shown the need for further research on teacher evaluation (Geo et al., 2012; Gitomer et al., 2014). The present research continues to rest repeatedly on the extended assumption that states and districts stand at the starting gate of implementation challenges and face the necessary deep shifts in culture and beliefs that remain on the distant horizon, if not completely out of focus.

**Teacher Evaluation and School Improvement**

The literature review reveals a diversity of opinions within the research community regarding the relationship between new generation teacher evaluations and school improvement.
Marshall (2013) contended that the growing consensus for using teacher evaluation reform in school improvement efforts stems from the research stating that quality of instruction is the single most important factor in student achievement. Papay (2012) stated that quantitative research supports the popular notion that teachers are the most important school-level factor in promoting student achievement. However, studies show that there are wide variations in teacher’s ability to raise student achievement (Aaronson, Barrow, & Sander, 2007). According to the NCTQ (2011), teacher evaluation reform has become the primary focus in the quest to secure an effective teacher in every classroom. In fact, teacher evaluation currently ranks over other innovative school improvement education initiatives to increase teacher effectiveness, student achievement, and school improvement (NCTQ, 2011). On the other hand, some researchers take the position that teacher evaluation acts as a low-leverage strategy to improve teacher effectiveness, student achievement, and school improvement (Dufour & Mattos, 2013; Fullan, 2011).

A study by Maslow and Kelley (2012) investigated how teacher evaluation data advanced systemic teaching quality. The researchers conducted a cross case analysis to identify themes in teacher evaluation practices in seven high schools. Maslow and Kelley looked specifically for ways teachers and administrators use teacher evaluation practices to promote system learning, feedback, and increased performance. Most notably, they found that schools using teacher evaluation effectively surfaced as trends of shared vision, shared values of good instruction, and strong cultures that embrace teacher collaboration with an emphasis on continuous improvement. This study also suggested that school cultures that predominantly attend to student behaviors do not have the capacity to implement teacher evaluation practices that promote high levels of feedback utilization. These outcomes can offer present day limits and opportunities for new
generation teacher evaluation systems and their capacity to promote school improvement and self-reflection in teaching behaviors.

According to Emstad (2011), post-conferences targeted at self-reflection are a powerful device through which teachers can link teaching behaviors with student achievement outcomes. Consequently, more attention can be given to teachers’ perceptions of critical elements and reflections, which can be addressed during the pre-observation and/or post-conference. Range, Young, and Hvidston (2013) investigated predictors for reflection within initial-contract teachers in relationship to beliefs about the importance of the post-conference versus the pre-conference session. Range et al. (2013) found that predictors of reflection in non-tenured teachers were influenced by the following seven variables: trust, sharing the observation report, teacher’s lesson focus, constructive feedback, identified standards, sequence for conference, and growth areas. Interestingly, Range et al. asserted that tenured teachers only view two variables, linking professional development to the post-conference and the provision of positive comments by the principal, as predictors of reflection. These findings better enable principals to use formative supervision, based on proactive processes, to impact teachers’ instructional practice (Marshall, 2013; Range, Young, & Hvidston, 2013).

A recent study by Hallinger, Heck, and Murphy (2014) investigated the relationship between school resource allocations and teacher evaluation. Specifically, the study investigated allocated resources such as additional teacher and principal time, increased paperwork, and financial performance incentives, which were used for new teacher evaluation, and the effects that these resources had on school improvement results. This study was empirical and employed a research review methodology. An exhaustive search conducted on Google Scholar included all relevant sources in the United States. The authors categorized the findings, designs, and
conceptual models for each relevant study. A critical evaluation study focused on pattern conceptualization, methods, and analysis of trend sets across studies and statistical models. A critical evaluation of the research literature regarding the capacity of new generation teacher evaluation systems to produce school improvement efforts yielded limited evidence.

Hallinger et al. (2014) presented a strong argument that the new generation of teacher evaluation models encompasses layers of past and present accountability research concerning the strength of teacher effects on student learning. Additionally, Louis, Dretzke, and Wahlstrom (2010) accumulated research that places teachers at the focal point of school improvement efforts. Hamilton (2011) emphasized research revealing that workplace conditions within schools need to dramatically change before linkages between teacher evaluation and school improvement can be clearly defined. The growth of the corporate human resource culture in public education moved these new generation teacher evaluation systems to center stage in the effort to improve education (Hallinger et al., 2014).

**Teacher Evaluation and Human Capital Management Systems**

The ideas expressed by researchers investigating the links between new teacher evaluations, professional development, and school improvement have led to broader conceptualizations of the link between new teacher evaluations and human resource management. Konstantopoulos (2012) reviewed the research on teacher effects and contended that lack of clarity in teacher-effects research should inhibit human capital-related decisions that could seriously affect teachers’ employment and pay.

Master (2014) fostered questions about the predictive relationship between teacher evaluation ratings and principal personnel decisions. Master collected data from public charter schools rated as highly effective that exist alongside state public schools under a single
centralized district management team. The charter district served 5,000-10,000 students. District data included teacher characteristics, personnel decisions, and teachers’ evaluation ratings over a three-year period. Five hundred six teachers were included in the study. A principal component factor analysis conducted across initially standardized individual indicators identified a single heavy loaded factor for later standardization. Master explored school-wide distribution averages of teacher ratings to see if some schools’ teachers were rated higher or lower by a group. This study involved two separate logistic regression models—one with a single evaluation characteristic and a vector of teacher demographic specifics and one with just a vector of teacher demographics. Master employed exploratory factor analysis, multiple-factor prediction of personnel decision, and a linear regression model to predict each teacher’s performance rating.

With few exceptions, Master (2014) found that teacher evaluation ratings strongly predicted the charter district’s personnel decisions. In addition, results emphasized the different characteristics that administrators’ value and the fact that these characteristics prioritize differently among varying teacher assignments. Master advanced the notion that administrators’ formative evaluation results predicted future personnel decisions. Importantly, the researcher went far beyond attendance rates, value-added measures, and the generic behaviors or characteristics of the teachers that administrators find valuable. However, the limitations of this study are found in the charter characteristics of the study’s setting. The dynamics of charter school workplace culture and norms differ greatly from K-12 public environments; thus, differences in cultural norms make this study’s findings exclusively applicable to charter schools. Yet this study could offer merit for further research into deconstructing the teacher characteristics valued by administrators. The logical aspects of discerning teacher characteristics
that administrators find valuable trumps arguments regarding generalizability due to its exclusivity to charter schools (Master, 2014).

Behrstock-Sherratt and Jacques (2012) presented a research brief on aligning professional development policies with teacher evaluations to increase comprehensive approaches to human resource management. Behrstock-Sherratt and Jacques illustrated how the connection between an aligned evaluation system and professional development system facilitates a broader connection to form a comprehensive human capital management system (HCMS) that empowers an evaluation system to promote systemic and organizational change efforts (Behrstock-Sherratt & Jacques, 2012). A fully aligned HCMS reaches the entire spectrum of teacher-effectiveness polices, from conception to retirement, and includes key stages such as pre-service, certification, recruitment, hiring, onboarding, leadership, standards for exiting teachers, and equitable access to quality teachers (Behrstock-Sherratt & Jacques, 2012).

The NCTQ (2011) extended Behrstock-Sherratt and Jacques’s (2012) research. The NCTQ scrutinized the capacity of states to provide a comprehensive state evaluation system fully aligned to professional development and HCMS. The NCTQ also substantiated that most states are in the infancy stages of aligning teacher evaluation and professional development to architect a comprehensive HCMS. The controversial challenges of implementing OTES, or any other state’s evaluation model, requires principals to consciously apply an ethical lens to the decision-making processes that link teacher evaluation to professional development, systemic school improvement efforts, and human resource management (Behrstock-Sherratt et al., 2013).

**Teacher Evaluation Tools and Roles**

Even the best professionals may attribute their evaluations to luck, if they believe the observation tool used is not a valid measure of their teaching (Lavigne, 2014). Low confidence
levels in the validity of observation tools by administrators or teachers creates the potential for the teacher evaluation systems to become rituals and diminishes opportunities to improve instruction, schooling, professional development, and personnel decisions (Ramirez, Clouse, & White-Davis, 2014).

A lack of clear instructional observation tools that help define the critical aspects of teaching that lead to improved practice may be closely related to the history of highly undifferentiated teacher evaluations (Gitomer et al., 2014). Findings provided by preceding studies (e.g., Hill, Charalambous, & Kraft, 2012) widely documented the conceptual aspects of the multiple sources of variance in observation scores and supported Gitomer’s et al. (2014) claim that the tradition of highly undifferentiated teacher evaluation scores resulted from evaluator uncertainty in judging intricate instructional components that led evaluators to focus on readily observable themes of organization. Gitomer et al. further contended that organizational themes legitimately lead to higher ratings more often than accurate judgments of deep instructional concepts. Gitomer et al. challenged Weisbergs et al.’s (2009) theoretical assumption that historically high numbers of undifferentiated teacher evaluation results stemmed from a national lack of administrative willingness to act on the differences in teaching effectiveness.

Hill and Grossman (2013) proposed that the connection between teacher evaluation and improved instruction does not automatically appear as improved value added scores for teachers who operate alone and have agnostic expectations for students. Preliminary studies on feedback and coaching-based observations (Allen et al., 2011) have provided hope for using evaluation systems that involve sophisticated observation tools to improve teaching quality and student outcomes (Hill & Grossman, 2013). Unfortunately, the optimism surrounding teacher evaluation
observation tools has been dampened by findings that change rarely results from any large-scale education reform requiring more than regulatory changes juxtaposed onto existing routines and practices (Hill & Grossman, 2013).

Hill and Grossman’s (2013) research found that systems designed to improve teacher evaluation must contain the following three features:

- Teacher evaluation systems using observation tools must make subject-specific observation instruments that provide tangible guidance on desirable teaching practices available to teachers.
- New teacher evaluation systems must draw content experts within districts into the process of teacher evaluation and concentrate expertise on improving instruction.
- States and districts need to design systems that foster accurate and usable feedback from observations specifically designed to improve instruction.

This system prioritizes teacher learning over accountability and requires additional elements not found in the original teacher evaluation reform policies (Hill & Grossman, 2013). Therefore, Hill and Grossman (2013) contended that the infrastructure of improvement systems must complement the existing framework of teacher evaluation.

Gitomer et al. (2014) documented a fundamental challenge in using teacher evaluation observation tools to improve teaching and learning in their study of 82 introductory-level algebra teachers in middle and high schools. Five raters scored previously-coded live and recorded video observations. The video observations received double coding, and researchers assigned scores to ten dimensions, for example, instructional technique, emotional supports, classroom organization, and classroom environment. The observer outcomes were placed against the master coders’ outcomes. Lastly, the participating teachers filled out a self-report questionnaire,
Classroom Assessment Scoring System (CLASS-T), to assess their own skills on CLASS-S observation protocol dimensions. Each lesson produced data aggregated into three specific domain scores for emotional support, classroom organization, and instructional support; then, the averages of these domain scores were calculated across lessons to create scores for each classroom. Pearson correlations were used to provide an understanding of the relationship between classroom-level scores (CLASS-S), observations, and self-reported (CLASS-T) scores. The raters showed the highest agreement on the classroom observation domain. Additionally, teacher self-reports related positively to the same classroom observation domain. The domain most out of rating sync was instructional support, shedding light on the need to further examine the instructional support needs of both teachers and principals.

Gitomer et al. (2014) exposed both the complexity involved in and the potential opportunities of aligning teacher evaluation observation data and professional development through an observation-evaluation tool. These authors’ findings suggest that a shared definition of quality instruction remains unattained. The ability for teachers to rate themselves consistently across different dimensions, for example, instructional technique, instructional supports, and/or classroom organization, by utilizing the same observation tools as the principals employ creates challenges in the evaluation system (Gitomer et al., 2014). Gitomer et al. (2014) acknowledged the need to acquire the missing language to clearly define “improvement” that lends to teachers’ receptiveness to the feedback and personalized professional development provided by principals.

Papay (2012) similarly argued that, in addition to being reliable and valid, new teacher evaluation systems should also increase principals’ support capacity, thus leading to improved teaching quality and development. Papay examined value-added measures and standards-based evaluations and emphasized using new teacher evaluation systems in a wider context to increase
teacher instructional capacity. Papay (2012) contended that teacher evaluations from standards-based observations are a better predictor of student performance than teachers’ value-added ratings. Although there has been relatively little research on the reliability of standards-based observation tools to predict student performance on standardized tests, Papay’s argument offered new insight. Additionally, corroborating the views of Papay (2012), Gitomer et al. (2014) assert that observation tools lend to increasing both student learning and instructional capacity.

Whitehurst, Chingos, and Lindquist (2011) agree with Papay’s (2012) position that observational classroom data should play a larger role in the evaluation system and deserve more research attention. Whitehurst et al. studied four urban districts’ use of classroom observations as a part of the overall teacher evaluation system, which used a mixed-methods approach. The researchers used individual student achievement data that were linked to individual teachers from four urban districts with enrollment sizes ranging from 25,000 to 110,000 students. The authors provided an empirical analysis of the four districts’ teacher evaluation data, classroom observation data, and student achievement data. The authors correlated teacher evaluation and student achievement data to find the predictive relationship between the two data sources. Overall, the authors found that each of the four districts’ evaluation systems operated within an acceptable range of reliability and validity, as defined by previous teacher evaluation research. Whitehurst et al. (2011) also expressed optimism toward the new teacher evaluation systems’ capacity to measure teacher performance. Whitehurst et al. (2011) further stated that the technical work on developing classroom observations makes up a much smaller body of research than do value-added measures.

Hill, Charalambous, and Kraft (2012) extended the research by investigating the observational systems of mathematical quality of instruction (MQI) instrument ratings for
reliability and validity of teacher evaluation results. The researchers sought to uncover issues about the design and improvement of observation systems versus investigating specific teachers’ scores on formal evaluations. Hill et al. (2012) found that the generalizability theory was a significant asset in developing observational systems. Reliable observation systems are at the core of a legitimate foundation of instructional feedback and/or professional development provisions for teachers resulting from evaluation data (Hill et al., 2012). Hill and Grossman (2013) also delineated observation challenges as needing further research and cautioned that a one-size-fits-all evaluation system inevitably diminishes instructional substance.

Range et al. (2013) highlighted the unintended consequences of past one-size-fits-all classroom observations and the present day push to ground observation protocols into clearly articulated models of effective teaching wreaking havoc on the implementation of new evaluation designs. This pattern has existed among states in which many state agencies opted to customize a single existing instrument, or worse, completely recreated a new instrument from scratch (Hill & Herhily, 2011). The literature review revealed consensus on the importance of moving away from a generalized vision of good teaching to an evidenced-based observation protocol that drills deeply into the pedagogical content knowledge necessary to increase student achievement (Hill & Herhily, 2011; Range et al., 2013; Whitehurst, Chingos & Lindquist, 2011). A discussion regarding old and new classroom observation protocols must address how the research field has attended to rater bias. The literature on classroom observations overwhelmingly suggests that the characteristics of observers, teachers, classrooms can expose raters to bias (MET, 2012; Bill & Melinda Gates Foundation, 2012).

The MET (2012) project examined strict scoring systems in teacher evaluation, rigorous evaluator and teacher training, calibration efforts, certification requirements, and evaluator
monitoring efforts to reduce rater bias among the observers. According to Doherty and Jacobs (2013), replicating this rigorous training was a huge undertaking for most states, even though they recognized the critical element of evaluator training in observation and standards-based evaluation. Few states have implemented the measures required to secure the quality training evaluators need. For example, Doherty and Jacobs (2013) found only 13 states require a certification process for their evaluators. Ohio belongs to this select group of states taking a new approach to classroom observations.

The literature review showed indisputable support for Danielson’s framework for teaching, the CLASS, and newer generation protocol instruments that measure complex competencies (Hill & Herhily, 2011; Range et al., 2013). New approaches to evaluate teacher performance through the use of highly structured and stringent observation protocols in hopes of crystalizing useful professional development recommendations (Kane et al., 2014). The MET study (2012) initially employed a huge scale study that used complex observation rubrics that allowed researchers to examine the impact of professional development. Kane et al. (2014) reported that the effect size for the association between teacher practice and student learning seen in MET was small and estimated a four percent difference in student learning due to teacher practice after professional development.

**Teacher Evaluation and Teacher Development Platforms**

New generation teacher evaluation systems acted as a catalyst for technological support and on-demand, web-based professional development platforms that enhance principals’ capacities to use the observation process to personalize professional development (Watters, 2012). Online capabilities are developing at a remarkable pace. Online capabilities include teacher evaluation tools that create customized evaluation rubrics; systems for teachers and
principals to collaboratively create goals; and forums for teachers, mentors, evaluators, and guests to share documents, videos, and resources. Jason DeRoner, CEO of TeachBoost (affiliated with ImagineK12), described his online teacher evaluation tool as a teacher development platform (Watters, 2012).

A recent study by Stax Inc. (2014) supports Alvoid and Black’s (2014) hypothesis that principals need technological supports that allow them to record, share, and synthesize instructional data. According to an April 2014 analysis by Stax Inc., a positive impact was measured for the new teacher effectiveness platform, TeachBoost, on teacher observations and professional development practices. The study quantified the time that the platform saved school leaders and confirmed that the teacher effectiveness platform facilitated higher frequency and more meaningful observations. Furthermore, Stax Inc. (2014) contended that principals demonstrated a higher capacity to use the information extracted from observations for coaching and decision-making.

The School Improvement Network commissioned EdNexus to study each state’s policy and funding procedures for providing meaningful teacher professional development linked to teacher evaluations (Culver & Hayes, 2014). School Improvement Network conducted a nationwide survey focused on the professional learning components of states’ teacher evaluation policies. All 50 states’ departments of education were contacted by phone and/or email from August through November 2013. According to Culver and Hayes (2014), all but four states’ (Colorado, Idaho, Illinois, and Wyoming) teacher evaluation systems included a professional learning provision for teachers. Culver and Hayes (2014) clearly stated the position that personalized professional development demands for teachers will be a principal responsibility:
Additionally, most states defer control over evaluation and its components to local education agencies. Consequently, whereas most state agency personnel could provide an overview of their evaluation systems’ professional learning component, few could do more than speculate about how professional learning, including the development and implementation of individualized teacher PD aligned with evaluation systems, is playing out locally. Most states either will not collect or have not begun collecting such data. The former stance speaks to the predominance of local control and, perhaps, states’ reliance on local control as a rationale for decoupled professional development efforts between departments of education and LEA’s. (p. 6)

Culver and Hayes (2014) further contended that only eight state education agencies confidently reported how districts had recently dealt with individualized provisions for teachers’ professional development. Hence, interest has increased in best practice approaches to evaluation-based professional learning for teachers by principals crushed with the daunting task (Stax, 2014). Kentucky’s DOE provided an example of a state agency committed to sharing the task of implementing individualized professional development aligned to teacher evaluation results. Principals had access to features within the Kentucky DOE’s online teacher evaluation platform that helped funnel aligned support resources to teachers through high-quality, on-demand professional development videos of education experts demonstrating best practices in the classroom (Culver & Hayes, 2014).

Shaha and Ellsworth (2013) conducted a multi-state, quasi-experimental study of the effects of internet-based, on-demand professional learning on student performance. A sample of 169 schools were randomly selected, and within those schools, teachers with a high participation in professional development were selected to participate in the study. Shaha and Ellsworth’s
(2013) results explicitly showed that students of participating teachers experienced an 18.9% increase (p<.001) in math scores, versus 4.2% (p<.01) for their districts, and a 15.3% increase (p<.001) in reading versus 2.5% (p<.01) for their districts, equating to statistically significant 4.2 fold (p>.001) and 6.1 fold advantages (p>.001), respectively. Within the high demands scope of the principals’ work in delivering personalized professional development to every teacher, Shaha and Ellsworth’s (2013) results helped principals give credibility to on-demand PD for improving teacher effectiveness and student performance.

**Teacher Evaluation and Principals’ Roles**

Education reformists focused on teacher evaluation policy call for principals implementing the new generation teacher evaluations to keenly hone skills related to instructional leadership purposed for increasing student achievement. The new generation of teacher evaluation systems has forced principals to develop these new skills centered on data, curriculum, pedagogy, human capital development, and maintenance of the traditional managerial skills of discipline, community relations, and conflict resolution (Alvoid & Black, 2014). According to a 2011 survey of American school leaders, almost 70% of principals reported that their job dramatically changed from just 5 years prior, and 75% of those school leaders reported an increase in job complexity, which, in turn, caused stress and lowered job satisfaction (Markow, Lee, & Interactive, 2013). High achieving districts on the cutting edge of innovation have provided pathways for redefining the roles of the principals in the face of new generation teacher evaluations. Alvoid and Black (2014) extracted a series of case studies to illustrate the innovative paths districts use to train and support principals to meet the challenges of providing instructional coaching, feedback, and personalized professional development.
The policy push for teacher evaluation caused a series of concerns about principals’ abilities and willingness to implement the new generation of teacher evaluation systems with fidelity. Only recently has research on the relationship between the new generation teacher evaluation systems and principals’ changing roles begun to emerge. Principals undoubtedly carry the implementation load of new generation teacher evaluation systems; yet, DeMonte and Pennington (2014) found principal supports in the implementation process seemed to be associated with mere curious afterthoughts. According to the National Association of Secondary School Principals [NASSP] and the National Association of Elementary School Principals [NAESP] (2014):

The needs of principals are overlooked in current ESEA provisions as states and districts currently spend less than 4% of Title II ‘allowable use’ of funds on principal professional development. As federal regulation requires principals to direct a variety of new reform initiatives in schools, this drastic misalignment of policy and practice cannot persist (p.4).

