

A PRINCIPAL EVALUATION PROGRAM: PERSPECTIVES FROM THE FIELD
ABOUT REFLECTIVE PRACTICE AND PROFESSIONAL DEVELOPMENT

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctorate of Philosophy

with a

Major in Education

in the

College of Graduate Studies

University of Idaho

by

Stephen A. Robertson

Major Professor: Kathryn Canfield-Davis, Ph.D.

Co-Major Professor: Paul Goldman, Ph.D.

Committee Members: Mary E. Gardiner, Ph.D.; Michael Kroth, Ph.D.;

Department Administrator: Kathryn Canfield-Davis, Ph.D.

June 2016

AUTHORIZATION TO SUBMIT DISSERTATION

The dissertation of Stephen Robertson, submitted for the degree of Doctor of Philosophy with a major in Education and titled, "A PRINCIPAL EVALUATION PROGRAM: PERSPECTIVES FROM THE FIELD ABOUT REFLECTIVE PRACTICE AND PROFESSIONAL DEVELOPMENT," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____

Kathryn Canfield-Davis, Ph.D.

Co-Major Professor: _____ Date: _____

Paul Goldman, Ph.D.

Committee Members: _____ Date: _____

Mary Gardiner, Ph.D.

_____ Date: _____

Michael Kroth, Ph.D.

Department

Administrator _____ Date: _____

Kathryn Canfield-Davis, Ph.D.

Abstract

The purpose of this basic qualitative study was to describe how well the “promise” of Idaho’s new model of performance evaluation of public school principals (IDAPA 121) compares with its actual impact on principals’ reflective practice, professional development, and leadership practices. Research questions that shaped the study explored (a) the impact of the professional practice standards and student growth measure of IDAPA 121, and (b) what circumstances contributed most significantly to reflective practice and professional development of Idaho school principals.

A review of the literature revealed five key themes and trends that shaped this study’s research questions and methodology: (a) perceived problems with traditional principal evaluation before the federal Race to the Top and ESEA Waiver program; (b) the promise of a reformed principal evaluation system to improve school leadership, teaching and student learning; (c) best practice recommendations to provide a meaningful and impactful evaluation experience for principals; (d) criticisms of the new model of principal evaluations; and (e) what the literature does and does not tell us about principal evaluation and its effect on reflective practice, professional development and leadership behaviors and practices.

Data were collected through semi-structured interviews of ten active public school principals from two suburban Idaho school districts. Five participants were elementary level principals and five were middle school principals. Data analysis revealed four major findings:

1. Participants experienced evaluation under IDAPA 121 as a perfunctory, isolated end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice.

2. The student growth component of IDAPA 121, and the prospect of being evaluated partly on student growth had no impact on the participants' leadership focus or practices.
3. Participants believe that policy makers and the school policy related to their performance evaluation are out of touch with what happens in schools.
4. Participants had clear ideas for improving the system so principals would experience more meaningful reflective practice and professional development.

A number of possible explanations for these findings are explored. These include (a) operational issues with implementation of IDAPA 121, and (b) deeper, more substantive issues that raise questions about the viability of using standards-based evaluations to drive effective professional development, and possible conflicts between IDAPA 121's summative and formative objectives. Recommendations for possible policy changes and further research conclude this study.

Acknowledgments

Completion of this study would have been impossible without the support of many incredible family members, friends and colleagues. Special thanks go to the following:

My major professor, Dr. Kathy Canfield-Davis, who stepped in just when I needed her with the perfect combination of rigor, caring and humor; my ‘other’ major professor Dr. Paul Goldman, whose experience, energy, and love of ideas helped me find a much-needed focus that I could carry to completion; Dr. Mary Gardiner who offered helpful suggestions to improve my proposal and opened my eyes to the power of qualitative research. Dr. Michael Kroth who brought a valuable organizational development perspective and experience.

My amazing family – this accomplishment is truly yours as much as it is mine. My incredible wife, Susan, has supported my dreams from day one, pushed me when I needed a gentle nudge, and picked-up the slack too many times to count. You are the best life-partner one could hope for. My father, Art, and parents-in-law Jim and LouAnn Dahl helped me in countless ways and provided tireless support to Susan and our kids. Our amazing children, Eli, Sophia, and Sean have encouraged and cheered me on even when they knew it meant I would be less available than they would have liked. Dad is finally done! My sisters Jeanne and Diana have always been there for me, believing in me, and sharing their kind and loving wisdom.