Previous studies revealed trends in concerns related to principals’ abilities to implement the new generation teacher evaluation systems with fidelity and included the following concepts: a) the lack of evidence that the time principals are reallocating to teacher evaluation systems leads to improved instruction and learning, b) issues of inadequate training specifically geared toward building the capacity for principals to differentiate performance and engage in intensive instructional coaching, c) problems posed by unclear and narrowly defined rubrics for observations coupled with insufficient feedback forums or ways to support meaningful teacher-principal relationships, and (d) time constraints caused barriers to making substantive, meaningful teacher evaluations and instructional coaching feasible (NASSP & NAESP, 2014). Additionally, DeMonte and Pennington (2014) contended principals’ professional-development
activities fell into similar weak training forms that received wide criticism in research studies on teacher professional development.

Arguments for redesigning the roles of principals and district leadership to promote principal leadership in new teacher evaluation implementation derived from Alvoid and Black (2014). In their newly released report from the Center of American Progress, “The Changing Role of the Principal: How High-Achieving Districts Are Recalibrating School Leadership,” Alvoid and Black (2014) recommended professional development for principals and improving district-wide leadership frameworks. The report also included the following recommendations: a) district leadership should redesign school organizational charts and job descriptions to ensure principals are receiving adequate support, b) central office administration must develop instructional leadership capacity around the principal, c) district leadership must focus principal training on coaching teachers, d) the district must provide regular opportunities for principals to gather around self-selected problems in their own practices, e) district leadership must develop partnerships with universities and nonprofit organizations to recruit and train future principals, f) the development and training of principals on district-wide teaching and leadership frameworks must be prioritized, and g) the district must provide the necessary technological supports to allow administrators to record and share instructional data (Alvoid & Black, 2014).

Summary

The literature review of teacher evaluation encompassed thought-provoking discussion and controversy. Yet, agreements about major trends within the topic of teacher evaluation exist between leading researchers. The findings of teacher evaluation studies before 2000 generally agreed with the most recent reviews of the literature with regard to the complexity involved in measuring teacher effectiveness and the lack of agreement about what constitutes good or
effective teaching (Papay, 2012; Weisberg et al., 2009). In most recent studies, teacher effectiveness remains tenable only through nuances in the terminology found in standardized observation protocols and model frameworks and no longer plausible in the one-dimensional view of effective teaching found in studies prior to the New Teacher Project’s (2009). The most recent research placed less emphasis on the lack of a definitive common language for effective teaching and instead focused on using the existing knowledge base on effective teaching to build a common vision of quality teaching as a first step in comprehensive teacher evaluation reform.

Accumulated research indicates that the current research-based observation protocols focusing on identifying practices that are most likely increase student learning (Hull, 2013), potentially providing a higher-quality measure of a teacher’s effectiveness and pathways to teacher performance improvement. However, the research repeatedly emphasized that improvement pathways only open when evaluators and teachers engage in shared conversations that build a clear image of quality teaching (Geo, Bell, & Little, 2008; Gitomer et al., 2014).

Evaluating teachers on student achievement, even partially, has been the subject of much debate in the literature, yet a succinct literature review revealed an increasing body of research supporting the direct relationship between teacher effectiveness and student learning (Kane et al., 2014; Papay, 2012). A review of the literature illustrated that most states adhere to the substantial body of research indicating that an effective teacher evaluation system employs multiple measures, including student achievement data (Bill & Melinda Gates Foundation, 2013; Goe et al., 2008). Most states evaluate teachers using multiple measure combinations of observation protocols and value-added models with no state currently evaluating teachers on student achievement test scores alone (Hull, 2013). Some states incorporate additional measure combinations that include student artifacts, student/parent surveys, lesson plan reviews, teacher
self-assessments, teacher portfolios, and measures of professional learning. A significantly high proportion of related studies in the field revealed concerns about the limited knowledge of these additional tools’ accuracy in measuring teacher effectiveness.

Stakeholder inclusion in teacher evaluation surfaced as a key to successfully designing and implementing accurate and reliable teacher evaluation systems. The research unanimously found that an inclusive design and implementation process are essential elements in building a quality teacher evaluation system. Most states legislate how to measure teacher effectiveness; however, central to the theme of successful implementation is the importance of listening to teachers, principals, content specialists, union representatives, board members, parents, and students (Behrstock-Sherratt et al., 2013). The research acknowledged the complex interdependent relationship between states and districts. Most districts do not possess the necessary resources to engineer a new evaluation system from scratch.

The literature is replete with references to the need to reinvent teacher evaluation as a part of a teaching and a learning system that supports continuous improvement for teachers, principals, students, and schools (Behrstock-Sherratt et al., 2013; Kane et al., 2014; Papay, 2012). Most of the literature suggests that more positive results occur when teachers feel included and genuinely engaged in the reform process (Behrstock-Sherratt et al., 2013). Paulson (2012) shared a leading policy analyst’s description of how failure to acknowledge the importance of teachers’ engagement in the process hindered RttT evaluation reform across the country. According to Paulson (2012), Elena Silva, a former senior policy analyst at Education Sector for Advancement in Teaching, stated, “The movement around teacher evaluation reform was pushed fairly boldly and quickly around the accountability issue and that was a mistake. It left teachers behind and didn’t attend to the vulnerabilities and insecurities teachers feel” (para.
Marshall (2014) and Behrstock-Sherratt et al. (2013) substantiated that teachers prefer principals who frequently observe them and offered detailed feedback as opposed to principals who enter the classroom once a year and offer generic comments. Behrstock-Sherratt (2013) repeatedly asserted that teachers also want a voice in the structure of teacher evaluation systems.

A succinct review of the literature emphasized the complexity in building an evaluation system that meets the needs of stakeholders. Behrstock-Sherratt et al., (2013) clearly stated the path to understanding the complexities of teacher evaluation reform: “It is only through the inclusion of many voices that the complexities inherent in teacher evaluation can be unearthed and addressed” (p. 5). Although teachers’ voices are currently almost non-existent in teacher evaluation reform, arguments for the inclusion of teacher’s ideas and opinions are rapidly gaining support. Darling-Hammond’s (2014) essential criteria for effective teacher evaluation systems is as follows:

Panels of teachers and administrators should oversee the evaluation process to ensure it is thorough and high quality, as well as fair and reliable. Such panels have been shown to facilitate more timely and well-grounded personnel decisions that avoid grievances and litigation. Teachers and school leaders should be involved in developing, implementing, and monitoring the system to ensure that it reflects good teaching well, that it operates effectively, that it is tied to useful learning opportunities for teachers, and that it produces valid results. (p. 12)

The available evidence suggests that as we continue to learn about replicating, creating, and evaluating the nuances within teaching and improved student achievement, the opinions of teachers, that is, those closest to the work, must be encouraged and used (Behrstock-Sharrett et al., 2013). Additionally, devoting more attention to teacher involvement and increased teacher
investment in policy creates opportunities for meaningful feedback, professional learning, and rating accuracy (Behrstock-Sherratt et al., 2013).

Another basic assumption pertinent to the study of teacher evaluation is that teacher evaluation results should be linked to professional development opportunities. The assumption that evaluation-based professional development should be prioritized varies among states and districts but remains consistent in the research (Archibald, Coggshall, Croft, & Goe, 2011; Bill & Melinda Gates Foundation, 2010; Council of Chief State School Officers, 2011; Kane & Staiger, 2012). Only 31 states align evaluation results to professional development provisions for all teachers (Culver & Hayes, 2014). The research extended beyond the benefits of aligning teacher evaluation results to professional development. Some studies highlighted the advantages of aligning teacher evaluation to comprehensive human capital management strategies to explore potential benefits. Although this interest is rapidly gaining support, there has been little investigation of the alignment between teacher evaluation results and larger human capital management systems in the field (Behrstock-Sherratt & Jacques, 2012).

Arguments for aligning teacher evaluation results, professional development, and recently extended human resource alignment are often repeated in the literature (Konstantopoulos, 2012). However, adherence to formalized procedures for aligning teacher evaluation results and professional development with extensions into human capital management define a weakness in the literature. Consequently, online platforms such as Teachscape, also a partner in the MET project (2012), offered online and mobile tools, professional learning resources, talent management solutions, and expert services claiming measurable growth results in teacher practice, school leadership, and student achievement. For example, Teachscape claimed, “With Teachscape’s observation and evaluation management, professional learning, and
talent management systems, administrators can strategically manage and develop their educators, along with non-teaching personnel, resulting in more high skilled staff, increased retention, and improved student outcomes” (“The Teachscape Story,” 2014). The research revealed that similar new start-up online platforms partnered with commercial research firms, for example, Stax Inc., and results have supported these products.

These online platforms also contain the capacity to assist with augmenting observation data by utilizing self-reflection tools, artifacts, and learning modules that allow access among principals/evaluators, teachers, and content specialist. The dynamic nature of the up-and-coming online teacher development platforms could potentially fill the gaps identified in the literature review, specifically, the real-life application of specific research-based observation protocols and the flexibility to upload continuously developing state or district specific observation rubrics. Recognition of this phenomenon offers the possibility that online teacher development platforms potentially provide the observation-evaluation tools for principals/evaluators and teachers to construct the necessary common language and vision of effective teaching in an equally personalized and precise path that leads to improved instruction.
CHAPTER III. METHODOLOGY

This survey addressed the research questions pertaining to Ohio principals’ intentions and practices toward OTES professional growth plans and used the theory of planned behavior as a framework (Ajzen, 1985, 1991). In 2013-2014, the entire state moved to a teacher evaluation system that mandated that districts allocate financial resources to support professional development (H.B. 153, 2011; Ohio Department of Education OTES, 2012). Every teacher evaluation system across Ohio must include provisions for professional development to accelerate and continue teacher growth, underscoring the urgency to explore principals’ intentions and practices toward OTES professional growth plans found in this research proposal (H.B. 362, 2014; Ohio Department of Education OTES, 2014.)

This chapter provides a description of the research design, participants in the research, instrumentation and data sources, data collection procedures, the research questions, and the methods used to analyze the data collected in this study. Finally, assumptions regarding the methodology of this study were presented.

Research Design

This non-experimental quantitative research study was designed to predict principals’ intentions and practices toward OTES professional growth plans and employed a cross-sectional self-report methodology grounded in Ajzen’s (1985, 1991) theory of planned behavior. A non-experimental design was a logical choice for this study because the research questions contained independent variables not subject to manipulation (Johnson & Christensen, 2010). Self-report data from principals’ attitudes, subjective norms, perceived behavior control, intentions, and practices toward OTES professional growth plans were used to study the relationships among the
variables. The self-report method often provides the only available method for researchers interested in investigating individuals’ belief systems (Paulhus & Vazire, 2007).

According to Ajzen (2014), most methods developed for data collection and analysis with the theory of planned behavior use quantitative approaches. Ajzen’s (1985) theory of planned behavior offers an explanatory model for a broad variation of behavioral intention (Lee, Cerreto, & Lee, 2010). Multiple meta-analysis provided support for the efficacy of the theory of planned behavior as a predictor of intentions and behaviors (Armitage & Conner, 2001).

Recent applications in the theory of planned behavior in education studies supported the use of a cross-sectional self-report methodology. Stone, Jawahar, and Kisamore (2009) used the theory to investigate academic misconduct; Weng, Weng, and Tsai (2014) employed a web-based and paper self-reported questionnaire to study predictors of college students’ participation in online teacher evaluations; Youngyool Oh (2003) investigated Korean teachers’ willingness to change traditional teaching practices with the theory of planned behavior; and Landeck (2006) employed the theory of planned behavior to explore principals’ intentions to promote diversity awareness.

An online questionnaire collected the self-report data on principals’ attitudes, subjective norms, perceived behavior control, intentions, and practices toward OTES professional growth plans. The framework presented in this study for evaluating construct validity was amassed from a similar research design by Yan and Cheng (2015). The design structure Yan and Cheng used to evaluate construct validity is best described by Bond and Fox (2015) as the “Rasch-Messick” framework. This framework combines Messick’s (1989, 1995) broad construct validity theory with Rasch’s (Rasch, 1960) measurement practice to employ a series of quality control checks, including reliability estimates and rating scale functionality.
This study, like the model framework study design provided by Yan and Cheng (2015), used the theory of planned behavior guidelines presented by Ajzen (2006) to construct a questionnaire for the current study. As suggested by Ajzen (2014), the direct theory of planned behavior measures (attitudes, subjective norm perceptions, and perceived behavior control) was developed in the formative research stage to ensure the general questionnaire’s suitability for the behavior and population of interest.

In regards to construct validity, the researcher applied Rasch measurement techniques both *priori* and *post hoc* to Yan and Cheng’s (2015) adapted instrument to capitalize on the Rasch model’s predictive and inferential (Bond & Fox, 2015) usefulness in securing satisfactory psychometric properties of the scales. This study’s approach toward construct validity derived from Bond and Fox’s (2015) description of a repeated loop of theory and practice conducted using the Rasch model: “The process of construct validation works at the interface between the development of a data collection instrument and the empirical data so collected” (p. 314). In other words, Rasch measurement played an on-going supervisory role in allowing the researcher to ascertain the extent to which the data actually measure the constructs under examination (Bond & Fox, 2015).

Yan (2014) advocated the use of Rasch analysis over traditional analysis formulated on classical test theory because of the Rasch analysis’s benefits in working with ordinal data collected through Likert-type response scales. Principals’ measures on seven latent traits, which are interval measures, were subjected to traditional analysis of variance (ANOVA) and hierarchical multiple regression to describe their attitudes, subjective norms, perceived behavior control, intentions, and practices toward OTES professional growth plans and the relationship among these variables in the context of Ohio schools was examined (Yan, 2014). Lastly, a path
analysis was employed to test the fit between Ajzen’s (1985) TPB model and the observed set of correlations between the variables in the model. Path analysis repeatedly and appropriately identifies common applications to test the fit between Ajzen’s model and the correlations between variables within the model (Igram, Cope, Harju, & Wuensch, 2001).

Seventeen high school, middle school, and elementary school principals from a convenience sample participated in the initial piloted questionnaire. The OTES credentialing requirements for high, middle, and elementary school principals holds no difference among principal assignment. These unilateral credentialing requirements for all principals allow the research questions to focus on the heterogeneous group of principals and simultaneously honor the theory of planned behavior’s homogeneous intentions toward utilization (Lee, Cerreto, & Lee, 2010).

Participants

The population for this survey research included school principals in Ohio. The K-12 brick and mortar educational system in Ohio includes public, nonpublic, community, and joint vocational schools. The target population participants selected for this survey research included K-12 brick and mortar public school principals in Ohio as the population for this study. The population of K-12 brick and mortar public school principals in Ohio served as the target population because they are the largest group directly responsible for ensuring the implementation of OTES professional growth plans. The total K-12 yield of public school principal population in Ohio equals approximately 3,175, based on the Ohio Educational Directory System (OEDS, 2014). The population of nonpublic school principals found exclusion due to nonpublic schools’ exclusion from OTES.
Ohio has eight typology categories assigned to districts, labeled one to eight. Included in Table 3.1 is a description of the eight categories of Ohio school district typologies (ODE Typology, 2013).

Table 3.1

2013 Ohio School District Typology

<table>
<thead>
<tr>
<th>Typology Code</th>
<th>Major Grouping</th>
<th>Descriptor</th>
<th>Districts within Typology</th>
<th>Students within Typology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rural</td>
<td>High student poverty, small student population size</td>
<td>124</td>
<td>170,000</td>
</tr>
<tr>
<td>2</td>
<td>Rural</td>
<td>Average student poverty, very small student population size</td>
<td>107</td>
<td>110,000</td>
</tr>
<tr>
<td>3</td>
<td>Small Town</td>
<td>Low student poverty, small student population size</td>
<td>111</td>
<td>185,000</td>
</tr>
<tr>
<td>4</td>
<td>Small Town</td>
<td>High student poverty, average student population size</td>
<td>89</td>
<td>200,000</td>
</tr>
<tr>
<td>5</td>
<td>Suburban</td>
<td>Low student poverty, average student population size</td>
<td>77</td>
<td>320,000</td>
</tr>
<tr>
<td>6</td>
<td>Suburban</td>
<td>Very low student poverty, large student population size</td>
<td>46</td>
<td>240,000</td>
</tr>
<tr>
<td>7</td>
<td>Urban</td>
<td>High student poverty, average student population size</td>
<td>47</td>
<td>210,000</td>
</tr>
</tbody>
</table>
The Ohio Department of Education used several data sources, including 2010 census-data, to classify similar districts based on shared demographic and geographic traits to create Ohio’s 2013 typology classifications (ODE, 2013). Ohio removed five districts from the typology classifications and designated these districts as null due to special circumstances and insufficient data. The following districts comprised the aforementioned exclusions: College Corner Local School District, Kelley’s Island Local School District, Middle Bass Local School District, North Bass Local School District, and Put-in-Bay Local School District. Ohio’s one through eight typology classifications served as the basis for the stratified sampling frame of districts in the state. According to Lee, Benoit-Bryan, and Johnson (2012), sampling frame quality determines the extent of coverage error, and generalizability issues emerge when sampling frames fail to fully represent a target population. The researcher made every effort to ensure full representation of the target population.

Raykov and Marcoulides (2006) summarized the complexity of determining sample size: “Although there is universal agreement of researchers that the larger the sample relative to the population the more stable the parameter estimates, there is no agreement as to what constitutes large, due to the exceeding complexity of this matter” (p. 30). Researchers can employ at least three ways to determine a study’s sample size: 1) collect as many participants as possible, 2) use rules of thumb, or 3) employ formalized sample size planning methods (Lai & Kelley, 2011). This study employed all three of Lai and Kelly’s (2011) sample size options.

The researcher first intends to collect as many participants in this survey research as possible. Secondly, the rules of thumb as presented by Tabacknick and Fidell (2007) suggested
the following formula for calculating sample size in hierarchical multiple regression: N > 50 + 8m (N = number of Participants m = number of IVs). Based on this calculation, 98 respondents were required for the study. The sampling size web-based program G*Power 3.1 produced an opportunity to validate the desired priori sample size (Mayr, Erdfelder, Buchner, & Faul, 2007). According to Francis et al. (2004), the theory of planned behavior studies assumes a moderate effect size. An effect size of 0.3, probability error of 0.05, and priori power of .8 were entered into the G*Power 3.1 calculator and yielded an estimated total sample size of 152 for a goodness-of-fit test. The priori power of .8 yielded a low estimate, and a priori power of .95 yielded a total sample size of 232.

A recent study by Jacobs and Jacobs (2012) investigated the response rates of school principals and reported a low principal response rate, 15.4%, for web surveys with no paper advance invitation, no incentive, and no reminder follow-ups. The researchers recorded a response rate of 18.4% for web surveys with advanced paper invitations, no incentives, and no reminder follow-ups. Assuming response rates as conservative as 15.5%, a minimum of 1,500 advanced invitations will need to be mailed out and 1,500 web questionnaires will need to be deployed through e-mail for an anticipated yield rate of 232 returned surveys.

As required by The University of Findlay’s Institutional Review Board (IRB), the researcher submitted the final questionnaire to the IRB committee. Consideration of the ethical criterion applied to voluntary participation in the study’s design formed the foundation for informed consent. Informed consent ensures participants understand the nature of the research and can competently and voluntarily decide whether or not to participate. This assurance protects the participants’ autonomy and the investigator.

The researcher provided prospective participants with all required reasonable information
during the informed consent process. Assurances for the informed consent process included an invitational letter via the email solicitation in the pilot study and the email solicitation and follow-up reminders in the final study. The researcher clearly articulated to all possible participants that they can quit the questionnaire at any time. The invitation postcard, pilot questionnaire, and final questionnaire focused on OTES professional growth plan concepts that normally occur in the daily professional life of principals. Therefore, the study design and related questionnaire documents present little risk of affecting participants’ psychological status.

This survey research design separated the data from the participants using an online survey system service called QuestionPro. The QuestionPro web-based survey service helped ensure anonymity after the initial submission of potential respondents list. All participants indicated consent and agreement to complete the survey by accessing the link. The web-based survey service secured anonymity without the need for hard copy paper files, which increased the possibility of compromising respondents’ anonymity. This study received approval from the IRB committee and the IRB committee categorized the study as minimal risk to the subjects with no adverse effects on the rights and welfare of the subjects.

Instrumentation and Data Sources

DeVillis (2012) suggested that researchers seek existing instruments when he advised, “It is important and efficient to be quite certain that a suitable measurement alternative does not already exist” (p. 185). As a result, the major constructs in the theory of planned behavior investigated in this study ultimately were assessed by the contextual adaptation of seven scales previously developed by Yan and Cheng (2015) in the A Teacher’s Conceptions and Practices of Formative Assessment Questionnaire (TCPFAQ). Most importantly, Yan and Cheng’s (2015) seven scales within the questionnaire reflect the conceptualization of Ajzen’s (1991) theory of
planned behavior applied to the phenomenon of a top-down educational reform initiative. The questionnaire attempted to construct an understanding of educators’ attitudes, intentions, and practices regarding the mandated initiative in a very similar context.