My many good friends who have hung in there with me over this long journey, always remembering to ask ‘how’s school going’ and whose support I felt no matter the geographic distance between us. Gary Trubell, Mike Cole, Rob Gallo, Nancy Young, Alex Ide, Britt Ide, Pat Baffaro, and Steve Guadagno deserve special appreciation.

Finally, the twelve public school principals who generously participated in my pilot study and this main study. Truly, this would have been impossible without them.

Dedication

I dedicate this study to my Mom and first teacher, Anne Leone Robertson. She had a beautiful gift for helping other people discover and bring out the very best in themselves. If we could instill her hopeful, kind, and supportive spirit into the evaluation process for our school leaders, what a wonderful world it would be.

Table of Contents

Authorization to Submit Dissertation	ii
Abstract	iii
Acknowledgements	v
Dedication	vi
Table of Contents	vii
List of Figures	xii
List of Tables	xiii
Chapter One	1
Introduction	1
Background of the Study	3
Problem Statement	6
Statement of Purpose	7
Research Questions	7
Significance of the Study	7
Definition of Terms	8
Conceptual Framework	9
Limitations and Delimitations	10
Researcher Perspective	10
Organization of the Study	11
Summary	12
Chapter Two: Review of the Literature	14
Introduction	14

Perceived problems with traditional principal evaluation	14
The promise of new models of principal evaluation	18
Best practice recommendations	19
Criticisms in the literature	22
Effect of evaluation on principal leadership behaviors	26
Summary	29
Chapter Three: Research Methods	32
Introduction	32
Rationale for qualitative research design	33
Rationale for basic qualitative research design	34
Participant selection and recruitment	36
Selection	36
Recruitment	38
Summary of information needed	40
Overview of research design	41
Data collection methods	42
Data analysis	44
Simultaneous data collation and analysis	45
Coding	46
Second cycle coding	48
Matrix displays	48
Jotting	50
Analytic memoing	51

Assertions and propositions	51
Validity	53
Reliability	55
Ethics	57
Summary	58
Chapter Four: Participants' Experience with IDAPA 121	60
Introduction	60
Setting	61
Major Findings	61
Participants experienced evaluation under IDAPA 121 as a perfunctory, isolated end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice	62
The student growth component of IDAPA 121, and the prospect of evaluated partly on student growth had no impact on the participants' leadership focus or practices	73
Participants believe that policy makers and school policy related to their performance evaluation are out of touch with what happens in schools	82
Participants had clear ideas for improving the system so principals would experience more meaningful reflective practice and professional development	85
Summary	102
Chapter Five: Interpretation of Participants' Experience with IDAPA 121	104
Introduction	104

Analytic Theme One: The relationship between the ‘professional practice standards’ of IDAPA 121 and reflective practice, professional development, and leadership practice	106
Analytic Theme Two: The relationship between IDAPA 121 and participants’ leadership focus and practice	122
Analytic Theme Three: The relationship between the prospect of being measured on student growth and participants’ leadership focus and practices	124
Analytic Theme Four: Meaningful professional development	129
Summary	133
Chapter Six: Conclusions and Recommendations	135
Introduction	135
Non-impact of IDAPA 121.....	135
The student growth component of IDAPA 121	136
Non-evaluative development alternative	137
Recommendations for policy changes	138
Recommendations for further research	139
Summary	140
References	141
Appendix A: Summary of IDAPA 121	152
Appendix B: Interview Protocol.....	155
Appendix C: Informed Consent.....	159
Appendix D: Second Cycle Codes	162
Appendix E: Analytic Memo (Composite Portrait)	164

Appendix F: Summary of Findings Submitted to Participants..... 169

Appendix G: Member Check Results..... 172

List of Figures

Figure 1: Conceptual Framework	9
Figure 2: Timing of Events in Evaluation Process	112
Figure 3: Professional Development Framework	131

List of Tables

Table 1: Participant Characteristics.....	41
Table 2: Matrix Display Example	49
Table 1: Participant Overview	61

Chapter One

There's no such thing as a high-performing school without a great principal. It is impossible. You simply can't overstate their importance in driving student achievement, in attracting and retaining great talent to the school.