This study’s self-report questionnaire was contextually adapted from the seven scales contained in Yan and Cheng’s (2015) A Teacher’s Conceptions and Practices of Formative Assessment Questionnaire and addressed the purpose of this study. The adaptive instrument development relied on standard attitude scaling procedures to assess affective and instrumental attitudes, subjective norm perceptions, perceived behavior control, intentions, and practices in a direct manner, as suggested by Ajzen (2014). The adaptive instrument development was equally influenced by the following DeVillis (2012) scale development guidelines: a) clearly determine what you want to measure, b) generate an item pool, c) determine the format for measurement, d) have initial item pool reviewed by experts, e) consider inclusion of validation items, f) administer items to a development sample, g) evaluate the items, h) optimize scale length.

The adaptation of the seven scales started with enrolling nine OTES experienced educators, who were not included in the study group, selected from a convenience sample. Specifically, the convenience sample included six OTES-experienced principals and three OTES-expert trainers who train principals and teachers on the evaluation system for the state of Ohio. Each participant was asked to list the following information about their use of OTES professional growth plans: a) the advantages and disadvantages, b) persons and groups who would approve and disapprove, and c) the elements that make it easy or difficult. Next, from the information generated, the individuals from the convenience sample assisted the researcher in identifying and wording the key communication concepts that related to evaluator components of OTES professional growth plans. The experts’ responses, offered in one-on-one conversations
with the researcher, formed the base of prevailing determinants of attitude toward OTES professional growth plans, subjective norm, and perceived behavior control (Ajzen’s, 2014). Finally, seven separate scales designed to assess principals’ attitudes and practices toward OTES professional growth plans emerged from the formative data generated.

The questionnaire contained two separate scales targeted at measuring two attitudinal subcomponents within the theory of planned behavior. The attitude component within the theory of planned behavior is widely believed to contain two specific types of attitudes, referred to as affective and instrumental (Ajzen, 2014; Conner & Armitage, 1998; Rhodes & Courneya, 2003; Rosenberg, 1956). The affective attitude component encompasses descriptions such as enjoyable or not enjoyable in relationship to the behavior under investigation (Rhodes & Courneya, 2003). The instrumental attitude component refers to the utility perception of the behavior under investigation and is often represented on semantic differential scales as harmful or beneficial (Rhodes & Courneya, 2003). The two-component attitude structure finds acceptance across multiple attitude measurement methodologies in research and extends beyond the theory of planned behavior (Crites, Fabrigar, & Petty, 1994; Olson & Zanna, 1993; Rhodes & Courneya, 2003). The four items on the affective attitudinal scale were used to assess the theory’s direct measure of principals’ affective attitudes toward OTES professional growth plans. These items explored how principals feel about completing OTES professional growth plans (Fishbein & Ajzen, 2010). The three items on the instrumental attitude scale investigated whether principals view the behavior as productive (Fishbein & Ajzen, 2010).

According to Ajzen (2006), the subjective norm construct represents the perceived social pressure to participate or not to participate in a behavior. In Armitage and Conner’s (2001) efficacy meta-analysis of the theory of planned behavior, the authors postulated that the
subjective norm component of the theory provided the weakest predictor of intentions. Trafimow and Finlay (1996) provided evidence that the subjective norm component possesses predictive power and independently establishes actions driven by subjective norm components across 30 behaviors. According to Armitage and Conner (2001), the likely explanation for low performance of the subjective norm component resided in faulty performance of single item measures. Ultimately, the subjective norm showed a reasonably strong relationship with intention when accurately measured with multiple-item scales (Armitage & Conner, 2001). The direct measure of the subjective norm component included four items designed to examine principals’ perceptions about the importance others’ opinions toward OTES professional growth plans. The direct measurement of subjective norm used the four items to specifically estimate the principals’ perceptions of the opinions toward OTES professional growth plans held by state officials, superintendents, teachers, and other principals.

The questionnaire also contained two separate scales targeted at measuring the perceived behavior control subcomponents recognized within the theory of planned behavior. According to Ajzen (2006), perceived behavior control refers to the degree of perceived ease or difficulty of performing a behavior, and perceived behavior control assumedly reflects past experience as well as any anticipated barriers. Ajzen (2006) maintained that, conceptually, no differences exist between perceived behavior control and self-efficacy because both refer to people’s belief that they are capable of performing a given behavior, whereas Bandura (1997) expressed the view that self-efficacy measures offer more clearly defined and operationalized aspects compared to perceived behavior control attempts that consist of mixed approaches.

In this respect, Armitage and Conner (2001) clarified that self-efficacy appears to be the preferred measure of perceived control in their meta-analysis of efficacy of Ajzen’s (2006)
theory of planned behavior; however, Armitage and Conner (2001) cited the need for further research on the impact of different operationalization approaches of perceived control on intention and behavior. Trafimow, Sheeran, Conner, and Finlay’s (2002) 11-study empirical review of self-efficacy and controllability confirmed that these two concepts showed reliably distinguishable across a wide range of behaviors, and the authors further contended that self-efficacy serves as a superior predictor over controllability for both intentions and behaviors.

As a result of the literature review, measures for the perceived behavior component of Ajzen’s (1991) theory of planned behavior relied on two distinct scales of self-efficacy and controllability (Rhodes & Courneya, 2003) in the study’s questionnaire. The three items on the controllability scale estimated the extent to which principals perceived that participation in OTES professional growth planning falls under their control (Ajzen, 2006). Bandura’s (1997) self-efficacy measurements included five items that were used to gauge the graded series of potential obstacles principals face in implementing OTES professional growth plans for teachers.

The intentions of principals toward OTES professional growth plans were assessed through six items contained in the intention scale adapted from Yan and Chang (2015). The principals practice behaviors of OTES professional growth plans were estimated by four items on the behavior scale, also adapted by Yan and Change (2015).

All seven scales modeled by Yan and Cheng (2015) were administered to a pilot convenience sample of 17 assistant principals and principals. Additionally, all participants in the pilot held OTES credentialing certificates and demonstrated at least one year of OTES implementation experience. The pilot group’s size restricted the usefulness of any extracted statistical data with the exception of an estimated time investment of six minutes. The pilot questionnaire’s usefulness surfaced in the identification of issues within item wording through
open-ended feedback sections.

The pilot questionnaire requested that principals rate each item on a 6-point scale. The 6-point scale mimicked Yan and Cheng’s (2015) scale with response options ranging from strongly disagree (1) to strongly agree (6). During the data analysis stage, reverse coding of negatively worded items ensured consistency in item interpretation. In other words, a higher score consistently represented a high level on the explored latent trait (Yan, 2014).

The results presented in Table 3.2 indicated that all scales showed high measures of internal consistency in the pilot work. The following rules of thumb regarding Cronbach’s alpha reliability coefficient proposed by George and Mallery (2003) supported the pilots’ internal consistency attempt: “ > .9 – Excellent, > .8 – Good, > .7 – Acceptable, > .6 – Questionable, > .5 – Poor, and < .5 – Unacceptable” (p. 231). Nunnally (1978) concurred with George and Mallery that instruments should have a Cronbach’s alpha (α) of .70 or higher for a scale to have reliable measures of internal consistency. Researchers cautioned against using Cronbach’s α in isolation (Cortina, 1993). The seven scales were analyzed separately in SPSS 22.0.

Table 3.2

<table>
<thead>
<tr>
<th>Scale</th>
<th>Number of Items</th>
<th>Cronbach’s α Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective Attitude</td>
<td>4</td>
<td>0.750</td>
</tr>
<tr>
<td>Instrumental Attitude</td>
<td>3</td>
<td>0.768</td>
</tr>
<tr>
<td>Subjective Norm Scale</td>
<td>4</td>
<td>0.802</td>
</tr>
<tr>
<td>Controllability Scale</td>
<td>3</td>
<td>0.685</td>
</tr>
<tr>
<td>Self-Efficacy Scale</td>
<td>5</td>
<td>0.782</td>
</tr>
<tr>
<td>Intention Scale</td>
<td>6</td>
<td>0.729</td>
</tr>
<tr>
<td>Behavior Scale</td>
<td>4</td>
<td>0.744</td>
</tr>
</tbody>
</table>
Ajzen (2014) urged researchers to establish the convergent and discriminant validities of the theory’s measures in the pilot work. Both Ajzen (2006) and DeVellis (2012) underscored the importance of ensuring items share a common cause and consequence toward the same latent variable, and Ajzen (2014) emphasized the importance of attitude, subjective norm, and perceived behavior control in assessing separate constructs and in not measuring the same underlying construct. To assess the convergent and discriminant validity of the seven scales, an item analysis to measure the correlation of each item to its scale was conducted. In support of the instrument’s validity, the results indicated that the items show a high correlation with their own scale. Each of the individual scales retained the appropriate levels of internal consistency. The notion that each scale can be deemed internally consistent remained tenable because of the limited number of pilot participants. The small pilot numbers over simplified the scale’s operable functions and further reiterated the researcher’s application of Rasch measurement techniques to the instrument after data were fully collected.

At the end of the pilot questionnaire, an open-ended item gave the participants the opportunity to provide input regarding the questionnaire. The research participants in the pilot saw the following sentence on the final screen: “Thank you for your valuable time. If you have any comments or information to add to this questionnaire, please use the space provided.” The open-ended data received from the pilot study proved helpful in refining the questions’ wording, sequence, and format and the electronic launching nuances and general word clarity within the questionnaire. The final questionnaire did not retain the open-ended item.

Table 3.3
Final Survey Instrument
Affective Attitude Scale
Q1 I like OTES professional growth plans.
Q2 OTES professional growth planning is an enjoyable process.
Q3 OTES professional growth planning encourages teachers and evaluators to help each other improve.
Q4 OTES professional growth planning is worth my effort.

Instrumental Attitude Scale
Q5 OTES professional growth planning can encourage teacher to engage in self-directed learning.
Q6 OTES professional growth planning can raise teachers’ interest in teaching.
Q7 OTES professional growth planning can offer a fair appraisal of a teacher’s professional development needs.

Subjective Norm
As far as I know, the following stakeholders believe that OTES professional growth planning should be implemented.
Q8 Officials in the state education department
Q9 The superintendent of my district
Q10 The teachers in my school
Q11 My administrative colleagues

Controllability Scale
Q12 I can decide the frequency of implementing OTES professional growth plans.
Q13 I can decide the time table for implementing OTES professional growth plans.
Q14 I can decide whether or not to implement OTES professional growth plans.

Self-Efficacy Scale
Q15 I can integrate OTES professional growth plans into the OTES evaluation cycle process.
Q16 I have received sufficient training to implement OTES professional growth plans.
Q17 I can assist teachers in designing appropriate goals for OTES professional growth plans.
Q18 I have enough time to implement OTES professional growth plans.
Q19 I have sufficient supporting materials (e.g., online modules, handbooks, DVD) to implement OTES professional growth plans.
Q20 I have sufficient skills to implement OTES professional growth plans.

Intention Scale
Q21 I am willing to try and implement OTES professional growth plans.
Q22 I am willing to integrate OTES professional growth plans into the evaluation cycle.
Q23 I am willing to assist teachers in designing appropriate goals for OTES professional growth plans.
Q24 I am willing to put effort toward implementing OTES professional growth plans.
Q25  I am willing to adjust traditional approaches to professional development to meet the requirements of OTES professional growth plans.
Q26  I am willing to encourage teachers to participate in professional growth plans.

Behavior Scale
Q27  In the past six months, how often have you engaged in developing OTES professional growth plans?
   1=Everyday 2=Almost every day 3 =Most days
   4 =A number of days, but less than half of the time
   5=Some days 6=Never

Q28  Please estimate how frequently you have had conversations with individual teachers specifically regarding their personal OTES professional growth plans in the past six months.
   1=Very frequently  2= Frequently  3=Sometimes  4=Seldom  5=Rarely  6=Never

Q29  Please estimate the amount of financial resources you have specifically allocated toward individual teachers’ OTES professional growth plans (not improvement plans) in the past six months.
   1= more than $20,000   2=between $20,000-$15,000  3=$15,000-$10,000
   4=$10,000-$5,000  5=less than $5,000
   6=no financial resources allocated specifically toward individual teachers’ OTES professional growth plans

Q30  How often have you provided time for teachers to specifically develop the skills necessary to achieve the goals stated in their individual OTES professional growth plans in the past six months?
   1= Everyday  2=Almost every day 3=Most days
   4=A number of days, but less than half of the time
   5=Some days 6=Never

Demographically Descriptive Items
Q31  Please click the link to identify your district’s typology code.
Q32  Please select the category that best describes your OTES caseload.
   1-7 teachers  8-15 teachers  16-23 teachers  24-31 teachers >32 teachers

The final two questions measured two of the participants’ demographic characteristics (typology classification and OTES caseload) and are considered by Ajzen (2006) as background factors. According to Ajzen (2006), demographic variables indirectly influence intentions and behavior by affecting behavioral, normative, and/or control beliefs. Furthermore, the theory of
planned behavior components assumes mediation of the effects of background factors on intentions and actions (Ajzen, 2006).

**Data Collection Procedures**

The data collection procedures in this study sought to control for the common error found within survey research, including coverage error, measurement error, nonresponse error, and processing error. During data collection procedural planning, the researcher attended to the factors involved in appropriate mode selection to help reduce measurement errors, as proposed by Lee, Benoit-Bryan, and Johnson (2012). Tourangeau and Yan (2007) reported that sensitive topics that intrude on participants, for example, criminal acts, sexual behavior, and voting behavior, can cause potential issues that threaten disclosure and often lead to social desirability bias measurement errors, which self-administered surveys through mail and web servers mitigate. Lee, Benoit-Bryan, and Johnson (2012) state that social desirability bias proved less of a concern in public administration research arenas exploring organizational attitudes, and implored researchers in public administration arenas to focus on reducing measurement error stemming from loss of quality data that fuels nonresponse bias. Therefore, this self-administered web-based survey mode placed focus less on threat reduction from social desirability bias measurement error (Fowler, 2002); instead, the researcher increased attention on reducing the threat to non-sampling error with attention focused toward nonresponse bias reduction (Lee, Benoit-Bryan & Johnson, 2012).

Increasing the response rate is the best strategy to minimize nonresponse error, specifically nonresponse bias, in survey estimates (Fowler, 2002). The contact delivery mode (Paraschiv, 2013) section of the data collection process opened with the distribution of a paper version pre-notification postcard invitation sent by mail to all potential participants. In the past,
researchers presented evidence that sending a pre-notification letter or postcard raises response rates, increases the trust between the participants and the researcher, and decreases the chance of discarded surveys (Fox, Crask, & Kim, 1988). The pre-notification announcement was distributed on Friday, September 18, 2015 and took place to prime the launch of the web-based survey on October 2, 2015. The survey window closed on October 30, 2015.

The web-based survey platform QuestionPro assisted the researcher in the technical design and distribution of the survey. QuestionPro’s capabilities included branching, skip logic, piping, and advanced question types that support specific scales most investigators use within theory of planned behavior frameworks. Additionally, participants who received the email invitation with the embedded link only needed to ignore the access link to opt out of the survey.

QuestionPro also allowed users to embed reference data within the survey to assist the participants in decision-making. The researcher embedded the reference data of Ohio’s 2013 Typology at the end of the survey as the demographic collection component. The demographic data collection portion of the survey entailed one question asking participants to identify the category of district typology membership. The embedded reference data helped participants accurately identify with the appropriate typology membership classification, as most participants likely hold minimal awareness of district typology classifications. Additionally, the researcher asked participants to quantify their individual OTES caseload, that is, the number of teachers the principal was responsible for evaluating.

QuestionPro provided the researcher with the ability to customize an introductory e-mail cover letter that coincided with the paper version pre-notification postcard invitation mailed three weeks prior to the launch of the web-based survey. Both the pre-notification postcard and the e-mail cover letter introduced the researcher, the topic of the study, the survey window’s open and
close dates, contact information for the primary investigator, Institutional Review Board contact information, ethical guidelines followed by the researcher, implied consent terms, and the assurance of confidentiality. The pre-notification postcards used Vista Print mailing services via the U.S. Postal Service. The e-mail subject heading contained the name of the survey. The researcher used a university email address to increase trust, as its suffix, @findlay.edu, is easily recognizable as originating from a higher education institution.

The electronic invitations to participate in the survey arrived by e-mail to potential participants on Friday, October 2, 2015 between noon and four o’clock in the afternoon. Paraschiv (2013) studied the timing of web-based survey delivery in relationship to response rates and provided evidence that this time interval registered a very active response activity. The first follow-up reminder to access the survey was scheduled to activate on Friday, October 16, 2015 during the same time interval between noon and four o’clock in the afternoon. Paraschiv (2013) reported Fridays showed the strongest response rate in regards to weekday relationship to response activity.

QuestionPro streamlined the process of sending follow-up reminders by automatically e-mailing reminders to non-participants without requiring the researcher to access or cross-reference response lists. Millar and Dillman (2009) encouraged researchers to employ automatic selective e-mail reminders to assist with increasing response rates while simultaneously ensuring that participants who have already completed the questionnaire are not offended. The researcher scheduled one additional follow up reminder spaced two weeks apart after the initial reminder. The second follow up reminder was distributed by email on Friday, October 23, 2015.

The option to access the survey appeared at the end of the invitational page through a live survey link. The landing page included a definition of the OTES professional growth plan,
directions to the survey, an optional link for participants to view Ohio’s OTES professional
growth plan template, and the 32-item questionnaire. Once the survey was completed, the
participants were presented with a very short demographic section containing seven items with a
direct link to the typology classification matrix.

QuestionPro provided a professional survey layout with an attractive graphic design on
the invitational e-mail, landing page, and subsequent survey pages. According to Walston,
Lissitz, and Rudner (2006), graphic design, professional style presentation, and intuitively
inviting platforms positively impact response rates. Conversely, respondent motivation,
measurement, and survey completion proved negatively impacted when survey platform options
rigidly force participants to answer every option (Millar & Dillman, 2009). As a result, the
proposed survey platform did not require participants to complete every item prior to moving
through the survey.

Research Questions

This study sought to discern whether or not the Ohio Teacher Evaluation System model
presented by the Ohio Department of Education is effectively meeting its intended goal to have
principals help teachers find the professional development they need, at the individual level, to
make improvements through the use of professional development plans. More specifically, this
study sought to determine principals’ intentions and behaviors toward OTES professional growth
plans using the framework of the theory of planned behavior (Ajzen, 1985) and strived to answer
the following questions:

1. Can principals’ intentions toward OTES professional growth plans be predicted by
   principals’ attitude, subjective norms, and perceived behavior control toward OTES
   professional growth plans?
2. Can principals’ OTES professional growth plan practices be predicted by intentions and perceived behavior control toward OTES professional growth plans?

3. Do principals’ intentions and practices differ among Ohio’s typology classifications and/or principals’ OTES caseload?

By virtue of the theory of planned behavior (Ajzen, 1991), this research pursued to measure a) principals’ attitudes toward OTES professional growth plans; b) principals’ perceptions regarding subjective norms (the influence of persons important to the principal when performing the behavior), if they executed OTES professional growth plans; and c) self-efficacy ratings (rated ability to perform the behavior) the principal reported for OTES professional growth plan tasks (Ajzen, 1991).

Data Analysis

This study aimed to employ Rasch analysis (Rasch, 1960) and path analysis. The Rasch Measurement Software Winsteps® 3.9’s data analysis approach allowed the psychometric properties of seven adapted scales adapted from Yan and Cheng (2015) to be thoroughly examined and the calculation of principals’ measures on each of the underlying traits (Yan, 2014). Yan (2014) advocated for the use of Rasch analysis, as opposed to traditional analysis, based on classical test theory because of its ability to compensate for ordinal data collected through Likert-type response scales. Wright (1997) took the position that ordinal data indicate ordering exclusively with no ability to indicate any proportional levels of meaning, making them inappropriate for traditional analyses that require linear, interval scale data input (Yan, 2014). Yan and Cheng (2015) accurately summarized the evidence and discussed the problem of researchers arriving at misleading conclusions after applying conventional analytical techniques to raw scores that are ordinal in nature. Rasch analysis converts ordinal raw data for interval
logit scale measures to provide more objective results and linear measurements that are otherwise not obtained (Bond & Fox, 2007; Linacre, 2012; Yan, 2014). The literature is replete with empirical references to the successful application of Rasch analysis in educational settings and social science settings to tackle tough assessment issues (Bond & Fox, 2007; Panayides, Robinson & Tymms, 2010; Yan & Cheng, 2014).

The Winsteps® 3.9 Rasch analysis software enabled the researcher to establish an interval metric prior to the calibration of person measures and item difficulty onto a single unidimensional latent trait scale that later facilitated direct comparisons between person measures and item difficulties (Linacre, 2006; Yan & Cheng, 2015). Yan and Cheng (2015) provided a clear explanation that,

In Rasch analysis, person measures are independent from the item used, with item difficulties being independent from the sample recruited because the estimates are calibrated within a common metric rather than against a single test situation (for person ability estimates) or a particular sample of test takers (for item difficulty estimates). (p. 130)

The Rasch rating scale analysis was used to examine the psychometric properties of the seven scales and calibrated principal (person) measures on each of the seven latent traits, including attitude (two subcomponents), subjective norm, perceived behavior control (two subcomponents), intention, and behavior (Yan & Cheng, 2015).

The seven scales’ psychometric properties were explored through Rasch person/item reliability, item fit statistics, the amount of variance explained by each of the five scale measures, and step threshold; and, the Rasch sequence of the scale properties was directly attributed to Yan and Cheng (2015). The Rasch person/item reliability offered the capacity to estimate
person/item ordering along the latent trait metrical structure (Bond & Fox, 2007; Yan & Cheng, 2014). The item fit statistics approximated the degree to which the data match the Rasch Model measurement specifications (Yan & Cheng, 2015) using outfit and infit mean squares (MNSQs) as indices. Linacre (2012) states that variance explained by Rasch measures stands for the proportion of variance in the observed data explainable by the Rasch measures. Wright and Linacre (1994) recommended that MNSQs show in the range of 0.6-1.4 to indicate a productive measurement for survey data with rating scales. Lastly, the researcher employed step thresholds to examine the functional appropriateness of each category in the rating scale (Yan & Cheng, 2015).

Next, the Rasch calibrated principal (person) measures were examined by path analysis using analysis of moment structures (AMOS) through SPSS 22.0 using Winsteps® 3.9’s conversion procedures and maximum likelihood for model estimation. The AMOS 22-driven path analysis was used to investigate the relationships among the theory of planned behavior components (Arbuckle, 2013; Yan, 2014). Yan and Cheng (2015) explained that this approach to data analysis is useful in measuring unobserved traits, whereas the traditional structural equation model usually uses confirmatory factor analysis to satisfy the utility of measurement model. Waugh and Chapman (2005) opposed the application of factor-analysis to non-linear item scores and substantiated Yan and Cheng’s (2014) justification for the Rasch model. A parallel analysis by Salzberger (2012) described Rasch analysis and factor analysis as mutually exclusive models. The Rasch Model, by assessing fit, examines whether observed person raw scores can be transformed into linear person measures and observed item raw scores into linear item measures, with factor analysis measures as the input (Waugh & Chapman, 2005).
Salzberger (2012) further clarified that in the Rasch model, non-fitting data imply that item scores show as non-linear or not ordinal, and that instead, data represent arbitrarily assigned numerals to response options. Salzberger (2012) contended that proponents of Steven’s (1946) scale measurement definition mistake establishing measurement and what denotes a factor analyst’s is only proof of measurement at the item level, or factor analysis necessitates the measure it claims to provide. Yan and Cheng’s (2014) exclusion of factor analysis measurement, and substitution of Rasch-calibrated measures, in the structural equation modeling finds further justification from Salzberber (2012), who asserted:

Today, there does not seem to be any need for conducting a factor analysis on raw data prior to a Rasch analysis. In fact, researchers might feel the need to run a confirmatory factory analysis (CFA) after the Rasch assessment of a scale in order to use measures in a structural equation model (SEM). However, Rasch measures can be integrated in SEM quite easily. (p. 5)

Ajzen (2006) confirmed path analysis application by designating multiple regression or structural equation analysis as the means for establishing estimates of relative importance among the three predictors (attitude, subjective norm, and perceived behavior control) within his theory. The principals’ Rasch-calibrated measures on the theory of planned behavior components were presented through path analysis to allow the researcher to explore the two main research questions. The first question explored if principals’ intentions toward OTES professional growth plans can be predicted by attitude, subjective norms, and perceived behavior control toward OTES professional growth plans. The second question investigated if principals' practices toward OTES professional growth plans can be predicted by their intentions and perceived behavior control toward OTES professional growth plans.
A basic assumption in the literature, pertinent to the application of path analysis, is that multiple indices should be sued to measure model fit in studies that employ path analysis as an analytical technique (Kline, 2005). The researcher used a model chi-square ($\chi^2$) and relative chi-square ($\chi^2$/df), comparative fit index (CFI), root mean square error of approximation (RMSEA), and Tucker-Lewis Fit (TLI). According to Garson (2008), the value of CFI and TLI may vary from 0 to 1, with values exceeding 0.90 indicative of a good fit. The level of acceptable fit for RMSEA, as rendered by Hu and Bentler (1999), is a value less than 0.8. Kline (2005) offered a value of less than 3 to determine acceptable fit using chi-square.

The covariate in this study was based on principal typology classification membership. More importantly, principal typology classifications was used to investigate possible distinguishable characteristics related to principals’ intentions and practices toward OTES professional growth plans.

As previously discussed, ordinal data was collected by the Likert-type scales and was transformed into interval measures through Rasch analysis (Yan, 2014). Principals’ Rasch-calibrated measures on the seven scales was also subjected to traditional inferential statistics. Inferential statistics was undertaken to investigate the two demographically descriptive characteristics among the principals responding. A one-way analysis of variance (ANOVA) was used to explore the two demographically descriptive characteristics and identify whether any significant differences in principals’ intentions and/or practices exist within principals’ typology categorical membership and OTES case load. Principals’ Rasch-calibrated measures on the seven scales were subjected to traditional descriptive statistics.

The means and standard deviations of the principals’ measures on the seven scales along with Pearson correlations of the theory of planned behavior constructs provided the lens through
which the interval Rasch-calibrated measures were viewed. Yan (2014) states, “In Rasch analysis, the mean of item difficulties is arbitrarily set to zero and the interpretation of item difficulties and person measures are based on pair-wise comparisons between items and persons. Therefore, person measure higher than zero indicates a positive response, while person measures lower than zero indicate a negative response” (p. 132).

The goal of this study was to examine the predictive utility of the theory of planned behavior to predict principals’ intentions and practices toward OTES professional growth plans. Table 3.4 provides an outline of the research questions and corresponding statistical methods adopted in this study.

Table 3.4

Research Questions and Corresponding Statistical Tests

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Statistical Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can principals’ intentions toward OTES professional growth plans be predicted by principals’ attitudes, subjective norms, and perceived behavior control regarding OTES professional growth plans?</td>
<td>Principals’ Intentions</td>
<td>Direct Measures of Principals’ Attitude, Subjective Norm, and Perceived Behavior Control</td>
<td>Rasch Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Path Analysis</td>
</tr>
<tr>
<td>Can principals’ OTES professional growth plan practices be predicted by intentions and perceived behavior control toward OTES professional growth plans?</td>
<td>Principals’ Practices</td>
<td>Direct Measures of Principals’ Intentions and Perceived Behavior Control</td>
<td>Rasch Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Path Analysis</td>
</tr>
</tbody>
</table>
Do principals’ intentions and practices differ among the typology classifications?

<table>
<thead>
<tr>
<th>Principals’ Intentions and Practices</th>
<th>Typology</th>
<th>ANOVA</th>
</tr>
</thead>
</table>

**Assumptions**

The basic assumption throughout this quantitative study was that principals’ intentions and practices toward OTES professional growth plans are difficult, but possible, to achieve. Or, at the very least, it is possible to find meaningful indicators regarding principals’ intentions and practices toward OTES professional growth plans through this quantitative study. The researcher expected principals to respond independently from each other to the online questionnaire.

The researcher selected the Rasch model as a method of analysis for this study, as this study sought to examine the psychometric properties of the adapted scale and to calculate principals’ measures on each of the latent traits. Unidimensionality is often highlighted as the most essential assumption within the Rasch model; unidimensionality indicates that one attribute or latent trait will be measured at a time (Bond & Fox, 2015). In this study, the latent variables of principal intentions, practices, attitudes, subjective norms, and perceived behavior control were independently measured. Existing scales, produced by Yan and Cheng (2015), underwent limited pilot adaptation for context; as a result, the principal components analysis of residuals was employed to ensure that the latent variable items adequately fall along a unidimensional scale.

As modeled by Yan and Cheng (2015), this study first sought to fully investigate the psychometric properties of the seven scales from the Rasch measurement perspective. Any items
that did not fit the Rasch model (both infit and outfit MNSQ being higher than 1.4) were singly deleted from the scale in the misfit order; then, re-application of Rasch analysis will continue until all remaining items showed sufficient fit to the Rasch model (Yan & Cheng, 2015). The items gained this careful removal from the scale to lessen the item deletion impact on the accuracy of the estimates model outputs (Hart & Wright, 2002; Yan, 2014; Yan & Cheng, 2015). Consequently, Shaw (1991) attested, “One impediment to acceptance of the Rasch model as a viable method of measurement in psychology and education has been the concept of fit” (p. 132). A firm commitment to the Rasch model provided the researcher with the opportunity to more deeply understand the instrument used to collect the data.

The main assumption within the theory of planned behavior posits that individuals are rational in considering their actions and the implications of those actions (Fishbein & Ajzen, 1975). There is considerable support (Armitage & Conner, 2001) for the capacity within the theory of planned behavior’s framework to predict a variety of behaviors; however, the theory of planned behavior model is based on the deductive reasoning of others. The theory of planned behavior assumes that prediction of behavior from intentions improves as the measurement of intention and the opportunity to display the behavior are presented closer in time (Gochman, 1997). As a result, this study coordinated the launch of the online final survey with the approximate window of OTES professional growth plan peek operations.

According to Gochman (1997), the comparative importance of intention and perceived behavioral control in the prediction of behavior is assumed to vary across situations and behaviors. Gochman explained that perceived behavior control is comparable to perceived self-efficacy (Bandura, 1986) because it involves beliefs that one has the assets and the opportunities to perform a behavior or accomplish a goal. Conversely, several researchers have presented
evidence that reinforces a separation between self-efficacy and perceived behavior control (Dzewaltowki, Noble & Shaw, 1990; Rodgers, Conner & Murray, 2008; Terry & O’Leary, 1995). This study’s instrument was adapted from Yan and Cheng’s (2014) *A Teacher’s Conception and Practices of Formative Assessment Questionnaire*, and the instrument contained a self-efficacy scale and a scale to measure for perceived behavior control. This study assumed the position that perceived behavior control and self-efficacy maintain conceptual differences, and, unless specifically stated, all perceived behavior control references indicate two distinct subcomponents—self-efficacy and controllability.

Several assumptions fall within the nature of the self-reported questionnaire used in the study. The researcher assumed that participating principals responded honestly, and that principals did not attempt to manage how they would appear through their survey responses. The researcher assumed that participating principals possess the required introspective ability to offer their best effort in responding to each question with honesty and accuracy. The researcher assumed that all participating principals will understand the operational definition of OTES professional growth plans presented at the onset of the questionnaire in an effort to establish a common predisposition for principals’ interpretations (Yan & Cheng, 2015).
CHAPTER IV. RESULTS

This dissertation is a research study about Ohio public school principals and the factors influencing their intentions toward and practices of OTES professional growth plans. The study is based on the perceptions of Ohio public school principals who represent Ohio’s eight district typology classifications. After the focus group research, a survey was created and reviewed and a pilot tested for validity and reliability with maximum brevity. Following survey enactment, application of Rasch analysis further examined the operation of the instrument’s scales designed to measure principal constructs of the theory of planned behavior and specific demographic variables. Furthermore, Rasch-calibrated person measures, in place of traditional test theory total score measures, provided the input data for path analysis.

The first two research questions were addressed through a path analysis in AMOS (version 23) using the structure based on Ajzen’s (1991) theory of planned behavior model. The path analysis applied endogenous variables of behavior and intention regressed on the exogenous variables of attitude, subjective norms, and perceived behavioral control. Ajzen’s (1985) classic theory of planned behavior model finds application in past studies involving implementation of educational initiatives in the literature review (Yan, 2014). The standardized path coefficients generated in the AMOS path analysis are also presented in this chapter.

The employment of several common model-fit measures were used to assess the model’s overall goodness-of-fit within the AMOS program. Yan (2014) and Yan and Cheng (2015) guided the tests selection that included: chi-square, root mean square error of approximation (RMSEA), normalized fit index (NFI), Tucker Lewis Index (TFI), incremental fit index (IFI), relative fit index (RFI), goodness of fit (GFI), adjusted goodness of fit index (AGFI), and the comparative fit index (CFI). A one-way analysis of variance (ANOVA) was conducted to
address the final research question posed to exploit any potential differences among principals’ intentions and behaviors toward OTES professional growth plans within Ohio’s typology classifications.

**Characteristics of the Sample**

All Ohio public school principals, serving in brick-and-mortar sites received an invitation to participate in this study. A total of 2,212 questionnaires were electronically distributed across all of Ohio’s school district typologies, and 235 were completed and returned.

The online survey platform, QuestionPro, provided a detailed view of the respondents’ activity regarding the questionnaire by allowing the researcher to learn the following: 647 principals viewed the survey, 340 principals started the survey, 105 principals dropped out of the survey after starting the survey, and the average time to take the survey equaled 3 minutes. QuestionPro also provided information regarding the type of devices that participants used to respond to the survey. Among the responding principals, 4.53% used a tablet, 16.92% of respondents used a smartphone, and 78.55% respondents used a desktop or laptop computer.

Among the 235 respondents designated as completed by QuestionPro, 48 respondents were extracted due to one or more missing data points found within the item requiring more than one response, and 17 respondents were extracted as outliers during the Rasch analysis process. The current study focused on the remaining 170 Ohio principals with 100% completed questionnaires to avoid partial credit models, missing data complications, and non-outlier data points to investigate principals’ intentions and practices regarding OTES professional growth plans based on their firsthand conceptions.

The final survey question requested principals to identify the Ohio typology classification code equating to the district of service through a series of dropdown boxes. The 170 principals’
dataset for this current study comprised representation from all eight of Ohio’s 2013 typology classifications as shown in Table 4.1 below.

Table 4.1

2013 Ohio Typology Classification Principals’ Representation

<table>
<thead>
<tr>
<th>Typology Code</th>
<th>Major Grouping</th>
<th>Full Descriptor</th>
<th>Raw Number of Principals</th>
<th>Percent of Principals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rural</td>
<td>High Student Poverty and Small Student Population</td>
<td>21</td>
<td>12.4%</td>
</tr>
<tr>
<td>2</td>
<td>Rural</td>
<td>Average Student Poverty and Very Small Student Population</td>
<td>17</td>
<td>10.0%</td>
</tr>
<tr>
<td>3</td>
<td>Small Town</td>
<td>Low Student Poverty and Small Student Population</td>
<td>19</td>
<td>11.2%</td>
</tr>
<tr>
<td>4</td>
<td>Small Town</td>
<td>High Student Poverty and Average Student Population</td>
<td>26</td>
<td>15.3%</td>
</tr>
<tr>
<td>5</td>
<td>Suburban</td>
<td>Low Student Poverty and Average Student Population Size</td>
<td>28</td>
<td>16.5%</td>
</tr>
<tr>
<td>6</td>
<td>Suburban</td>
<td>Very Low Student Poverty And Large Student Population</td>
<td>25</td>
<td>14.7%</td>
</tr>
<tr>
<td>7</td>
<td>Urban</td>
<td>High Student Poverty and Average Student Population</td>
<td>16</td>
<td>9.4%</td>
</tr>
<tr>
<td>8</td>
<td>Urban</td>
<td>Very High Student Poverty and Very Large Student Population</td>
<td>18</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

**Instrument Validity and Reliability**

The psychometric properties of the seven scales used with this sample were initially examined from a Rasch measurement perspective as modeled by Yan (2014) and Yan and Cheng
Rasch analysis is widely used in educational assessment and social sciences for tackling tough assessment issues (Bond & Fox, 2014; Knutson, 2011; Yan (2014), Yan & Cheng, 2015). Many state departments of education, including the state of Ohio, used Rasch analysis to analyze high-stakes tests created by No Child Left Behind (Boone, Staver & Yale, 2014). The reason Rasch analysis gains momentum in application resides in the Rasch model’s ability to overcome the problem of arriving at misleading conclusions from the application of conventional analytical techniques to ordinal data or raw scores (Linacre, 2006). Rasch analysis overcomes problems encountered with assumptions that ordinal data behave like linear data with the ability to convert ordinal data into equal interval linear data appropriate for parametric statistical analysis (Yan, 2014; Wright, 1997; Bond & Fox, 2014; Linacre, 2006).

Rasch person and Rasch item fit statistics that included infit and outfit mean squares (MNSQ) tested the level to which the data match the measurement criteria of the Rasch model as in Yan (2014) and Yan and Cheng’s (2015) studies. These Rasch person/item reliability measures coupled with the variance explained by the measures gauged the quality of scales (Yan, 2014; Yan & Cheng, 2015). The goal for the instrument used in this study sought to obtain a set of scales for which Cronbach’s alpha for person/item reliability measures registered at least .85, with no item exceeding the more demanding range of 0.6 to 1.4 as the MNSQ fit statistics criteria for the Rasch model.

According to Wright and Linacre (1994), the calculations of MNSQ produces an average near 1.0, and values larger than 1.0 indicate too much unexplained variance in the data, whereas values smaller than 1.0 show that the model hyper-predicts the data, leading to inflated reliability statistics. Yan (2014) highlighted that previous studies designated different acceptable ranges for establishing good model data fit based on Wright, Linacre, Gustafson, and Martin-Löf’s
(1994) claim that no stone-carved rules regarding acceptable ranges for good model data fit indicators exist. Linacre (2012) regarded the range of 0.5 to 1.5 as an indicator of productive measurement, whereas Wright and Linacre (1994) offered the range of 0.6 to 1.4 as reasonable person and item MNSQ ranges for infit and outfit in survey data with rating scales. Items with infit and outfit values of greater than 1.4 indicate the item is unhealthy in some way, or is vaguely defined (Smith, 2004). The Rasch infit and outfit measures for each of the items are presented in Table 4.2 below.

Table 4.2. Rasch-Analysis Item Fit Indices

<table>
<thead>
<tr>
<th>Question Item:</th>
<th>Rasch-Analysis Fit Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INFIT</td>
</tr>
<tr>
<td>1. I like OTES professional growth plans.</td>
<td>0.88</td>
</tr>
<tr>
<td>2. OTES professional growth planning is an enjoyable process.</td>
<td>0.82</td>
</tr>
<tr>
<td>3. OTES professional growth planning encourages teachers and evaluators to help each other improve.</td>
<td>0.92</td>
</tr>
<tr>
<td>4. OTES professional growth planning is worth of my effort.</td>
<td>0.76</td>
</tr>
<tr>
<td>5. OTES professional growth planning can encourage self-directed learning of teachers.</td>
<td>1.08</td>
</tr>
<tr>
<td>6. OTES professional growth planning can raise teachers’ interest in teaching.</td>
<td>1.02</td>
</tr>
<tr>
<td>7. OTES professional growth planning can offer a fair appraisal of individual teacher’s professional development needs.</td>
<td>1.59</td>
</tr>
<tr>
<td>8. As far as I know, the officials of the state education department believe that OTES professional growth planning should be implemented.</td>
<td>1.49</td>
</tr>
<tr>
<td>9. As far as I know, the superintendent of my district believes that OTES professional growth planning should be implemented.</td>
<td>0.96</td>
</tr>
<tr>
<td>10. As far as I know, the teachers in my school believe that</td>
<td></td>
</tr>
</tbody>
</table>
OTES professional growth planning should be implemented. 0.72 0.77

11. As far as I know, my administrative colleagues believe that OTES professional growth planning should be implemented. 0.66 0.60

12. I can decide the frequency of implementing OTES professional growth plans. 1.14 1.0

13. I can decide the timing of implementing OTES professional growth plans. 0.92 0.83

14. I can decide whether or not to implement OTES professional growth plans. 0.88 0.81

15. I can integrate OTES professional growth plans into the OTES evaluation cycle process. 1.63 1.58

16. I have received sufficient training to implement OTES professional growth plans. 0.82 0.75

17. I can assist teachers in designing appropriate goals for OTES professional growth plans. 0.84 0.82

18. I have enough time to implement OTES professional growth plans. 1.37 1.38

19. I have sufficient supporting materials (e.g., online modules, handbooks, DVDs) to implement OTES professional growth plans. 0.91 1.01

20. I have sufficient skills to implement OTES professional growth plans. 1.04 1.05

21. I am willing to try to implement OTES professional growth plans. 1.07 0.82

22. I am willing to integrate OTES professional growth plans into the evaluation cycle. 0.85 0.80

23. I am willing to assist teachers in designing appropriate goals for OTES professional growth plans. 0.94 0.70

24. I am willing to put effort toward implementing OTES professional growth plans. 0.71 0.63

25. I am willing to adjust traditional approaches to professional
development to meet the requirements of OTES professional growth plans.  

<table>
<thead>
<tr>
<th>Question</th>
<th>MNSQ 1.50</th>
<th>MNSQ 1.34</th>
</tr>
</thead>
<tbody>
<tr>
<td>26. I am willing to encourage teachers to participate in professional growth plans.</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>27. In the past 6 months, how often have you engaged in developing OTES professional growth plans?</td>
<td>0.94</td>
<td>0.77</td>
</tr>
<tr>
<td>28. In the past 6 months, please estimate how frequently you have had conversations with individual teachers specifically regarding their personal OTES professional growth plans.</td>
<td>0.73</td>
<td>0.83</td>
</tr>
<tr>
<td>29. In the past 6 months, please estimate the amount of financial resources you have allocated specifically toward individual teachers’ OTES professional growth plans (not improvement plans).</td>
<td>1.34</td>
<td>1.29</td>
</tr>
<tr>
<td>30. In the past 6 months, how often have you provided time for teachers to specifically develop the skills necessary to achieve the goals stated in their individual OTES professional growth plans?</td>
<td>1.16</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Linacre (2012) ascertained that item outfit MNSQ and person outfit MNSQ exhibits a useful sensitivity toward identifying outliers. As a result, during the initial analysis, priority to item outfit MNSQ and person outfit MNSQ offered the most efficient path to identifying person outliers and correcting issues of item fit (Boone, Staver, & Yale, 2014; Linacre, 2012). Of the total sample (n = 187), approximately 9.1% (17 participants) were identified as misfits or outliers for removal.