- Arne Duncan, U.S. Secretary of Education, November 2010

Introduction

Performance evaluation of school principals is not new in the American public school system (Wallace Foundation, 2009; Radinger, 2014). Yet until recently, principal evaluation was not a high priority for policy makers or school officials (Lashway, 2003; Darling-Hammond, LaPointe, Meyerson, Orr, & Cohen, 2007). Consequently, principal evaluations were largely administrative exercises that did not align to national or local policy priorities or professional standards (Goldring, Porter, Murphy, Elliot, & Cravens, 2007). Moreover, the quality of those evaluations was often quite poor (Goldring, et al., 2007; Lashway, 2003). Not surprisingly, evaluations were reported to have little impact on leadership practice or development (Davis, Kearney, Sanders, Thomas, & Leon, 2011; Wallace Foundation, 2009).

As the accountability movement gained momentum in the early to mid- 2000s, policy makers began to view principal evaluations as a strategic lever that could improve teacher performance and student achievement across the nation (Lashway, 2003; Portin, Feldman, & Knapp, 2006). This renewed interest in principal evaluations culminated in federal school reform efforts of U.S. President Barack Obama's administration that drove significant changes in principal evaluation programs across the nation. These included the Race to the Top program and a program that granted waivers from certain provisions of No Child Left Behind.

Race to the Top (RTT) was the Obama administrations \$4.35 billion competitive

grant program that launched in 2009 with a goal to “encourage and reward States that are creating the conditions for education innovation and reform” (U.S. Department of Education, 2009). This program awarded points to states for adopting a number of school reform initiatives, including the implementation of rigorous principal evaluation programs that were tied to standards and that included student growth measures as a “significant factor” in the principal’s evaluation (U.S. Department of Education, 2009). Forty states, including Idaho, applied for Phase 1 of RTT that ran from January to March 2010 (McNeil, 2010).

In 2011, the Obama administration announced its ESEA Waiver program that provided additional incentives for States to revise their principal evaluations. The ESEA Waiver program offered flexibility from certain provisions of the federal No Child Left Behind law (NCLB now commonly referred to as ESEA) for any state that agreed to comply with four specific policies (U.S. Department of Education, 2012). One of these policies centered squarely on the evaluation of teachers and principals, and required states to adopt and implement “principal evaluation and support systems that:

- (1) will be used for continual improvement of instruction;
- (2) meaningfully differentiate performance using at least three performance levels;
- (3) use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students ... and other measures of professional practice ...;
- (4) evaluate teachers and principals on a regular basis;
- (5) provide clear, timely, and useful feedback, including feedback that identifies needs and guides professional development; and
- (6) will be used to inform personnel decisions. (U.S. Department of Education, 2012,

p.7).

As of November 1, 2014, forty-three States, including Idaho, the District of Columbia and Puerto Rico received ESEA Waivers; two additional States have waiver requests pending with the US Department of Education. Within a span of less than five years, principal evaluations that include student growth data “as a significant factor” and “other measures of professional practice” became the de-facto standard for school districts in Idaho and across the United States (Clifford, Hansen and Wraight, 2014; McNeil, 2010; Guilfoyle, 2013).

Background of the Study

In March 2011, Idaho lawmakers enacted three education laws that were collectively known as ‘Students Come First’. Consistent with RTT and requirements of the ESEA Waiver program, these laws required at least 50% of school principal evaluations to be based on student achievement, and the balance of the evaluation to be based on parental input and observation (Idaho State Department of Education, 2012). The Students Come First laws did not mandate any specific practice standards for principal evaluations. In December of 2011, the Idaho State Department of Education convened the Administrator Evaluation Focus Group to begin the work of “crafting a Statewide Framework for Administrator Performance” that would be heavily focused on instructional leadership (Idaho State Department of Education, 2012, p. 150). The Idaho State Board of Education planned to solicit public comment on the recommendations of this Focus Group in the fall of 2012.

In February 2012, the Idaho State Department of Education submitted its ESEA Waiver application, touting the Students Come First laws and their requirement that 50% of the principal evaluation be based on measures of student growth. The U.S. Department of Education conditionally approved Idaho’s ESEA Waiver application in October 2012. Less

than one month later, Idaho voters repealed the Students Come First laws, presumably due to other controversial parts of the legislation that curtailed collective bargaining, introduced merit pay measures, and required certain technological investments in all schools. With the repeal of Students Come First, Idaho's eligibility for the ESEA Waiver was threatened, and Idaho's then Superintendent of Public Instruction, Tom Luna, asked the US Department of Education to suspend its review of Idaho's waiver application to allow Idaho time to regroup (US Department of Education, 2013).

Following repeal of the Students Come First laws, Idaho convened an Educator Evaluation Task Force that was charged with analyzing ESEA Waiver requirements and recommending changes to principal evaluation that would enable Idaho to qualify for the ESEA Waiver. Their recommendations were approved by the Idaho State Board of Education in April 2013 and implemented, with immediate effect, through changes to Idaho's administrative code (Idaho State Department of Education, 2013).

Idaho's new principal evaluation requirements are set forth in Idaho Administrative Code section 08.02.02.121 (hereinafter, IDAPA 121). Under these requirements, all Idaho school districts and charter schools must evaluate their principals on: (1) a 'professional practice' component that aligns with the Idaho Standards for Effective Principals (ISEP) and (2) a 'student achievement' measure. The professional practice standards cover three domains (school climate, collaborative leadership and instructional leadership) and must comprise 67% of the total evaluation; the student achievement component must account for 33% of the evaluation. Appendix A provides a complete overview of the substantive and procedural requirements of IDAPA 121. The 2014-2015 school year marked the second year that IDAPA 121 was implemented across Idaho.

In its amended ESEA Waiver Application from June 2013, Idaho's State Department of Education explained two primary and related goals in adopting IDAPA 121. The first goal was to "ensure that each Local Educational Agency develops and adopts an evaluation and support systems (sic) that will improve student achievement and the quality of instruction for all students in the classroom." The second goal is to "promote reflective practice and the development of ongoing, personalized development plans leading to improved support for turning around low-performing schools and measurably increasing student achievement for all students" (Idaho State Department of Education, 2013, p. 200).

These goals of improved instruction, increased student achievement, and professional development for principals are consistent with the espoused purposes and promises of principal evaluation in the research and policy literature (Portin, et al., 2006; Catano & Stronge, 2007; Wallace Foundation, 2009). Halverson, Kelley, and Kimball, (2004) declared that performance evaluations of principals hold great promise for improving the quality of principal leadership, classroom instruction and learning outcomes for our students. A decade later, Clifford, et al. (2014) struck a similar note when it observed, "principal evaluation has long held promise for improving principal effectiveness, fostering learning and reflection, and increasing accountability for job performance" (p. 5).

On December 10, 2015, President Obama signed into law the Every Student Succeeds Act (ESSA) which reauthorized ESEA and will curtail many of the most controversial elements of NCLB (U.S. Department of Education, 2016). These included aspects of NCLB which prompted Idaho and other states to apply for the ESEA Waiver program in the first place. As a result, the state of Idaho must continue to comply with its ESEA Waiver requirements until August 2016, at which time ESSA will formally replace NCLB. The

2016-2017 school year has been described as a “transition year” from NCLB and the ESEA Waivers program to ESEA under ESSA (Corbin, 2015).

Problem Statement

Despite the strong consensus regarding the importance of principal leadership (Edmunds, 1979; Robinson, Lloyd, & Rowe, 2008; Wallace Foundation, 2013), and “the promise” of more effective principal evaluation models that improve leadership, teaching and learning, these promises are largely unexplored in the research literature (Clifford et al., 2014; Superville, 2014). As Radinger (2014) observed, “more research is needed on both formal appraisal approaches and actual appraisal practices, as well as on the effects of appraisal on school leaders’ practices and behaviours” (p. 379). Other prominent commentators in the field of principal evaluation reform agree (Davis et al., 2011; Clifford & Ross, 2011; Clifford, Behrstock-Sherratt, & Fetters, 2012).