The Rasch analysis identified item 7 (OTES professional growth planning can offer a fair appraisal of a teacher’s professional development needs) from the instrumental attitude scale and item 15 (I can integrate OTES professional growth plans into the OTES evaluation cycle process) from the self-efficacy scale as misfitting to the Rasch model by MNSQ falling outside of the acceptable range (Yan, 2014). Two items in the original behavior scale showed fit to the
Rasch model with MNSQ values inside the acceptable range; however, the items were excluded because including the two items produced a Cronbach’s alpha of zero for the behavior scale. The two excluded items are identified as item 29 (“Please estimate the amount of financial resources you have specifically allocated toward individual teachers OTES professional growth plans—not improvement plans—in the past 12 months,”) and item 30 (“How often have you provided time specifically for teachers to develop skills necessary to achieve the goals stated in their individual OTES professional growth plans in the past year?”)

Hart and Wright (2002) cautioned that Rasch analysis must be reapplied to all remaining items after removing items individually and sequentially to ensure the accuracy of remaining item estimations and the accuracy of the error rate estimations. As modeled by Yan (2014) and Yan and Cheng (2015), the items identified as misfitting were removed from the scale one at a time as identified by the misfit order with Rasch analysis re-ran until all remaining scales presented sufficient fit to the Rasch model with reasonable Cronbach’s alpha. The sequenced extraction of item 29 increased Cronbach’s alpha from 0 to .49, whereas the extraction of both item 29 and item 30 increased Cronbach’s alpha to .71 for the behavior scale. The behavior scale did not meet the study’s goal of .85 Cronbach’s alpha; however, the behavior scale’s Cronbach’s alpha of .71 does fit Field’s (2005) desired level. Furthermore, Kline (2000) argued that values below .70 might surface with scales that use only two or three items to investigate psychological constructs. All other items in the remaining four scales had acceptable fit measures according to the Rasch model. Table 4.3 presents the summary of the psychometric properties of all scales as modeled in Yan and Cheng (2015).

It can be observed from Table 4.3 that the Rasch person/item reliabilities for all scales are higher than 0.80, with the exception of the person reliability for the behavior scale. The tables
reflect that most of the scales produced reasonably good psychometric properties, whereas the behavior scale only showed an adequate quality. Additionally, the satisfactory reliability coefficients produced within the scope of the short number of items per scale deserved mention. Rasch measures explained over 70% of the variables observed in the data for all scales. This moderately high proportion of variance in the observed data explained by the Rasch measures suggested that the Rasch model provided a useful prediction of item and person performance for the seven scales (Yan, 2014). In summary, according to Yan (2014) the modestly good psychometric properties of the seven scales support their use with the sample in this study.

Table 4.3 Psychometric Properties of Measurement Scales for Measuring Components of TPB

<table>
<thead>
<tr>
<th>Scale</th>
<th>Number of Items</th>
<th>Rasch Person Reliability</th>
<th>Rasch Item Reliability</th>
<th>Variance Explained by Measures</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>4</td>
<td>0.91</td>
<td>0.97</td>
<td>82.0%</td>
<td>.93</td>
</tr>
<tr>
<td>IAT</td>
<td>2</td>
<td>0.85</td>
<td>0.98</td>
<td>76.4%</td>
<td>.85</td>
</tr>
<tr>
<td>SNO</td>
<td>4</td>
<td>0.81</td>
<td>0.99</td>
<td>73.4%</td>
<td>.77</td>
</tr>
<tr>
<td>PBC</td>
<td>3</td>
<td>0.80</td>
<td>0.98</td>
<td>70.1%</td>
<td>.81</td>
</tr>
<tr>
<td>SEF</td>
<td>5</td>
<td>0.86</td>
<td>0.98</td>
<td>70.0%</td>
<td>.85</td>
</tr>
<tr>
<td>INT</td>
<td>6</td>
<td>0.88</td>
<td>0.87</td>
<td>70.1%</td>
<td>.92</td>
</tr>
<tr>
<td>BEH</td>
<td>2</td>
<td>0.74</td>
<td>1.00</td>
<td>83.2%</td>
<td>.71</td>
</tr>
</tbody>
</table>

Wright Maps show the hierarchy of survey items while simultaneously showing the hierarchies of both persons and items (Boon, Staver, & Yale, 2014). The items easiest to endorse drop toward the bottom of the hierarchy, and items most difficult to endorse are represented toward the top of the item hierarchy (Bond & Fox, 2015). The Wright Maps also identify the mean of the items along with the mean of the persons (Boon, Staver & Yale, 2014). The Wright Maps indicate person and item means by the letter “M”, the person means are marked with an “M” on the left side, and the item means are marked with an “M” on the right side. The maps also identify one standard deviation with the letter “S”, and the letter “T” represents two standard
deviations. The Wright Maps offer a visual analysis of the persons on the left side and the items on the right side. The distance between the items on the hierarchy of the map, or logit scale, visually demonstrates how the items function in relationship to each other (Boon, Staver, & Yale, 2014). Bond and Fox (2012) explained the logit scale is an interval scale and emphasized the distance between the logits are of equal size.

The person item map in Figure 4.1 demonstrates the closeness of the person mean to the item mean. According to Boon, Staver, and Yale (2014), measurement shows strength and precision when items are targeted to the mean of person. The item hierarchy is also consistent with previous studies regarding principals’ negative attitudes toward OTES professional growth plans (Kowalski & Dolph, 2015), with all attitude items represented toward the top of the more-difficult-to-endorse item hierarchy.

*Figure 4.1 Person Item Map*

*Figure 4.1.* The symbol “#” represents 2 and the symbol “.” represents 1.
Category-use statistics were examined to assess the category functioning within the seven scales to determine the rating scale’s utility. According to Bond and Fox (2015), category frequencies and average measures are the easiest way to assess category functioning. Category frequencies summarize the distribution of all responses across all categories and allow the researcher to quickly gain a basic understanding of the rating scale use (Bond & Fox, 2015). Linacre (1995) explained that average measures illustrate the average of the ability estimates of all persons included in the sample who selected that particular response category and the average calculated across all observations in that category. The average measures are expected to increase monotonically across the response categories (Bond & Fox, 2015). Bond and Fox (2015) conveyed that average measures that increase monotonically across response categories indicate that participants with stronger attitudes support progressively higher categories, whereas participants with weaker attitudes support progressively lower categories. All category frequencies (observed count) and average measures for the six-category rating scale found in each of the seven scales functioned with adequate category frequency and the expected monotonicity of average measures.

Wright and Masters (1982) explained that threshold calibrations offer an estimated view of how difficult the endorsement of one response category over the adjacent response category behaves for respondents. Threshold calibrations, also called step calibrations, when healthy, behave in the same monotonic trend as the average measures do (Bond & Fox, 2015), and thresholds that do not produce monotonic increases across the rating scales designate a disordering (Linacre, 2012). According to Linacre (2012), the disordering of these estimates may indicate that the category is relatively rarely observed.
The subjective norm (SNO) rating scale and the self-efficacy (SEF) rating scale each revealed one pairwise measure of the transition that did not ascend in value up the rating scale, between step 2 and step 3. All other step thresholds advanced monotonically with the rating scale category, indicating that the six category rating scale functioned adequately and, for the most part, higher performance categories corresponded to higher measures on the latent trait under investigation (Bond & Fox, 2007; Linacre, 2006; Yan & Cheng, 2015). The five-step thresholds for the six-category rating scale are presented in Table 4.4 below.

Table 4.4.
Step Thresholds of Measurement Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>-7.01</td>
<td>-2.82</td>
<td>-1.73</td>
<td>2.26</td>
<td>9.29</td>
</tr>
<tr>
<td>IAT</td>
<td>-5.18</td>
<td>-2.16</td>
<td>-1.16</td>
<td>1.79</td>
<td>7.15</td>
</tr>
<tr>
<td>SNO</td>
<td>-2.80</td>
<td>-0.41</td>
<td>-1.11</td>
<td>0.35</td>
<td>3.97</td>
</tr>
<tr>
<td>PBC</td>
<td>-5.02</td>
<td>-0.59</td>
<td>-0.58</td>
<td>0.41</td>
<td>5.79</td>
</tr>
<tr>
<td>SEF</td>
<td>-2.49</td>
<td>-0.58</td>
<td>-1.27</td>
<td>0.37</td>
<td>3.98</td>
</tr>
<tr>
<td>INT</td>
<td>-3.14</td>
<td>-2.31</td>
<td>-2.24</td>
<td>0.71</td>
<td>6.98</td>
</tr>
<tr>
<td>BEH</td>
<td>-8.71</td>
<td>-3.85</td>
<td>-0.45</td>
<td>2.51</td>
<td>10.50</td>
</tr>
</tbody>
</table>

As previously modeled by Yan (2014), descriptive statistics from this study provided a lens to view the interval Rasch-calibrated measures of principals on the seven constructs related to the theory of planned behavior. According to Yan and Cheng (2015), in Rasch analysis, the mean of the item difficulties is subjectively set to 0, and the understanding of item difficulties and person measures are based on pairwise comparisons between items and persons. Consequently, person measures higher than 0 specify a positive response, whereas person measures lower than 0 specify a negative response (Yan, 2014; Yan & Cheng, 2015).
Overall, principals presented a noticeably high level of intention to conduct professional 
growth plans (INT Mean = +5.33, SD = 5.48). Principals’ subjective norm perceptions about how important others’ (state department, superintendent, other principals, and teachers) regard professional growth plans were also considerably positive (SNO Mean = +1.46, SD = 2.79). Principals’ mean person measures on the self-efficacy scale (SEF Mean = +0.02, SD = 1.99) just scarcely registered. These results suggested that principals claimed low confidence levels toward implementing OTES professional growth plans while simultaneously holding high intentions to implement OTES professional growth plans. According to Yan (2014) the large standard deviations associated with principals’ measures on the Intention scale inferred that there exist considerable individual differences amongst these principals.

Principals’ measures on the Affective Attitude (AAT Mean = -0.78, SD= 4.28), Instrumental Attitude (IAT Mean = -0.69, SD = 3.27) scales regarding OTES professional growth plans were only marginally lower than zero. Principals’ measures regarding the perceived behavior control (PBC Mean = -2.42, SD = 2.88) scale registered as the most negative. This signaled that principals held slightly negative to marginally neutral attitudes toward OTES professional growth plans and held a low sense of control toward OTES professional growth plans. Additionally, it should be highlighted that principals had a negative mean measure on behavior toward OTES professional growth plans (BEH Mean = -1.04, SD = 0.78). This implied that according to principals’ perceptions, conversations with individual teachers about OTES professional growth plans and allocated energy toward planning OTES professional growth plans are infrequent. The descriptive statistical means and standard deviations of principals’ measures on the scales along with Pearson correlations among the theory of planned behavior-related constructs of interest are presented in Table 4.5 below.
As can be seen from Table 4.5, the typical constructs used in the theory of planned behavior are highly correlated, and all the correlations among the predictors registered above 0.8. Field (2009) cautioned that correlations between predictors above 0.8 warrant regard as a signal of multicollinearity. Field (2009) explained that multicollinearity is a state of very high intercorrelations among the independent variables. Multicollinearity causes a type of disturbance in the data and, depending on the amount found present in the data, the statistical inferences made about the data may not be reliable (Dunlap & Kemery, 1987).

Alin (2010) suggested combining related predictor variables as the first step in attempting to solve multicollinearity issues. Therefore, prior to testing the extent to which multicollinearity existed within the data, affective attitude scale and instrumental attitude scale were combined into one attitude scale, and self-efficacy scale was combined into the perceived behavior control scale. Table 4.6 presents the newly collapsed theory of planned behavior components’ means, standard deviations, and person correlations. Consequently, a new path model surfaced as a
result of the collapsed scales. The new path model represented Ajzen’s (1985) original five components of the theory of planned behavior model—as opposed to the seven-component model represented by Yan and Cheng (2015) originally presented for this study.

Table 4.6. Means, Standard Deviations, and Correlations of Theory of Planned Behavior ComponentsCollapsed Scales

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>AAT</th>
<th>SNO</th>
<th>PBC</th>
<th>INT</th>
<th>BEH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>-0.49</td>
<td>3.69</td>
<td>----</td>
<td>0.96</td>
<td>0.94</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>SNO</td>
<td>+1.46</td>
<td>2.79</td>
<td>0.96</td>
<td>----</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>PBC</td>
<td>-0.58</td>
<td>1.10</td>
<td>0.94</td>
<td>0.91</td>
<td>----</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>INT</td>
<td>+5.33</td>
<td>5.48</td>
<td>0.97</td>
<td>0.93</td>
<td>0.91</td>
<td>----</td>
<td>0.92</td>
</tr>
<tr>
<td>BEH</td>
<td>-1.04</td>
<td>0.78</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.92</td>
<td>----</td>
</tr>
</tbody>
</table>

The calculation of both a tolerance test and variance inflation factor (VIF) indicated the presence of a multicollinearity problem within the data set. The results of the tolerance test and VIF were interpreted using Cohen, Cohen, West, and Aiken’s (2003) criteria for tolerance levels. According to Cohen, Cohen, West, and Aiken (2003), tolerance levels identified as greater than or equal to .01 and VIF values identified as greater than or equal to 10 point toward a multicollinearity problem. Table 4.7 presents the outcomes of the multicollinearity tolerance test and variance inflation factor below.

Table 4.7. Multicollinearity Tolerance Test and Variance Inflation Factor Test

<table>
<thead>
<tr>
<th>Variables</th>
<th>Tolerance</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude</td>
<td>0.02</td>
<td>36.95</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>0.06</td>
<td>15.42</td>
</tr>
<tr>
<td>Perceived Behavior Control</td>
<td>0.11</td>
<td>8.99</td>
</tr>
<tr>
<td>Intention</td>
<td>0.05</td>
<td>17.23</td>
</tr>
</tbody>
</table>
Next, the extent of the multicollinearity problem was explored. According to Dereny and Rashwan (2011) ridge regression analysis proves useful in computing the condition number that indicates the degree of multicollinearity. The condition number is the biggest eigenvalue divided by each corresponding eigenvalue, and because the eigenvalues are variances, the condition number is a ratio of variance (NCSS, 2015). Condition numbers larger than 1,000 indicate a severe multicollinearity problem, whereas condition numbers between 100 and 1,000 indicate a mild multicollinearity problem (NCSS, 2015). It can be seen from Table 4.8 that the multicollinearity problem identified in this Rasch-calibrated person measures data set is best classified as a mild problem.

Table 4.8. Ridge Regression Analysis Report of Eigenvalues of Correlations

<table>
<thead>
<tr>
<th>Number of Eigenvalue</th>
<th>Eigenvalue</th>
<th>Incremental Percent</th>
<th>Cumulative Percent</th>
<th>Condition Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.827345</td>
<td>95.68</td>
<td>95.68</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.092181</td>
<td>2.30</td>
<td>97.99</td>
<td>41.52</td>
</tr>
<tr>
<td>3</td>
<td>0.060900</td>
<td>1.52</td>
<td>99.51</td>
<td>62.85</td>
</tr>
<tr>
<td>4</td>
<td>0.019574</td>
<td>0.49</td>
<td>100.00</td>
<td>195.23</td>
</tr>
</tbody>
</table>

Principals’ Rasch-calibrated person measures on the theory of planned behavior components were subsequently subjected to path analysis using the analysis of moment structures (AMOS) 23.0 version. According to Hancock and Mueller (2004), path analysis adheres to a basic process of estimating relationships among measured variables represented as a path diagram then thoroughly analyzing the observed data relations for degree of fit with the original hypotheses’. Path analysis differs from traditional regression analysis methods use of one statistical test, RSquare, to determine the significance of the analysis (Hancock & Mueller, 2004). Instead, path analysis relies on several statistical tests to determine the acceptability of model fit to the data.
The path analysis application to the complete set of data was conducted in two phases. The first path analysis application encompassed the original conceptualized seven-scale model presented in the methods chapter of this dissertation. The first path analysis endogenous variables included the deconstruction of principals’ attitudes toward OTES professional growth plans into affective attitudes and instrumental attitudes, principals’ subjective norms for using OTES professional growth plans, and the deconstruction of principals’ perceived behavior control construct to include a separate component of self-efficacy.

The second path analysis included the traditional theory of planned behavior constructs endogenous variables of attitude, subjective norm, and perceived behavior control. Both path analyses included the theory of planned behavior constructs of intention and behavior as the exogenous variables and used the same sample of 170 principals’ Rasch-calibrated person measures.

**Research Question 1**

The results of the first path analysis demonstrated an overall unsatisfactory fit between the first conceptualized theory of planned behavior model and the observed data. According to Arbuckle (2014), the chi-square test indicates the amount of difference between the over-identified model and saturated model, or the difference between the conceptualized model and the perfect model. The chi-square statistic ($\chi^2 = 2323.135$, df = 13, $p = .000$) and the relative chi-square ($\chi^2 / \text{df} = 178.70$) achieved the minimum but were unsatisfactory.

The GFI, goodness of fit index, indicates proportional variance in the sample variance-covariance matrix that is accounted for by the model (Arbuckle, 2014) and should exceed .9 for a good model. This first path analysis model produces an unsatisfactory GFI of .278. The AGFI, adjusted goodness of fit index, represents an alternate GFI index in which the value of the index
finds adjustment based on the number of parameters in the model (Wuensch, 2016), and negative values indicate poor fit (Byrne, 2001). This proposed model’s AGFI = -.555. These two goodness-of-fit indices compared the conceptualized model with the independence model and not the saturated model (Byrne, 2001).

The baseline comparison fit indices help identify the difference between the two models’ chi-squares divided by the independence models’ chi-square (Byrne, 2001; Wuensch, 2016). The baseline comparison measures also failed to meet satisfactory fit criteria of values higher than .9 (Arbuckle, 2014) and included the normed fit index (NFI = .360), relative fit index (RFI = -.033), incremental fit index (IFI = .362), Tucker-Lewis index, (TLI = -.033), and the comparative fit index (CFI = .360).

The root mean square error of approximately (RMSEA) estimates lack of fit compared with the saturated model. Models with RMSEA of .05 or less indicate a good fit, and models with RMSEA of .08 or lower equal a reasonable error of approximation (Byrne, 2001). The first conceptualized model’s RMSEA registered much too high at .960 and indicated an inadequate fit to the saturated model (Arbuckle, 2014).

These results of the first path analysis do not support the appropriateness of the proposed theory of planned behavior based model in predicting Ohio’s principals’ intentions and practices regarding OTES professional growth plans. The failed model is presented in figure 4.2 below. The model’s unacceptable fit measures make the parameters, estimates, and relationships among the latent traits in the theory of planned behavior-based model inappropriate for examination.
Figure 4.2. Failed theory of planned behavior-based model 1. IA represents the instrumental attitude, AA represents the affective attitude scale, SNO represents the subjective norms scale, PBC represents the perceived behavior control, SEF represents the self-efficacy scale, INT represents the intention scale, and BEH represents the behavior scale.
Research Question 1 Answered

The second path analysis conducted targeted the primary research questions—to explore whether principals’ intentions toward and practices of OTES professional growth plans can be predicted by attitude, subjective norm, and perceived behavior control regarding OTES professional growth plans. The research question results suggested principals’ intentions toward and practices of OTES professional growth plans can be predicted by attitude, subjective norm, and perceived behavior control. The path analysis revealed that principals’ attitudes, subjective norms, and perceived behavior control accounted for 96% of the variance in principals’ intentions toward OTES professional growth plans, and principals’ intentions accounted for 94% of the variance in principals’ behaviors toward OTES professional growth plans.

The second path analysis model collapsed affective attitude and instrumental attitudes (IA/AA) together, and this model also collapsed self-efficacy into perceived behavior control (PBCSE). The second path analysis theory of planned behavior-based model’s goodness of fit estimates showed a satisfactory fit between this proposed theory of planned behavior-based model and the observed data.

The chi-square statistic ($\chi^2 = 4.508$, df = 2, $p = .105$) and the relative chi-square ($\chi^2 / df = 2.254$) were satisfactory. The root-mean square residual (RMR) indicated the amount by which the proposed model estimated variances and covariance’s differ from the observed variances, and covariance registered at .005 (Wuensch, 2016). According to Wuensch (2016), the smaller the RMR the better.

According to Kenny (2015), the Tucker Lewis index (TLI) and the comparative fit index (CFI) depend on the average size of the correlations in the data. Kenny (2015) explained that the TLI shows high when the average correlation between variables is high. The TLI (.994) and CFI
 (.999) fit statistics from the baseline comparisons were both higher than 0.95. The other fit statistics (GFI = .990, NFI = .998, RFI = .989, and IFI = .999) were higher than .95 and contributed to the overall satisfactory fit (Yan & Cheng, 2014).

The RMSEA of this model registered .086 and supports a reasonable error of approximation, or a mediocre fit (Bryne, 2001; Arbuckle, 1999). MacCallum, Browne, and Sugawara (1996) have used 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively. However, according to Kenny (2015), others have suggested 0.10 as the cutoff for poor fitting models. The RMSEA of .086 arguably fell within the adequate fit range. Additionally, the PCLOSE value of .207 indicated that the RMSEA supported an adequate fit. The RMR (.005) supports the fit of the model being tested as well.

Overall, these fit criteria results supported the appropriateness of the proposed theory of planned behavior-based model in predicting Ohio’s principals’ intentions and practices of OTES professional growth plans. These fit criteria provide the basic foundation for the exploration of the parameter estimates within the path analysis model.

**Research Question 2**

The second path model’s acceptable fit measures led to the exploration of the parameter estimates to explore the second research question regarding the predictable power of intentions and perceived behavior control on principals’ practices of OTES professional growth plans. Standardized regression weights of the paths from the two collapsed instrumental attitude and affective attitude (IA/AA) and subjective norm (SNO) to intentions (INT) were significant ($p < .01$). The standardized regression weight of the path from intention (INT) to behavior (BEH) also registered as significant.
The strongest predictor of intention was subjective norm (SNO of $\beta = 0.42$) followed by attitude (IA/AA of $\beta = 0.40$). The paths from perceived behavior control (PBCSE of $\beta = 0.17$) to intention and perceived behavior control to behavior were both not significant, and the path from perceived behavior control (PBCSE of $\beta = 0.00$) to behavior was zero. The standardized regression weight of the path from intention (INT) to behavior (BEH of $\beta = 0.97$) was significant ($p < .01$).

The direct and indirect coefficient effects of predictors on OTES professional growth plan practices (BEH) are presented in the Table 4.9 below. In this model, attitude (via indirect) and perceived behavior control (via direct and indirect) had a similar effect on OTES professional growth plan practices. This model also showed that subjective norm (via indirect) and intentions (via direct) had similar effect on OTES professional growth plan practices. Overall, these results indicated that the theory of planned behavior components had weak effects on principals’ OTES professional development practices.

Table 4.9 Effects of Each Theory of Planned Behavior Component on OTES Professional Growth Plan Practices (Behavior)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Direct effect</th>
<th>Indirect effect</th>
<th>Total effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude</td>
<td>.00</td>
<td>.16</td>
<td>.16</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>.00</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>Perceived Behavior Control</td>
<td>.08</td>
<td>.10</td>
<td>.18</td>
</tr>
<tr>
<td>Intentions</td>
<td>.08</td>
<td>.00</td>
<td>.08</td>
</tr>
</tbody>
</table>

This model accounts for 96% of the variances in principals’ intentions toward OTES professional growth plans and 94% of the variances in principals’ practices of OTES professional growth plans. Armitage and Conner (2001) determined through a meta-analysis on studies that applied the theory of planned behavior that the theory’s components explained 27% of the variance in behavior and 39% of the variance in intentions. In addition, when the behavior
measures were self-reports, as in this study, the theory of planned behavior components accounted for 11% more of the variance in behavior than when behavior measures were observed or objective (Armitage & Conner, 2001).

The hyper-exaggerated percentage of explained variance, especially when compared with the averages explained in Armitage and Conner’s (2001) meta-analysis, suggested the notion that this model suffers from overfit. It is conceivable that the current path model may require too much from the current data set (Babyak, 2004), and the effects of the high correlations among the theory of planned behavior components exhausted the power of the sample size, resulting in the overfit (Kleinbaum et al., 2013). The second path analysis was conducted using AMOS on the same data set of 170 participants. The second path analysis collapsed the instrumental attitude scale together with the affective attitude scale to form one attitude component; and, the second path analysis also collapsed the self-efficacy scale into the perceived behavior control scale to form the perceived behavior control component.

**Research Question 2 Answered**

The second path model’s acceptable fit measures make the parameters estimates and relationships among the latent traits in the theory of planned behavior-based model appropriate for examination to answer the second research question. The second research question investigated whether principals’ OTES professional growth plan practices could be predicted based on principals’ intentions and perceived behavior control toward OTES professional growth plans. Principals’ intentions toward OTES professional growth plans were a significant predictor of principals’ OTES professional growth plan practices; however, principals’ perceived behavior control was not a significant predictor of principals’ OTES professional growth plan practices. The second path model is found below in Figure 4.3.
Figure 4.3 Standardized Result of Second Path Analysis

![Diagram](image.png)

Figure 4.3. Standardized results of the second path analysis. IA/AA represents the instrumental attitude and affective attitude scales, SNO represents the subjective norms scale, PBCSE represents the perceived behavior control combined with the self-efficacy scale, INT represents the intention scale, and BEH represents the behavior scale.

** p < .01.

**Research Questions 3**

A one-way analysis of variance was conducted to determine whether principals’ intentions toward and practices of OTEP professional growth plans differ among Ohio’s typology classifications. Principals were classified into eight groups: rural high poverty (n = 21), rural average poverty (n = 17), small town low poverty (n = 19), small town high poverty (n = 26), suburban low poverty (n = 28), suburban very low poverty (n = 25), urban high poverty (n = 16), and urban very high poverty (n = 18).
There were four outliers found in the principals’ intentions and principals’ typology classification data set. Consequently, the one-way ANOVA was applied twice, with the four outliers and without the outliers, to ensure that the outliers posed no appreciable effect on the ANOVA. The results were compared, and the two results did not differ sufficiently for different conclusions to be drawn from the data. The data were normally distributed for rural average poverty, small town low poverty, suburban low poverty, urban high poverty, and urban very high poverty, as assessed by Shapiro-Wilk’s test ($p > .05$). The typologies of rural high poverty, ($p = .031$), small town poverty ($p = .015$), and suburban very low poverty ($p = .027$) violated the test of normality; however, Maxwell and Delaney (2004) contended that the one-way ANOVA arguably handles deviations from normality, type 1 errors only, with robustness. There was homogeneity of variances, as assessed by Levene’s test for homogeneity of variances ($p = .706$).

Principals’ intentions increased from rural high poverty ($M = 3.0$, $SD = 4.2$), to rural average poverty ($M = 7.2$, $SD = 5.0$), and then decreased to small town low poverty ($M = 5.4$, $SD = 6.2$), increased to small town high poverty ($M = 7.5$, $SD = 5.3$), decreased to suburban low poverty ($M = 6.4$, $SD = 5.7$), decreased again to suburban very low poverty ($M=5.2$, $SD 4.6$), and the last decrease to urban high poverty ($M = 3.4$, $SD = 4.7$) before the final increase to urban very high poverty ($M = 5.4$, $SD = 5.3$), in that order. Although, differences exist among principals’ intentions between the typology classifications, these differences were not statistically significant, $F (7,162) = 2.109$, $p = .045$ based on a $p < .001$.

Finally, a test of between-subjects effects was run to find the partial eta squared. The partial eta squared (.08) revealed that less than 10% of principals’ intentions are accounted for by principals’ assigned typology classification. Figure 4.4 below presents the analysis of variance within the typology classifications for principals’ intentions.
Figure 4.4. Mean principals’ measures of intentions towards OTES professional growth plans within typology classifications showed no statistically significant differences ($p > .001$). Error bars are at the 95% confidence intervals.

A one-way ANOVA was conducted next to determine whether principals’ practices of OTES professional growth plans differ among Ohio’s typology classifications to complete the answer set to the research questions. The principals’ practices (behaviors) and principals’ typology classification data set similarly possessed seven outliers. The same strategy used for the outliers in the intentions-typology data set was repeated for the outliers found in the practices-typology data set. The one-way ANOVA was applied with and without the seven outliers, and the outcomes produced the same conclusions.
The principals’ practice-typology data was normally distributed in the same partial pattern as the principals’ intentions-typology data. The data were normally distributed for rural average poverty ($p = .35$), small town low poverty ($p = .12$), suburban low poverty ($p = .40$), urban high poverty (.19), and urban very high poverty (.06), as assessed by Shapiro-Wilk’s test ($p > .05$). The typologies of rural high poverty, ($p = .040$), small town poverty ($p = .002$), and suburban very low poverty ($p = .000$) violated the test of normality. The principals’ practice-typology data set employed in the one-way ANOVA also relied on Maxwell and Delaney’s (2004) justification for ANOVA compensation of normality deviations. Levene’s test for homogeneity of variances confirmed that there was homogeneity of variances in the data set ($p = .555$).

Principals’ practices (behaviors) registered negatives means for all of the typology classifications: rural high poverty, (M = -1.3, SD = .78) rural average poverty, (M = -.86, SD = .67) small town low poverty, (M =-1.0, SD = 6.2) small town high poverty, (M = -.75, SD = .96) suburban low poverty, (M = -.87, SD = .83) suburban very low poverty, (M = -1.0, SD = .69) urban high poverty, (M = -1.2, SD = .51) and urban very high poverty (M = -1.2, SD = .63). The differences among principals’ practices (behaviors) between the typology classifications were not statistically significant, $F (7,162) = 1.794, p = .092$ based on a $p < .001$.

The test of between-subjects effects was also run on the principals’ practice-typology data set to find the partial eta squared (.07). Ohio principals’ typology classifications accounted for less than 10% of principals’ practices. Figure 4.5 below presents the analysis of variance within the typology classifications for principals’ practices.
Research Question 3 Answered

The final research question explored whether differences in principals’ intentions and practices exist among Ohio’s typology classifications. There were no statistically significant increases or decreases in principals’ mean intentions or practices among the typology classification within Ohio.

Summary

Ohio centralized the application of professional growth plans into the new teacher evaluation framework and outlined the way in which the OTES professional growth plans should
be conducted. However, few researchers have attempted to construct an understanding of OTES professional growth plans. Even fewer attempts have been made to understand the relationships among the variables that influence principals’ intentions and practices regarding OTES professional growth plans.

This study used the theory of planned behavior framework to first examine whether principals’ intentions toward OTES professional growth plans can be predicted by attitudes, subjective norm, and perceived behavior control toward OTES professional growth plans. Second, this study used the theory of planned behavior framework components to examine whether principals’ practices can be predicted by intention and perceived behavior control toward OTES professional growth plans. Finally, an exploration of potential differences among principals’ intentions and practices toward OTES professional growth plans within Ohio’s typology classifications emerged.

This study’s findings revealed that the first hypothesis was generally supported. Principals’ intentions toward OTES professional growth plans can be predicted by attitude, subjective norm, and perceived behavior control regarding OTES professional growth plans. The theory of planned behavior components accounted for around 96% of the variance in principals’ intentions toward OTES professional growth plans. The second hypothesis was also supported. Principals’ intentions and perceived behavior control regarding OTES professional growth plans accounted for 94% of the variances in principals’ OTES professional growth plan practices.

The results demonstrated that attitude and subjective norm were significant predictors of principals’ intentions toward OTES professional growth plans. Principals who had positive attitudes and positive subjective norm beliefs were more likely to have the intentions to conduct
OTES professional growth plans. These results supported the appropriateness of the theory of planned behavior as theoretical framework in explaining the formation of principals’ intentions toward OTES professional growth plans.

Principals’ intentions were most strongly predicted by subjective norm (SNO of $\beta = 0.42$); and also strongly predicted by attitude (IA/AA of $\beta = 0.40$). The predictive power of perceived behavior control (PBCSE of $\beta = 0.17$) was relatively weak. The results revealed that perceived behavior control was not a significant predictor of principals’ intentions to conduct OTES professional growth plans.

Principals’ intentions toward OTES professional growth plans possessed the most powerful predictive path to behavior (INT of $\beta = .97$). The predictive power of intentions was a significant predictor of behavior. The proposed theory of planned behavior-based model found principals’ intentions and perceived behavior control accounted for 94% of variance in principals’ OTES professional growth plans practices. Additionally, principals’ intentions’ toward OTES professional growth plans and practices of OTES professional growth plans did not differ significantly among Ohio’s typology classifications.

The path analysis conducted using the Rasch-calibrated person measures from the five adapted scales demonstrated an arguable goodness of fit based on various model fit scales ($\chi^2 = 4.508, df = 2, p = .105, \text{RMSEA} = .086$). However, the hyper-exaggerated percentages used to predict principals’ intentions and practices do not reflect alignment with past research applications of the theory of planned behavior.
CHAPTER V. CONCLUSIONS AND RECOMMENDATIONS

The purpose of this chapter is to summarize the research study and suggest recommendations for policy, practice, and future research opportunities. The first section of this chapter will discuss an overview of the entire study with emphasis on the objectives of the research and the methodology used to accomplish the quantitative analysis. A summary of the results for the three research questions will be described. The second part of the chapter will discuss policy and practice recommendations from the research findings. Lastly, proposed recommendations for future research as it relates to educator professional growth will be presented.

Review of the Study

A report by Kowalski and Dolph (2015) presented evidence that Ohio principals’ beliefs about the utility of OTES professional growth plans directly contradict the Ohio Department of Education’s thrust to centralize teacher professional growth plans as the crux of OTES. The main purpose of this study was to investigate the OTES professional growth plan implementation gap, described by Kowalski and Dolph (2015), that exists between policy makers and practitioners; and, to initiate a theoretical based assessment tool aimed at unmasking the current state of Ohio principals’ intentions and practices towards OTES professional growth plans.

This study applied the theory of planned behavior as the theoretical base to measure and explain principals’ intentions and practices toward OTES professional growth plans. The theory of planned behavior, which has been applied in various studies in the social sciences, (Yan & Cheng, 2015) was employed as the foundation of this research study. Additionally, an assessment tool based on the theory of planned behavior, and on Yan and Cheng’s (2015) Teacher’s Conceptions and Practices of Formative Assessment Questionnaire, was created to
assess Ohio principals’ intentions and practices toward OTES professional growth plans. This study employed Rasch analysis (Rasch, 1960) and path analysis as the two analytical methods; and, the study’s framework design was modeled after the work of Yan and Cheng (2015).

The Rasch principal calibrated person measures from the instrument were first applied unsuccessfully to the Principals’ OTES Professional Growth Plan Intentions and Practices model path analysis. Next, the same Rasch principal calibrated person measures were reapplied successfully in the traditional theory of planned behavior model path analysis for the examination of principals’ intentions and practices toward OTES professional growth plans. Both models included Ajzen’s (1985) foundational theory of planned behavior components of attitudes, subjective norm, and perceived behavior control.

The entire population of Ohio’s principals (2,431) who served full-time in K-12 facilities during the 2015-2016 school year, were asked to participate in the on-line survey. From the total sampling population 170 usable responses to the instrument emerged. The instrument was administered via QuestionPro, on-line survey software, through an e-mail solicitation that provided a link to the questionnaire.

Prior to answering the research questions application of Rasch analysis thoroughly examined the scales used to assess the five components in the theory of planned behavior framework regarding principals’ intentions and practices toward OTES professional growth plans. Results of the Rasch analysis on the instrument scales indicated the survey provided a valid and reliable instrument capable of empirically measuring principals’ intentions and practices toward OTES professional growth plans. After the Rasch measurement examination of the psychometric scales, Rasch calibrated person measures were subjected to path analysis to answer the first two research questions. Finally, one-way ANOVA’s conducted on principals’
intentions and practices within Ohio’s typology groupings framed the answers to the third research question. The research questions were:

1. Can principals’ intentions toward OTES professional growth plans be predicted by principals’ attitudes, subjective norm, and perceived behavior control toward OTES professional growth plans?

2. Can principals’ OTES professional growth plan practices be predicted based on their intentions and perceived behavior control toward OTES professional growth plans?

3. Do principals’ intentions and practices differ among Ohio’s typology classifications?

**Discussion**

A lot of policy attention in the educational arena has been focused on ways to evaluate individual teachers; and, (Kane, Kerr, & Pianta; 2014) minimal effort has been put toward the understanding of principals’ intentions and practices toward promoting individual professional growth planning within state mandated teacher evaluation processes. The research questions were formulated to quantify the predictive power of attitude, subjective norm, and perceived behavior control in efforts to explore principals’ intentions and practices toward OTES professional growth plans through the scientifically justified theory of planned behavior.

**Research Question 1.**

The findings of this study indicated evidence that it is conceivable that principals’ intentions toward OTES professional growth plans can be predicted by principals’ attitudes, subjective norm, and perceived behavior control toward OTES professional growth plans. The theory of planned behavior was operationalized from Yan and Cheng’s (2015) adapted Teacher’s
Conceptions and Practices of Formative Assessment Questionnaire scales developed to measure the five components in the theory of planned behavior framework. The theory of planned behavior framework components accounted for approximately 96% of the variance in principals’ intentions toward OTES professional growth plans and nearly 94% of the variance in principals’ practices of OTES professional growth plans.

The findings of earlier studies appear to be in general agreement that it is uncommon to find explained variance percentages, or R-squared values, above 50% when predicting human behavior (Frost, 2015). These high R-squared values provided an adequate base for an argument against the predictability of principals’ intentions toward and practices of OTES professional growth plans. The high R-squared values warrant reflection to determine the possibilities for the high-explained variance found in the model.

The extremity of the percentage of the explanation of variance, R-squared values, contained in the results of the first research question give rise to a reasonable argument that the path analysis model suffers from overfit. Babyak (2004) researched the subject of overfitting in regression-type models and put forth that results presented in overfitting models do not occur in the population and therefore will not replicate. According to Frost (2015) an overfit model is a model that is too complex for the data set. Thereby, the concept of overfit could possibly be linked to this study’s sample size. It is important to note that this study’s sample size, 170, met the modest functional requirement for the theory of planned behavior effect size assumptions of 0.3 (Francis et al., 2004) and a probability error of 0.05 with priori power of .8 yielding the minimal suggested sample size of 152 from the G* Power 3.1 calculator. The guidelines for powering a regression-type model help ensure that the model will have adequate power to detect
a relationship and provide a sensibly precise estimate of the strength of that relationship (Frost, 2015).

Recognition of the extremity in explained variance phenomenon in this study raises the possibility that the mild multicollinearity contributed to the inflated R-squared values. Many practitioners do not use the R-squared results due to the vulnerability of biased estimates attributed to sample size and instead use adjusted R-squared results (Frost, 2015). The adjusted R-squared results were slightly lower, yet consistent, with the R-squared results. The adjusted R-squared results explained 94% (compared to R-squared of 96%) of the variance in principals’ intentions and 93% (compared to R-squared of 94%) of principals’ behaviors toward OTES professional growth plans. The pervasive influence of multicollinearity on the explained variance values depicted the classic effects of mild multicollinearity.

Ramifications of the presence of mild multicollinearity in this study should not be oversimplified. It is conceivable that the presence of mild multicollinearity eliminated the significance register of the perceived behavior control predictor within the model. It is also questionable on how much the multicollinearity exhausted the statistical power of the analysis. The only corrective action taken toward the multicollinearity involved linearly combining the instrumental attitude with the affective attitude predictors and collapsing self-efficacy into the perceived behavior control predictor. The underlying rational for not pursuing additional multicollinearity corrective action included the confirmation of the Ridge Regression test that defined the multicollinearity problem as mild and the model satisfied the residual assumptions. Ultimately, there is ample justification in the research to support the claims that the presence of mild multicollinearity does not impact the goodness-of-fit statistics (Frost, 2015).
The results showed that attitude, instrumental and affective attitude combined, and subjective norm were significant predictors of principals’ intentions toward OTES professional growth plans. The descriptive analysis on the Rasch-calibrated principals’ measures indicated that principals had slightly negative attitudes toward OTES professional growth plans (Yan, 2014). In particular, principals expressed they did not like OTES professional growth plans and did not agree that OTES professional growth plans encouraged teachers and evaluators to help each other improve. Moreover, principals did not feel OTES professional growth plans were worth their effort, and principals did not agree that OTES professional growth plans encouraged teachers to engage in self-directed learning. These findings appear to be in general agreement with Kowalski and Dolph’s (2015) contention that Ohio principals view the OTES component of professional growth planning rather negatively.

It should be noted that the predictive power of subjective norm ($\beta = .42$) was slightly higher than the predictive power of attitude ($\beta = .40$); and, this runs contradictory to the largest proportion of studies that support the premises that subjective norm possesses the weakest predictive power toward intentions. Even though the predictive power of subjective norm outranked the predictive power of the attitude component, and this runs counter to the majority of the theory of planned behavior studies, these results are still compatible with the analytical framework of the theory of planned behavior.

Armitage and Conner (2001) emphasized that several researchers argued that the subjective norm component of the theory of planned behavior proved inadequate at predicting intentions. The underlying rational for the strong predictive power of subjective norm in this study might be attributed to the fact that OTES professional growth plans are a mandatory initiative in Ohio. Furthermore, principals may have ranked the subjective norm questions high
because of the social pressure principals feel as a result of the statewide mandate. Additionally, Yan (2014) discussed the influence of mandatory initiatives on the effectiveness of the application of the theory of planned behavior. Other researchers presented evidence that paralleled Yan’s (2014) claim that several studies give rise to the usefulness of applying the theory of planned behavior in mandatory initiatives to investigate human behaviors (Brown, Massey, Montoya-Weiss, & Burkman, 2002).