These gaps in the research literature leave open the question of what impact, if any, principal evaluation programs such as IDAPA 121 will have relative to the stated goals of increased reflective practice, and professional development for principals, improved instruction, and increased student achievement, (Davis, et al., 2011). Given the significant sums of federal and local dollars, time and attention being invested across the country and here within Idaho on the promise of principal evaluation reform, these research gaps loom large. Although it is too early in the implementation of IDAPA 121 to examine its long-term impacts on instructional quality or student achievement, this is a perfect time to “examine how the hopes of policy design meet the realities of actual practice” concerning IDAPA 121’s stated goal of “promot[ing] reflective practice and development of ongoing professional development plans” for Idaho principals (Idaho State Department of Education,

2013, p. 200). With the recent enactment of ESSA, this is especially true now that Idaho may have the option to revisit some or all aspects of IDAPA 121 during the 2016-2017 school year and beyond.

Statement of Purpose

The purpose of this study was to describe how well the “promise” of IDAPA 121 compared with its actual impact on participants’ reflective practice, professional development, and leadership practices.

Research Questions

Specific research questions that guided this inquiry were:

1. How do principals in Idaho describe the effect of the ‘professional practice standards’ of IDAPA 121 on their reflective practice and professional development?
2. How do principals in Idaho describe the effect of the ‘student growth component’ of IDAPA 121 on their reflective practice and professional development?
3. How does the experience of being evaluated under IDAPA 121 effect the leadership behaviors or practices of principals in Idaho?
4. How does the experience of being evaluated under IDAPA 121 compare with the experience of being evaluated before IDAPA 121?
5. What circumstances or experiences contribute most significantly to reflective practice and professional development for principals in Idaho?

Significance of the study

Policy makers have invested significantly in the promise of reformed principal evaluations such as IDAPA 121. These investments are being made even though little research has investigated their actual impact. With the investments being made in Idaho and

across the United States to reform principal evaluation, this study can increase understanding of the value of these investments, and shed further light on how closely the promise of this policy change matches the actual practice and experience of principals it was intended to support.

Definition of Terms

Evaluation: As it relates to public school leaders, is a process that school districts use to assess and review the job performance of the principal. Evaluations can take many forms and serve multiple purposes; sometimes also referred to as ‘assessment’ or ‘performance evaluation’ in the literature.

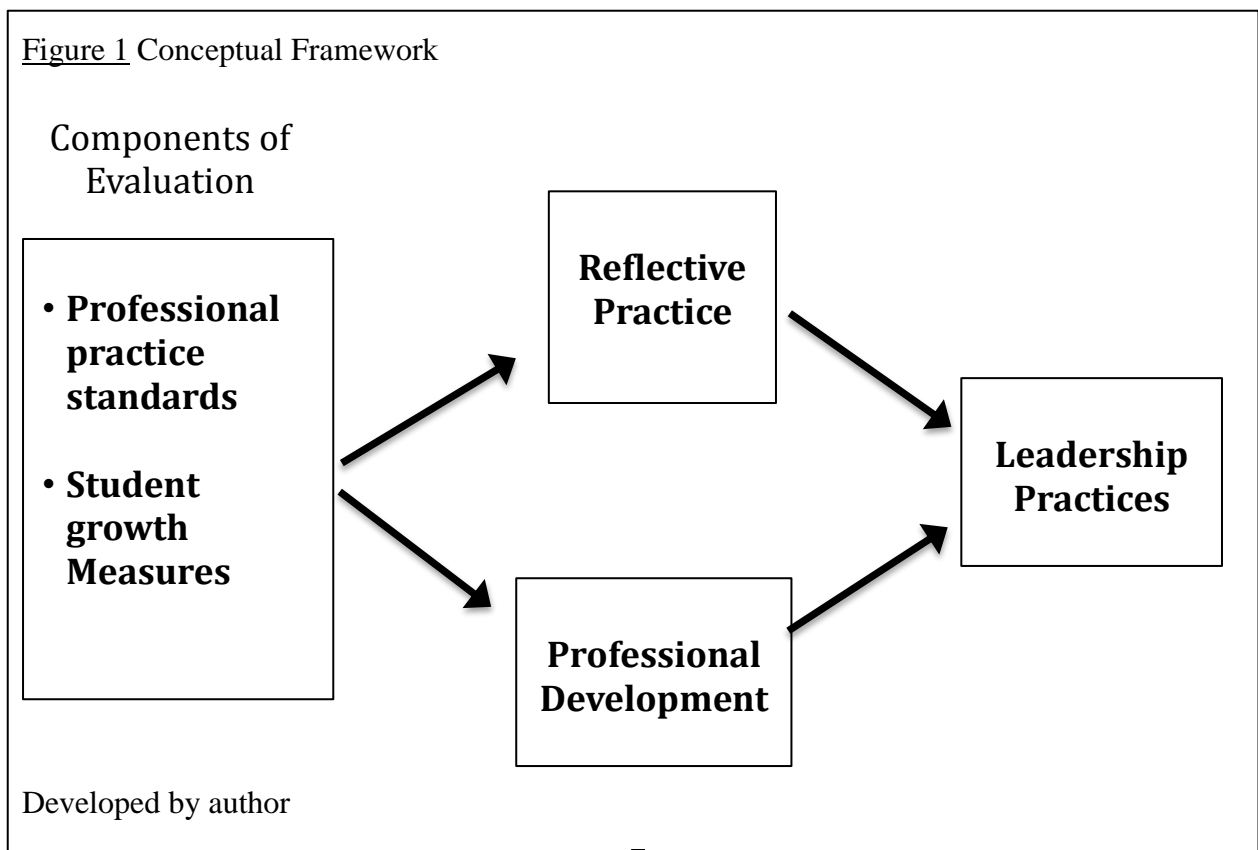
IDAPA 121: Is an Idaho administrative rule created in 2013 and set forth in Idaho Administrative Code section 08.02.02.121. It requires all school districts and charter schools in Idaho to include two primary elements within their principal evaluation programs, beginning in the 2014-2015 school year: (1) a ‘professional practice’ component that is organized into three Domains (school climate, collaborative leadership and instructional leadership) and (2) a ‘student achievement’ measure. See Appendix A for additional details.

Professional development: Is the process of obtaining the skills, qualifications, and experience that allow you to make progress in your career (Macmillan, 2014).

Reflective practice: Although there is no universally accepted definition of reflective practice (York-Barr, Sommers, Ghere and Montie, 2006), most definitions accord with the following description by Bright (1996): “a genuinely critical, questioning orientation and a deep commitment to the discovery and analysis of positive and negative information concerning the quality and status of a professional’s designed action” (p. 165).

Conceptual framework

Miles, Huberman, and Saldaña (2014) describe the conceptual framework as an explanation of the main things to be studied and the presumed relationships among them. Although it can be presented both graphically and in narrative form, Miles, Huberman and Saldaña (2014) assert the conceptual framework is “best done graphically, rather than in text” (p. 34). Following their advice, Figure 1 presents the conceptual framework for this proposed study. Conceptually speaking, this study was interested in two primary loci of inquiry. The first was the relationship between the two components of IDAPA 121 and reflective practice and professional development by principals. The second was the relationship between (a) reflective practice and professional development and (b) leadership practices.



Limitations and Delimitations

One limitation of this study arises from the fact, discussed more fully in Chapter Four, that study participants were drawn from two specific Idaho school districts. As a result, their experiences may reflect district specific implementations of IDAPA 121 that differ from other Idaho districts. A second limitation is that both districts are large and primarily suburban, and the experience of principals from smaller and/or urban districts may be different. A third limitation relates to the fact that all participants were elementary or middle school teachers, and the experiences of high school principals might be different from those of the study's participants. A fourth limitation arises from the data collection method of this study that focuses primarily on participant interviews. Although one hopes participants will be candid and truthful in their reflections and responses, there is no independent way to validate all of their information, especially as it relates to their personal perspectives and opinions.

In terms of delimitations, this study did not explore the impact of IDAPA 121 on teacher instruction, school climate or student achievement. Similarly, it did not investigate IDAPA 121 from the perspective of district supervisors who evaluated the performance of school principals or Idaho's Department of Education who helped implement IDAPA 121 in Idaho. Also, this study did not explore, in depth, the phenomenon of reflective practice as understood and practiced by participants.