The subjective norm raw score scales indicated principals’ feelings of social pressure from the Ohio Department of Education to perform OTES professional growth plans ranked high. Specifically, 84% of participating Ohio principals stated that officials at the Ohio Department of Education believe OTES professional growth plans should be implemented. Interestingly principals demonstrated lower feelings of social pressure to perform OTES professional growth plans from their districts’ superintendents with only 66% of participating Ohio principals stating they think their superintendent believes OTES professional growth plans should be implemented. Most noticeably, principals’ feelings of social pressure to perform OTES professional growth plans dropped again with only 31% of participating Ohio principals stating they believe the teachers in their district think they should implement OTES professional growth plans.

Fundamentally, the findings support the reasonable interpretation that principals’ intentions toward OTES professional growth plans can be predicted by principals’ attitudes and subjective norm toward OTES professional growth plans. The notion that principals’ intentions toward OTES professional growth plans can be predicted by principals’ perceived behavior control remains tenable with a non-significant path toward intentions ($\beta = .17$).
**Research Question 2.**

The general theoretical contexts of high R-squared value, power and sample size, and multicollinearity explanations for the first research question are also assumed to be fundamental to the second research question. The second research question sought to answer the following: Can principals’ OTES professional growth plan practices be predicted based on their intentions and perceived behavior control toward OTES professional growth plans? This study found principals’ intentions and perceived behavior control accounted for 94% of the variances in principals’ OTES professional growth plan practices. Principals’ intentions toward OTES professional growth plans strongly predicted principals’ practices of OTES professional growth plans and was significant (β = .97). Most noticeably, the pattern persist that the perceived behavior control component offered zero predictive power toward principals’ OTES professional growth plan practices.

According to Ajzen (1991) perceived behavior control is the extent to which people feel able to engage in the behavior. Perceived behavior control has two aspects: how much a person can exert control over the behavior; and how confident a person feels about executing the behavior (Ajzen, 1991). The perceived behavior control component is determined by the person’s beliefs about the power of both external and internal factors to perform the behavior (Yan, 2014). The more control, both externally and internally, a principal feels about implementing OTES professional growth plans, the more likely he or she will be to do so.

A closer inspection of the perceived behavior control component offers an essential dimension into the perceptions of principals. The perceived behavior control component sheds light on principals’ beliefs about their ability to control external factors related to OTES professional growth plans (I can decide whether or not to implement OTES professional growth plans).
plans), and the internal beliefs about one’s competence (I have sufficient skills to implement OTES professional growth plans). Principals’ Rasch calibrated person mean scores of perceived behavior control registered negatively at -0.58 with survey response scores supporting principals’ perceptions of their external control over OTES professional growth plans as very low. In effect, less then 34% of principals’ agreed that they had enough time to implement OTES professional growth plans. On the contrary, principals’ confidence toward the behavior registered high with 69% responding they had sufficient skills to implement OTES professional growth plans.

In this study, perceived behavior control is the ability to implement OTES professional growth plans. The underlying dynamics within the perceived behavior control component suggest external factors of controllability (time and resources) run contradictory to the internal factors (principals’ confidence in competence to perform OTES professional growth plans). This point is not peculiar to Conley and Glasman (2008) who contended that the balance between external supports and internal improvement pressures is a critical component in the evaluation system; and, the absence of this critical balance produces evaluation systems with little concern for professional growth. By virtue of this observed imbalance of external and internal control factors, it is a conceivable explanation for why the perceived behavior component generates such weak predictive power.

In spite of the weak predictive power of perceived behavior control, the results make a convincing case that principals’ OTES professional growth plan practices could be predicted from principals’ OTES professional growth plan intentions. Whereas, principals’ intentions toward OTES professional growth plans revealed a significant pathway to principals’ practice ($\beta = .97$) of OTES professional growth plans emerged. Additionally, intentions highly correlated with principals’ attitudes, subjective norm, and perceived behavior control. In this respect,
attitude, subjective norm, and perceived behavior control inter-correlate with one another and principals’ intentions reciprocally correlate with principals’ practices toward OTES professional growth plans. In other words, a principal with a positive attitude toward OTES professional growth plans, who recognizes outside influences to support OTES professional growth plans, and who has positive perceived behavior control is more likely to intend to invest effort toward OTES professional growth plans and actually exhibits effort toward OTES professional growth plans. These results substantiate a strong argument for the use of the theory of planned behavior theoretical framework as means to assess Ohio school principals’ intentions and practices toward OTES professional growth plans.

Research Questions 3.

The final research question of this study asked: Do principals’ intentions and practices toward OTES professional growth plans differ among Ohio’s typology classifications? This study found principals’ intentions and practices toward OTES professional growth plans do not differ among Ohio’s typology classifications. In comparison, the results to the final research question were consistent with the present review of the literature assumptions regarding most states and districts are at the beginning stages of the implementation process and experience similar challenges (Behrstock-Sherratt et al., 2013). Consequently, the factors that influence principals’ intentions and practices toward OTES professional growth plans are irrespective of typology classifications. Whereas, principals’ across the state of Ohio appear to have no significant experiential differences in their intentions and practices toward OTES professional growth plan regardless of student enrollment size, poverty levels, and setting.

The underlying commonalities among the typology classifications suggest there is evidence that Ohio principals’ have positive intentions toward OTES professional growth plans.
Principals’ positive levels of intentions are plausibly related to Kowalski and Dolph’s (2015) contention that 76% of Ohio principals identified teacher evaluation as one of their key responsibilities. In essence, this analysis implies that the commonality of principals’ high register of positive intentions across the state may be attributed more to a sense of duty and does not imply an endorsement of OTES professional growth plan. Kowalski and Dolph (2015) dispel the notion that Ohio principals’ positive intentions can be interpreted as an endorsement of OTES professional growth plans with the premise that 52% of Ohio principals disagree that the OTES professional growth plan requirement is effective.

Ohio principals’ practices toward OTES professional growth plan were measured through the allocation of time and money. Ohio principals’ practices toward OTES professional growth plans do not differ among Ohio’s typology classifications; however, the underlying commonalities across every typology classification registered negative mean principal practices toward OTES professional growth plans. Specifically, approximately 27% of Ohio principals responded that in the past twelve months they have frequently engaged in conversations with individual teachers regarding their personal OTES professional growth plans. On the contrary, approximately 42% of Ohio principals responded that in the past twelve months they have sometimes, 17% seldom, 11% rarely, and 3% never engaged in conversations with individual teachers regarding their personal OTES professional growth plans. More over, 57% of Ohio principals responded that no financial resources where specifically allocated toward individual OTES professional growth plans and 30% of Ohio principals reported less then $5,000.00 was allocated to OTES professional growth planning.

Principals’ positive intentions toward OTES professional growth plans and negative practice measures toward OTES professional growth plans is a finding that may bear directly on
the changing workload of principals. According to Alvoid and Black (2014) the new generation of teacher evaluations requires principals to acquire a new skillset that will provide them with the means to instructionally coach, offer meaningful feedback and personalized professional development.

DeMonte and Pennington (2014) discussed the problem of principals’ willingness and abilities to engage in the complex work of new generation evaluation systems; and, summarized evidence that principals’ willingness, or intentions, to carry the implementation load of the new evaluation systems are intact. On the contrary, principals’ supports to build the skills necessary to execute the robust changes surfaced as an after thought throughout the push to implement the next generation evaluation systems (DeMonte and Pennington, 2014). This study’s results tend to highlight the current research in the literature that substantiate principals have positive intentions toward integrating professional development into the evaluation system; however, the structures that facilitate principals to execute those intentions are not in place. The underlying dynamics of the contradiction between principals’ positive intentions and negative practice measures toward OTES professional growth plans must not go unchallenged.

**Conclusion**

The New Teacher Project’s (TNTP) 2009 groundbreaking report, “The Widget” Effect, sparked significant change, across the nation and state, to how teachers are evaluated for the main purpose of improving instruction. In this respect, the literature indicates that there is widespread agreement among educational researchers that teacher evaluation results should be linked to professional development opportunities (Archibald, Coggshall, Croft, & Goe, 2011; Bill & Melinda Gates Foundation, 2010; Council of Chief State School Officers, 2011; Kane & Staiger, 2012). In contrast, the notion that states and districts prioritize evaluation-based
professional development varies, and at best, remains tenable (Archibald, Coggshall, Croft, & Goe, 2011; Bill & Melinda Gates Foundation, 2010; Council of Chief State School Officers, 2011; Kane & Staiger, 2012). Studies conducted by Culver and Hayes (2014) point out that only 31 states attempt to align evaluation results to the professional development provisions of all teachers. The state of Ohio set forth that as a result of the evaluation process, teachers and evaluators should place focus on accelerating and continuing teacher growth through professional development (ODE, 2015). Accordingly, ODE (2015) mandated the provision of the professional growth plan; and the Professional Growth Plan document instructions read:

- Professional development should be individualized to the needs of the teacher, and specifically relate to his/her areas of refinement as identified in the teachers’ evaluation.
- The evaluator should recommend professional development opportunities, and support the teacher by providing resources (e.g., time, financial).

Once this mandate is established it becomes necessary to develop capabilities to quantitatively measure the mandate on both an individual user scale and the population user scale statewide. The ultimate effect of not having the capability to quantitatively measure principals’ intentions and practices toward OTES professional growth plans results in an absence of understanding about the implementation effects of OTES professional growth plans on instructional improvement.

Moreover, without an unbiased tool to measure principals’ intentions and practices toward OTES professional growth plans the initiative, which ODE touts as the crux of the entire evaluation framework, becomes susceptible to token implementation or application. Hence, the main purpose of this study was to attempt to develop a theoretically based instrument that possess the capacity to measure principals’ intentions and practices toward OTES professional
growth plans. The instrument was developed based on the theory of planned behavior and modeled after Yan and Cheng’s (2015) Teacher’s Conceptions and Practices of Formative Assessment Questionnaire; then, operationalized to explore the appropriateness of the applicability of the instrument to measure principals’ intentions and practices toward OTES professional growth plans.

This study significantly contributed to constructing an understanding of principals’ attitudes, intentions, and practices toward OTES professional growth plans. This study sustained that the theory of planned behavior is an applicable theoretical framework to clarify the formation of principals’ intentions and practices toward OTES professional growth plans.

Principals’ attitudes and subjective norm were found to be significant predictors of principals’ intentions toward OTES professional growth plans. Furthermore, principals’ intentions toward OTES professional growth plans demonstrated significant predictive power toward principals’ practices of OTES professional growth plans. However, principals’ perceived behavior control did not demonstrate any significant predictive power in the pathway to principals’ intentions; and, principals’ perceived behavior control demonstrated zero predictive power in the pathway to principals’ practices toward OTES professional growth plans. It is evident that the dimensional structure of the perceived behavior control component needs careful deconstructed to fully understand principals’ perceptions about their ability to perform OTES professional development growth plans. This conclusion does coincide with the controversy found in the review of literature that scrutinize the unitary latent variable into multidimensional attributes; and many theory of planned behavior studies suspect that the role of perceived behavior control as a predictor of intentions is overestimated (Kraft, Rise, Sutton, & Roysamb, 2005).
Recommendations

The application of the theory of planned behavior to the context of Ohio’s mandated initiative of OTES professional growth plans offers great strategic value in understanding principals’ attitudes, subjective norm, and perceived behavior control. Implicit in the theory of planned behavior is the idea that behavioral intention is a function of attitude about the behavior, subjective norm, and perceived behavior control; and, behavioral intention combined with perceived behavior control can predict the outcome behavior (Ajzen, 1991; Yan, 2014). The theory of planned behavior can assist policy makers and practitioners to better understand those who are held with the task of implementing OTES professional growth plans. Additionally, application of the instrument may offer assistance toward building an understanding of the outcomes of the initiative.

According to S. A. Brown et al. (2002) vitality lies in understanding the behaviors held by those who implement mandated initiatives because they, “can delay or obstruct the implementation, and resent, underutilize or sabotage the new system” (p.284). Behrstock-Sherratt, Rizzolo, Laine and Friedman (2013) also emphasized that top-down teacher evaluation reform initiatives, like OTES professional growth plans, are inherent with confrontation, resistance, and skepticism due to the lack of stakeholder participation, or a lack of engaging stakeholders in the process. The theory of planned behavior offers a theoretical framework to help policy makers and practitioners better understand principals’ behaviors within the context of the mandates of OTES professional growth plans. Furthermore, the theory of planned behavior provides a framework to assess the interventions applied to gain desirable outcomes of the tested attitudes, subjective norm, and perceived behavior control variables that influence intentions and behaviors. The theory of planned behavior intervention models offer avenues to unchartered, but
critical, territories of dialogue about the complexities involved in improving teacher professional
development and improving instruction.

The complexity of reconciling, or aligning, principals’ intentions and practices toward
OTES professional growth plans with the Ohio Department of Education’s implementation
vision of OTES professional growth plans is best described by Spirkin’s (2004) contention that
the resolving of a contradictory system is a means to move towards a new, replacement system.
The theory of planned behavior framework offers indispensable conditions for confronting and
re-conceptualizing the barriers to meaningful professional growth. For example, if ODE
attempts to enhance principals’ practices toward OTES professional growth plans, some actions
become necessary to better understand principals’ intentions resulting from their attitudes,
subjective norm and perceived behavior control. Thereby, creating a forum for principals to
develop a clear and deep understanding of their own progress and performance toward
facilitating professional growth for their teachers.

The largest proportion of the studies support the premise that the evidence base for
professional development in general is weak and laden with mixed results (Hill, Beisiegel, Jacob,
2013). There has been no evidence that the statewide implementation of OTES professional
growth plans have produced improved instructional practices. Herein lies the most significant
recommendation proposed by this study. The theory of planned behavior offers a framework to
assess the capacity of the professional growth plan to promote teacher growth at the local level of
schools. The strength and power of the theory of planned behavior is in its adaptive ability to
offer a structural understanding of both principals’ and teachers’ perceptions on and responses to
existing professional learning supports. Ultimately, the theory of planned behavior provides a
framework to foster much needed new dialogue about how to obtain and sustain educators’ professional growth.

**Future Research Opportunities**

The implications of this study should inform us further about supporting principals’ capacity to foster evaluation-based professional growth in teachers. The understanding of principals’ intentions and practices toward OTES professional growth plans in this study is, therefore, expected to tender a platform for future studies, especially in the domains of instructional leadership.

This study potentially holds important implications for future ramifications through the study’s provision of tools for replication of the study with Ohio teachers’ intentions and practices toward OTES professional growth plans. Additionally, the study could be replicated to build an understanding about the intentions and practices of state support staff, superintendents, and other district specialist in charge of curriculum and instruction toward improving teacher development. Moreover, the study replication could benefit from a mixed methods approach that delves into details of the participants’ intentions and practices toward fostering professional growth for one’s self and others.

The final recommendation for future research stems from TNTP’s 2015 report entitled “The Mirage”; and the presented evidence that our nation, states, and districts are spending a lot of money on helping teachers develop and virtually none of it is helpful. The significance of this study is that it dispels the widely held myth that educators already know how to help teachers improve; and the research base for what actually helps teachers improve is extremely thin (TNTP, 2015). Hence, there evidence to suggest future research should build on strong theoretical frameworks such as Ajzen’s (1991) theory of planned behavior. As a result, in
addition to replicating this study with other populations, it would be interesting to compare the
theory of planned behavior to the self-determination (Deci & Ryan, 1985) theory. Deci and
Ryan’s (1985) self-determination theory more specifically examines causality concerning
people’s inherent growth tendencies (Flaste, 1995) and might also add to our understanding of
the behaviors that forecasts professional growth.
REFERENCES


Alvoid, L., Black, Jr., W. L. (2014). *The changing role of the principal: How high-achieving districts are recalibrating school leadership*. Retrieved from:


Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International journal of testing, 1*(1), 55–86.


Johnson, S. M. (2012). Build the capacity of teachers and their schools: Improve the quality of teachers and teaching by investing in individual teachers and assessing their
effectiveness while also improving the schools where they work. *Phi Delta Kappan*, 94(2), 62.


Assessing teacher quality: Understanding teacher effects on instruction and achievement 
(pp. 33-48). New York: Teachers College Press.

Educational Policy. doi: 0895904811426866.

Kowalski, T. J., & Dolph, D. A. (2015). Principal dispositions regarding the Ohio teacher 
evaluation system. AASA Journal of Scholarship and Practice, 11(4). Retrieved from 
http://ecommons.udayton.edu/eda_fac_pub/16/ .

Thousands Oaks, CA: SAGE.

planned behaviour: Perceived behavioural control or affective attitude?. British Journal 

structural equation modeling: Sample size planning for narrow confidence intervals. 
Psychological Methods, 16(2), 127. doi: 10.1037/a0021764.

Landeck, E. S. (2006). Using a theory of planned behavior approach to assess principals’
professional intentions to promote diversity awareness beyond the level recommended by 
their district (Doctoral dissertation, Texas A&M University). Retrieved from 
oaktrust.library.tamu.edu/bitstream/.../LANDECK- DISSENTATION.pdf .


New America Foundation. (2014). *No Child Left Behind overview*. Retrieved from
https://www.febp.newamerica.net/background../no-child-left-behind-overview


Ohio Department of Education. (2010). *Fifth to first Ohio’s race to the top strategy Selection criteria and competition priorities* [Race to the Top Application for Phase 2 Funding CFDA Number: 84.395A]. Retrieved from

Ohio Department of Education. (2014). *New policy updates*. Retrieved from
http://education.ohio.gov.


Senunyeme, E., K., (2013). Business research methods. [PowerPoint presentation]. Retrieved from Regent University of Science and Technology website:


Educational Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.
APPENDIX A

THE UNIVERSITY OF FINDLAY
Institutional Review Board

Date:  September 21, 2015
To:    Dr. Kathleen Crates
CC:    Kelly Schooler
RE:    Predicting Ohio Principals’ Intentions and Practices toward State Evaluation-based Professional Growth Plans Using the Theory of Planned Behavior

Project Expiration date:  September 21, 2016

The University of Findlay Institutional Review Board (IRB) has completed its review of your project utilizing human subjects and has granted authorization. This study has been approved for a period of one year only. The project has been assigned the number 941.

In order to comply with UF policy and federal regulations, human subject research must be reviewed by the IRB on at least a yearly basis. If you have not completed your research within the year, it is the investigator’s responsibility to ensure that the Progress Report is completed and sent to the IRB in a timely fashion. The IRB needs to process the re-approval before the expiration date, which is printed above.

Understand that any proposed changes may not be implemented before IRB approval, in which case you must complete an Amendment/Modification Report.

Following the completion of the use of human subjects, the primary investigator must complete a Certificate of Compliance form indicating when and how many subjects were recruited for the study.

Please refer to the IRB guidelines for additional information. This packet can be obtained within blackboard under community section. Please note that if any changes are made to the present study, you must notify the IRB immediately. Please include that number on any other documentation or correspondence regarding the study.

Thank you very much for your cooperation. If you have any questions, please feel free to contact IRB at (419) 434-4640 or email irb@findlay.edu.

Sincerely,

[Signature]

Jennifer Fennema-Bloom, Ed.D.
Chair, Institutional Review Board

Cc:  IRB Office
APPENDIX B

OTES EVALUATOR SURVEY

Exploring Ohio Principals’ Practices
Toward OTES Professional Growth Plans

Share Your Voice

Survey window opens Friday, October 2nd, 2015

---

Dear Ohio Principal,

I am writing to ask for your help in learning more about principals’ practices toward OTES professional growth plans for my dissertation work at The University of Findlay. In the upcoming weeks you will receive an email requesting your participation in an online survey.

Your answers to the online survey questions will be anonymous. The survey takes approximately 10 minutes to complete and has been approved by The University of Findlay’s Institutional Review Board.

Please consider contributing your voice to the larger body of knowledge about principals’ practices toward OTES professional growth plans. Your insights as an educational leader in the state of Ohio are genuinely appreciated.

Best regards,

Kelly L. Schooller
Doctoral Candidate
schoolerh@findlay.edu
The University of Findlay
October 2, 2015

Dear Ohio Principal,

You have been invited to participate in a study of Ohio’s implementation of OTES professional growth plans. I hope to learn without prejudice principals’ intentions and behaviors regarding the implementation of OTES professional growth plans. You were selected as a possible participant in this study because of your experience with Ohio’s Teacher Evaluation System.

Your return of this survey is implied consent. The survey is designed to determine the effectiveness of Ohio’s professional growth plan implementation. It will take less than 10 minutes to complete. This survey has a Respondent Anonymity Assurance guarantee from the survey platform provider QuestionPro. All records of individual responses are unidentifiable in this study.

No benefits accrue to you for answering the survey, but your responses will be used to determine the effectiveness of Ohio’s professional growth plan implementation. Any discomfort or inconvenience to you derives only from the amount of time taken to complete the survey. The University of Findlay's Institutional Review Board approved this survey and the project has been assigned the number 941.

Your decision whether or not to participate will not prejudice any future relationships with The University of Findlay. If you decide to participate, you are free to discontinue participation at any time without prejudice. If you have any questions, please ask. If you have any additional questions later, contact Dr. Kathleen Crates at crates@findlay.edu.

Thank you for your time.