Researcher perspective

I came to this study with a rich and varied background of experience and opinions about school leadership, performance evaluations in general, and the specific details of IDAPA 121. While I have held a variety of positions in my career, including teacher, school

administrator, private practice attorney, corporate attorney, manager, and leader of corporate training and development programs, I have experienced few meaningful performance evaluations. As a performance evaluator, I hope that I have provided meaningful evaluations to others, yet acknowledge that doing so can be difficult for reasons that are both structural and practical. Thus, I came to this topic with some biases, including the expectation that many evaluations are poorly designed and implemented, and a sense that conducting evaluations well is not easy. I also came to this project with a strong bias against ‘one size fits all’ solutions to challenges, especially those as complex as evaluating and improving modern school leadership. Yet, I understood that the quality of this study depended on my ability to remain neutral, open and curious with participants, and aware of any personal assumptions, biases or conflicts that may influence or interfere with the voice of the participants. The process of naming my potential biases during the proposal stage of this study increased my self-awareness of these issues and helped me approach the interviews more openly. I returned to these issues throughout the interviews and data analysis, to ensure I was honoring the perspectives and voices of the participants and representing them faithfully in this study.

Organization of the Study

This study is organized into six chapters. Chapter One provides an introduction and background to the subject phenomena, followed by the research problem, statement of purpose and research questions, significance of study, definition of terms, conceptual framework, limitations and delimitations, researcher perspective and assumptions, and organization of the study. Chapter Two reviews the literature with a focus on five major themes and trends that helped to shape this study: (a) problems with traditional principal

evaluations, (b) the promise of new models of evaluation, (c) best practices to implement these new models, (d) criticisms of the new models, and (e) what a surprisingly limited body of research tells us about the effect of new principal evaluation models on reflective practice, professional development and leadership practices. Chapter Three details the research methods for this study and explains why a basic qualitative design was chosen. It defines the participant selection strategy, explains how data was collected and analyzed, and discusses how issues of validity, reliability and ethics were addressed. Chapter Four describes participants' experience with IDAPA 121 and is presented through four major findings. Chapter Five provides analysis and interpretation based on the major findings. Chapter Six offers recommendations for policy changes and additional research.

Summary

During the 2014-2015 school year, Idaho's public school principals began to experience a new, consistent model of evaluation that measures their performance relative to (1) specific leadership standards modeled on ISLLC-2008 and (2) student growth measures. Idaho is not alone in this change. This new evaluation framework was driven, in large part, by national trends and specific federal policy changes of the last five years that have given rise to similar reforms across nearly every state in the union. Despite the clear consensus about the importance of principal leadership and the value of strong, effective evaluation programs for principals, little research on the process of principal evaluation or its impact on school leadership and school improvement has been conducted. Given the contemporaneous implementation of Idaho's revised principal evaluation model and the significant gaps in the research literature regarding principal evaluation, this is an ideal time for a qualitative study that describes how well the promise of IDAPA 121 compares with the actual experiences of

Idaho principals. This is especially true with the advent of ESSA, and the choices that Idaho may have to continue with IDAPA 121 or to pursue other approaches to principal evaluation and professional development.

Chapter Two

Review of the Literature

Introduction

The purpose of this study was to describe how well the “promise” of IDAPA 121 compared with its actual impact on participants’ reflective practice, professional development, and leadership practices. In order to frame this study, an extensive literature review was conducted that focused on principal leadership, including instructional and collaborative leadership models, and principal evaluation policy and practice from before and after enactment of RTT and the ESEA Waiver program. Several key themes and trends emerged from this review to shape this study’s specific research questions and methodology. Five themes are detailed in this chapter. They are: (a) perceived problems with traditional principal evaluation before RTT and the ESEA Waiver program; (b) promise of a reformed principal evaluation system to improve school leadership, teaching and student learning; (c) best practice recommendations to provide a meaningful and impactful evaluation experience for principals; (d) criticisms of the new model of principal evaluations; and (e) what the literature does and does not tell us about principal evaluation and its effect on reflective practice, professional development and leadership practices.

Perceived problems with traditional principal evaluation

Although principal evaluations have received surprisingly little attention in the research literature of the last several decades, researchers consistently reported significant problems with both the substance and process of traditional models of principal evaluation. Shelton (2013) summarized what we know about traditional principal evaluations:

A comprehensive review of principal evaluations has found that current assessments typically are out of synch with what research has identified as the most important indicators of effective school leadership . . . Many state and district evaluations are not aligned to performance standards, valid and reliable methods for evaluations are few and far between, and little emphasis is given to evaluator training. In addition, few rigorous principal assessments exist that are intended for use in hiring, advancement and tenure decisions (p. 4).