Sincerely,

Kelly Schooler
Doctoral Candidate
The University of Findlay
schoolerk@findlay.edu
Table 1

New Generation Early Implementation Teacher Evaluation Results from Three States

<table>
<thead>
<tr>
<th>State</th>
<th>Percentage of Teachers Rated Effective or Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delaware</td>
<td>99%</td>
</tr>
<tr>
<td>Florida</td>
<td>97%</td>
</tr>
<tr>
<td>New York</td>
<td>95%</td>
</tr>
</tbody>
</table>
Table 3.3

2013 Ohio School District Typology

<table>
<thead>
<tr>
<th>Typology Code</th>
<th>Major Grouping</th>
<th>Descriptor</th>
<th>Districts within Typology</th>
<th>Students within Typology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rural</td>
<td>High student poverty, small student population size</td>
<td>124</td>
<td>170,000</td>
</tr>
<tr>
<td>2</td>
<td>Rural</td>
<td>Average student poverty, very small student population size</td>
<td>107</td>
<td>110,000</td>
</tr>
<tr>
<td>3</td>
<td>Small Town</td>
<td>Low student poverty, small student population size</td>
<td>111</td>
<td>185,000</td>
</tr>
<tr>
<td>4</td>
<td>Small Town</td>
<td>High student poverty, average student population size</td>
<td>89</td>
<td>200,000</td>
</tr>
<tr>
<td>5</td>
<td>Suburban</td>
<td>Low student poverty, average student population size</td>
<td>77</td>
<td>320,000</td>
</tr>
<tr>
<td>6</td>
<td>Suburban</td>
<td>Very low student poverty, large student population size</td>
<td>46</td>
<td>240,000</td>
</tr>
<tr>
<td>7</td>
<td>Urban</td>
<td>High student poverty, average student population size</td>
<td>47</td>
<td>210,000</td>
</tr>
<tr>
<td>8</td>
<td>Urban</td>
<td>Very high student poverty, very large student population</td>
<td>8</td>
<td>200,000</td>
</tr>
</tbody>
</table>
Table 3.4
Cronbach's Alpha Reliability Test

<table>
<thead>
<tr>
<th>Scale</th>
<th>Number of Items</th>
<th>Cronbach’s ( \alpha ) Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective Attitude</td>
<td>4</td>
<td>0.750</td>
</tr>
<tr>
<td>Instrumental Attitude</td>
<td>3</td>
<td>0.768</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>4</td>
<td>0.802</td>
</tr>
<tr>
<td>Scale</td>
<td>3</td>
<td>0.685</td>
</tr>
<tr>
<td>Controllability Scale</td>
<td>5</td>
<td>0.782</td>
</tr>
<tr>
<td>Self-Efficacy Scale</td>
<td>6</td>
<td>0.729</td>
</tr>
<tr>
<td>Intention Scale</td>
<td>4</td>
<td>0.744</td>
</tr>
<tr>
<td>Behavior Scale</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3

Final Survey Instrument

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affective Attitude</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 I like OTES professional growth plans.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2 OTES professional growth planning is an enjoyable process.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3 OTES professional growth planning encourages teachers and evaluators to help each other improve.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4 OTES professional growth planning is worth my effort.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Instrumental Attitude Scale**

Q5 OTES professional growth planning can encourage teacher to engage in self-directed learning.

Q6 OTES professional growth planning can raise teachers’ interest in teaching.

Q7 OTES professional growth planning can offer a fair appraisal of a teacher’s professional development needs.

**Subjective Norm**

As far as I know, the following stakeholders believe that OTES professional growth planning should be implemented.

Q8 Officials in the state education department

Q9 The superintendent of my district

Q10 The teachers in my school

Q11 My administrative colleagues

**Controllability Scale**

Q12 I can decide the frequency of implementing OTES professional growth plans.

Q13 I can decide the timetable for implementing OTES professional growth plans.

Q14 I can decide whether or not to implement OTES professional growth plans.

**Self-Efficacy Scale**

Q15 I can integrate OTES professional growth plans into the OTES evaluation cycle process.

Q16 I have received sufficient training to implement OTES professional growth plans.

Q17 I can assist teachers in designing appropriate goals for OTES professional growth plans.

Q18 I have enough time to implement OTES professional growth plans.

Q19 I have sufficient supporting materials (e.g., online modules, handbooks, DVD) to implement OTES professional growth plans.

Q20 I have sufficient skills to implement OTES professional growth plans.

**Intention Scale**

Q21 I am willing to try and implement OTES professional growth plans.
Q22  I am willing to integrate OTES professional growth plans into the evaluation cycle.
Q23  I am willing to assist teachers in designing appropriate goals for OTES professional growth plans.
Q24  I am willing to put effort toward implementing OTES professional growth plans.
Q25  I am willing to adjust traditional approaches to professional development to meet the requirements of OTES professional growth plans.
Q26  I am willing to encourage teachers to participate in professional growth plans.

Behavior Scale
Q27  In the past six months, how often have you engaged in developing OTES professional growth plans?
   1=Everyday 2=Almost every day 3 =Most days 4=A number of days, but less than half of the time 5=Some days 6=Never
Q28  Please estimate how frequently you have had conversations with individual teachers specifically regarding their personal OTES professional growth plans in the past six months.
   1=Very frequently 2= Frequently 3=Sometimes 4=Seldom 5=Rarely 6=Never
Q29  Please estimate the amount of financial resources you have specifically allocated toward individual teachers’ OTES professional growth plans (not improvement plans) in the past six months.
   1= more than $20,000   2=between $20,000-$15,000  3=$15,000-$10,000 4=$10,000-$5,000  5=less than $5,000 6=no financial resources allocated specifically toward individual teachers’ OTES professional growth plans
Q30  How often have you provided time for teachers to specifically develop the skills necessary to achieve the goals stated in their individual OTES professional growth plans in the past six months?
   1= Everyday  2=Almost every day 3=Most days 4=A number of days, but less than half of the time 5=Some days 6=Never

Demographically Descriptive Items
Q31  Please click the link to identify your district’s typology code.
Q32  Please select the category that best describes your OTES caseload.
   1-7 teachers  8-15 teachers  16-23 teachers  24-31 teachers  >32 teachers
### Table 3.4
Research Questions and Corresponding Statistical Tests

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Statistical Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can principals’ intentions toward OTES professional growth plans be predicted by</td>
<td>Principals’ Intentions</td>
<td>Direct Measures of Principals’ Attitude, Subjective Norm, and Perceived Behavior Control</td>
<td>Rasch Analysis</td>
</tr>
<tr>
<td>principals’ attitudes, subjective norms, and perceived behavior control regarding</td>
<td></td>
<td></td>
<td>Correlation</td>
</tr>
<tr>
<td>OTES professional growth plans?</td>
<td></td>
<td></td>
<td>Path Analysis</td>
</tr>
<tr>
<td>Can principals’ OTES professional growth plan practices be predicted by intentions</td>
<td>Principals’ Practices</td>
<td>Direct Measures of Principals’ Intentions and Perceived Behavior Control</td>
<td>Rasch Analysis</td>
</tr>
<tr>
<td>and perceived behavior control toward OTES professional growth plans?</td>
<td></td>
<td></td>
<td>Correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Path Analysis</td>
</tr>
<tr>
<td>Do principals’ intentions and practices differ among the typology classifications?</td>
<td>Principals’ Intentions and Practices</td>
<td>Typology</td>
<td>ANOVA</td>
</tr>
</tbody>
</table>
Table 4.1

2013 Ohio Typology Classification Principals’ Representation

<table>
<thead>
<tr>
<th>Typology Code</th>
<th>Major Grouping</th>
<th>Full Descriptor</th>
<th>Raw Number of Principals</th>
<th>Percent of Principals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rural</td>
<td>High Student Poverty and Small Student Population</td>
<td>21</td>
<td>12.4%</td>
</tr>
<tr>
<td>2</td>
<td>Rural</td>
<td>Average Student Poverty and Very Small Student Population</td>
<td>17</td>
<td>10.0%</td>
</tr>
<tr>
<td>3</td>
<td>Small Town</td>
<td>Low Student Poverty and Small Student Population</td>
<td>19</td>
<td>11.2%</td>
</tr>
<tr>
<td>4</td>
<td>Small Town</td>
<td>High Student Poverty and Average Student Population</td>
<td>26</td>
<td>15.3%</td>
</tr>
<tr>
<td>5</td>
<td>Suburban</td>
<td>Low Student Poverty and Average Student Population Size</td>
<td>28</td>
<td>16.5%</td>
</tr>
<tr>
<td>6</td>
<td>Suburban</td>
<td>Very Low Student Poverty And Large Student Population</td>
<td>25</td>
<td>14.7%</td>
</tr>
<tr>
<td>7</td>
<td>Urban</td>
<td>High Student Poverty and Average Student Population</td>
<td>16</td>
<td>9.4%</td>
</tr>
<tr>
<td>8</td>
<td>Urban</td>
<td>Very High Student Poverty and Very Large Student Population</td>
<td>18</td>
<td>10.5%</td>
</tr>
</tbody>
</table>
Table 4.2. Rasch-Analysis Item Fit Indices

<table>
<thead>
<tr>
<th>Question Item:</th>
<th>Rasch-Analysis Fit Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INFIT</td>
</tr>
<tr>
<td></td>
<td>MNSQ</td>
</tr>
<tr>
<td>1. I like OTES professional growth plans.</td>
<td>0.88</td>
</tr>
<tr>
<td>2. OTES professional growth planning is an enjoyable process.</td>
<td>0.82</td>
</tr>
<tr>
<td>3. OTES professional growth planning encourages teachers and evaluators to help each other improve.</td>
<td>0.92</td>
</tr>
<tr>
<td>4. OTES professional growth planning is worth of my effort.</td>
<td>0.76</td>
</tr>
<tr>
<td>5. OTES professional growth planning can encourage self-directed learning of teachers.</td>
<td>1.08</td>
</tr>
<tr>
<td>6. OTES professional growth planning can raise teachers’ interest in teaching.</td>
<td>1.02</td>
</tr>
<tr>
<td>7. OTES professional growth planning can offer a fair appraisal of individual teacher’s professional development needs.</td>
<td>1.59</td>
</tr>
<tr>
<td>8. As far as I know, the officials of the state education department believe that OTES professional growth planning should be implemented.</td>
<td>1.49</td>
</tr>
<tr>
<td>9. As far as I know, the superintendent of my district believes that OTES professional growth planning should be implemented.</td>
<td>0.96</td>
</tr>
<tr>
<td>10. As far as I know, the teachers in my school believe that OTES professional growth planning should be implemented.</td>
<td>0.72</td>
</tr>
<tr>
<td>11. As far as I know, my administrative colleagues believe that OTES professional growth planning should be implemented.</td>
<td>0.66</td>
</tr>
<tr>
<td>12. I can decide the frequency of implementing OTES professional growth plans.</td>
<td>1.14</td>
</tr>
<tr>
<td>13. I can decide the timing of implementing OTES professional growth plans.</td>
<td>0.92</td>
</tr>
<tr>
<td>14. I can decide whether or not to implement OTES professional growth plans.</td>
<td>0.88</td>
</tr>
<tr>
<td>Question</td>
<td>Rating</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>15. I can integrate OTES professional growth plans into the OTES evaluation cycle process.</td>
<td>1.63</td>
</tr>
<tr>
<td>16. I have received sufficient training to implement OTES professional growth plans.</td>
<td>0.82</td>
</tr>
<tr>
<td>17. I can assist teachers in designing appropriate goals for OTES professional growth plans.</td>
<td>0.84</td>
</tr>
<tr>
<td>18. I have enough time to implement OTES professional growth plans.</td>
<td>1.37</td>
</tr>
<tr>
<td>19. I have sufficient supporting materials (e.g., online modules, handbooks, DVDs) to implement OTES professional growth plans.</td>
<td>0.91</td>
</tr>
<tr>
<td>20. I have sufficient skills to implement OTES professional growth plans.</td>
<td>1.04</td>
</tr>
<tr>
<td>21. I am willing to try to implement OTES professional growth plans.</td>
<td>1.07</td>
</tr>
<tr>
<td>22. I am willing to integrate OTES professional growth plans into the evaluation cycle.</td>
<td>0.85</td>
</tr>
<tr>
<td>23. I am willing to assist teachers in designing appropriate goals for OTES professional growth plans.</td>
<td>0.94</td>
</tr>
<tr>
<td>24. I am willing to put effort toward implementing OTES professional growth plans.</td>
<td>0.71</td>
</tr>
<tr>
<td>25. I am willing to adjust traditional approaches to professional development to meet the requirements of OTES professional growth plans.</td>
<td>1.50</td>
</tr>
<tr>
<td>26. I am willing to encourage teachers to participate in professional growth plans.</td>
<td>0.96</td>
</tr>
<tr>
<td>27. In the past 6 months, how often have you engaged in developing OTES professional growth plans?</td>
<td>0.94</td>
</tr>
<tr>
<td>28. In the past 6 months, please estimate how frequently you have had conversations with individual teachers specifically regarding their personal OTES professional growth plans.</td>
<td>0.73</td>
</tr>
<tr>
<td>29. In the past 6 months, please estimate the amount of</td>
<td></td>
</tr>
</tbody>
</table>
financial resources you have allocated specifically toward individual teachers’ OTES professional growth plans (not improvement plans).

30. In the past 6 months, how often have you provided time for teachers to specifically develop the skills necessary to achieve the goals stated in their individual OTES professional growth plans?

<table>
<thead>
<tr>
<th>1.34</th>
<th>1.29</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.16</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Table 4.3 Psychometric Properties of Measurement Scales for Measuring Components of TPB

<table>
<thead>
<tr>
<th>Scale</th>
<th>Number of Items</th>
<th>Rasch Person Reliability</th>
<th>Rasch Item Reliability</th>
<th>Variance Explained by Measures</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>4</td>
<td>0.91</td>
<td>0.97</td>
<td>82.0%</td>
<td>.93</td>
</tr>
<tr>
<td>IAT</td>
<td>2</td>
<td>0.85</td>
<td>0.98</td>
<td>76.4%</td>
<td>.85</td>
</tr>
<tr>
<td>SNO</td>
<td>4</td>
<td>0.81</td>
<td>0.99</td>
<td>73.4%</td>
<td>.77</td>
</tr>
<tr>
<td>PBC</td>
<td>3</td>
<td>0.80</td>
<td>0.98</td>
<td>70.1%</td>
<td>.81</td>
</tr>
<tr>
<td>SEF</td>
<td>5</td>
<td>0.86</td>
<td>0.98</td>
<td>70.0%</td>
<td>.85</td>
</tr>
<tr>
<td>INT</td>
<td>6</td>
<td>0.88</td>
<td>0.87</td>
<td>70.1%</td>
<td>.92</td>
</tr>
<tr>
<td>BEH</td>
<td>2</td>
<td>0.74</td>
<td>1.00</td>
<td>83.2%</td>
<td>.71</td>
</tr>
</tbody>
</table>
Table 4.4.

Step Thresholds of Measurement Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>-7.01</td>
<td>-2.82</td>
<td>-1.73</td>
<td>2.26</td>
<td>9.29</td>
</tr>
<tr>
<td>IAT</td>
<td>-5.18</td>
<td>-2.16</td>
<td>-1.16</td>
<td>1.79</td>
<td>7.15</td>
</tr>
<tr>
<td>SNO</td>
<td>-2.80</td>
<td>-0.41</td>
<td>-1.11</td>
<td>0.35</td>
<td>3.97</td>
</tr>
<tr>
<td>PBC</td>
<td>-5.02</td>
<td>-0.59</td>
<td>-0.58</td>
<td>0.41</td>
<td>5.79</td>
</tr>
<tr>
<td>SEF</td>
<td>-2.49</td>
<td>-0.58</td>
<td>-1.27</td>
<td>0.37</td>
<td>3.98</td>
</tr>
<tr>
<td>INT</td>
<td>-3.14</td>
<td>-2.31</td>
<td>-2.24</td>
<td>0.71</td>
<td>6.98</td>
</tr>
<tr>
<td>BEH</td>
<td>-8.71</td>
<td>-3.85</td>
<td>-0.45</td>
<td>2.51</td>
<td>10.50</td>
</tr>
</tbody>
</table>
Table 4.5. Means, Standard Deviations, and Correlations of Theory of Planned Behavior Components Seven Scales

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>AAT</th>
<th>IAT</th>
<th>SNO</th>
<th>PBC</th>
<th>SEF</th>
<th>INT</th>
<th>BEH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>-0.78</td>
<td>4.28</td>
<td>---</td>
<td>0.98</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>IAT</td>
<td>-0.69</td>
<td>3.27</td>
<td>0.98</td>
<td>---</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>0.97</td>
<td>0.93</td>
</tr>
<tr>
<td>SNO</td>
<td>+1.46</td>
<td>2.79</td>
<td>0.95</td>
<td>0.95</td>
<td>---</td>
<td>0.90</td>
<td>0.97</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>PBC</td>
<td>-2.42</td>
<td>2.88</td>
<td>0.94</td>
<td>0.94</td>
<td>0.90</td>
<td>---</td>
<td>0.91</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td>SEF</td>
<td>+0.02</td>
<td>1.99</td>
<td>0.95</td>
<td>0.94</td>
<td>0.97</td>
<td>0.91</td>
<td>---</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>INT</td>
<td>+5.33</td>
<td>5.48</td>
<td>0.98</td>
<td>0.97</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
<td>---</td>
<td>0.94</td>
</tr>
<tr>
<td>BEH</td>
<td>-1.04</td>
<td>0.78</td>
<td>0.94</td>
<td>0.93</td>
<td>0.97</td>
<td>0.90</td>
<td>0.97</td>
<td>0.94</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 4.6. Means, Standard Deviations, and Correlations of Theory of Planned Behavior

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>AAT</th>
<th>SNO</th>
<th>PBC</th>
<th>INT</th>
<th>BEH</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>-0.49</td>
<td>3.69</td>
<td>----</td>
<td>0.96</td>
<td>0.94</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>SNO</td>
<td>+1.46</td>
<td>2.79</td>
<td>0.96</td>
<td>----</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>PBC</td>
<td>-0.58</td>
<td>1.10</td>
<td>0.94</td>
<td>0.91</td>
<td>----</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>INT</td>
<td>+5.33</td>
<td>5.48</td>
<td>0.97</td>
<td>0.93</td>
<td>0.91</td>
<td>----</td>
<td>0.92</td>
</tr>
<tr>
<td>BEH</td>
<td>-1.04</td>
<td>0.78</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.92</td>
<td>----</td>
</tr>
</tbody>
</table>
Table 4.7. Multicollinearity Tolerance Test and Variance Inflation Factor Test

<table>
<thead>
<tr>
<th>Variables</th>
<th>Tolerance</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude</td>
<td>0.02</td>
<td>36.95</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>0.06</td>
<td>15.42</td>
</tr>
<tr>
<td>Perceived Behavior Control</td>
<td>0.11</td>
<td>8.99</td>
</tr>
<tr>
<td>Intention</td>
<td>0.05</td>
<td>17.23</td>
</tr>
</tbody>
</table>
Table 4.8. Ridge Regression Analysis Report of Eigenvalues of Correlations

<table>
<thead>
<tr>
<th>Number of Eigenvalue</th>
<th>Eigenvalue</th>
<th>Incremental Percent</th>
<th>Cumulative Percent</th>
<th>Condition Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.827345</td>
<td>95.68</td>
<td>95.68</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.092181</td>
<td>2.30</td>
<td>97.99</td>
<td>41.52</td>
</tr>
<tr>
<td>3</td>
<td>0.060900</td>
<td>1.52</td>
<td>99.51</td>
<td>62.85</td>
</tr>
<tr>
<td>4</td>
<td>0.019574</td>
<td>0.49</td>
<td>100.00</td>
<td>195.23</td>
</tr>
</tbody>
</table>
Table 4.9 Effects of Each Theory of Planned Behavior Component on OTES Professional Growth Plan Practices (Behavior)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Direct effect</th>
<th>Indirect effect</th>
<th>Total effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude</td>
<td>.00</td>
<td>.16</td>
<td>.16</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>.00</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>Perceived Behavior Control</td>
<td>.08</td>
<td>.10</td>
<td>.18</td>
</tr>
<tr>
<td>Intentions</td>
<td>.08</td>
<td>.00</td>
<td>.08</td>
</tr>
</tbody>
</table>
Figure 2.1

Principals' OTES Professional Growth Plan Intentions and Practices
Figure 4.1 Person Item Map

Figure 4.1. The symbol “#” represents 2 and the symbol “.” represents 1.
Figure 4.2. Failed Theory of Planned Behavior-Based Model 1

Figure 4.2. Failed theory of planned behavior-based model 1. IA represents the instrumental attitude, AA represents the affective attitude scale, SNO represents the subjective norms scale, PBC represents the perceived behavior control, SEF represents the self-efficacy scale, INT represents the intention scale, and BEH represents the behavior scale.
Figure 4.3 Standardized Result of Second Path Analysis

Figure 4.3. Standardized results of the second path analysis. IA/AA represents the instrumental attitude and affective attitude scales, SNO represents the subjective norms scale, PBCSE represents the perceived behavior control combined with the self-efficacy scale, INT represents the intention scale, and BEH represents the behavior scale. ** p < .01.
Figure 4.4. Analysis of Variance for Intentions Typology Classifications

Figure 4.4. Mean principals’ measures of intentions towards OTES professional growth plans within typology classifications showed no statistically significant differences ($p > .001$). Error bars are at the 95% confidence intervals.
Figure 4.5. Analysis of Variance for Behaviors Typology Classifications