This studies in this section of the literature review describe problems associated with traditional principal evaluation programs across the U.S. and span the mid-1980s through 2013.

Harrison and Peterson (1986) studied the perspectives of principals and superintendents concerning evaluation programs in one southern state, with a focus on criteria, purpose, sources of information, and results. They reported significant difference between the perception of superintendents and principals in five key areas. First, they found that superintendents had more favorable perceptions of the evaluation processes than the principals who were being evaluated. Second, principals reported being much less clear about the process of evaluation than superintendents. Third, superintendents and principals reported widely different understandings of the purposes and priorities of the evaluation program. Whereas superintendents believed that instructional leadership practices should be the primary focus of evaluations, principals believed that operational management functions were most important. Fourth, principals were more likely than superintendents to believe that their evaluations were based on community opinions, as opposed to measurable

performance standards. Fifth, principals were more likely than superintendents to report weak communication concerning the purpose, process or outcome of their evaluations.

Davis and Hensley (1999) reported similar findings in their report on the perspectives of fourteen principals from school districts in northern California. These principals reported their formal evaluations were not helpful in “shaping or directing their professional development or in promoting school effectiveness” (p. 399). The principals described a lack of trust in their district office evaluators, and believed their evaluations were based on subjective opinions rather than objective performance criteria, and may have been influenced by a small number of disgruntled parents or teachers. Additionally, the principals complained about inconsistencies in evaluation procedures, in the qualifications and knowledge of the people conducting the evaluations, and in the sources of information that informed the evaluations themselves. Principals believed that evaluators spent little time observing their leadership practices, and that evaluations rarely included input from teachers, parents or students.

Stine (2001) examined principal evaluation policies and criteria from 17 school districts in Southern California and discovered significant inconsistencies across the districts. For example, some districts emphasized personal characteristics while others focused on leadership style, management skills or content expertise. Stine also found that few of the evaluation programs were aligned with professional growth and development plans, and most relied on a checklist approach.

Reeves (2005) surveyed principals across the U.S. and found that few principals believed that evaluations shaped their motivation or performance. They reported that criteria

for evaluation were not well established, few consequences were attached to the outcomes of their evaluation, and they did not receive helpful feedback.

Goldring, Cravens, Murphy, Porter, Elliot, and Carson (2009) analyzed principal evaluation documents from 68 urban school districts across 43 states and found significant problems with the content of those evaluation instruments. Among the problems reported, the evaluations did not emphasize the leadership behaviors necessary to improve instruction, and less than fifty percent of the evaluation models were aligned with professional standards. Additionally, few of the districts (2 out of 68) examined the validity or reliability of their evaluation protocols, a finding repeated by other researchers (Clifford and Ross, 2011).

In Idaho, Bingham's (2013) doctoral research study pre-dated implementation of IDAPA 121. Consistent with the findings of researchers from other states (Harrison and Peterson, 1986) Bingham reported that superintendent evaluators had a significantly more positive opinion of the value of principal evaluation process than did the principals who were being evaluated.

In response to these and similar findings, researchers and policy makers began to express the need for new models of principal evaluation. For example, Goldring, et al. (2007) noted a 'general agreement' among researchers that the current state of school leader assessment was lacking. Davis, et al. (2011) conducted a thorough review of the literature on principal evaluations for WestEd and concluded by expressing hope that implementation of new evaluation models would allow researchers to better understand what models of evaluation work best, and how they impact individual and organizational development. Clifford and Ross (2013) summarized the general consensus as follows, "it is time to rethink principal and assistant principal evaluation as a process to build individual leadership

capacity and school effectiveness” (p. 8). A host of other commenters echo this sentiment (Jacques, Clifford & Hornung, 2012; Clifford, et al., 2012; Shelton, 2013; Clifford et al., 2014).

The promise of new models of principal evaluation

As Halverson, et al. explained, performance evaluations of school principals hold great promise for improving the quality of school leadership, classroom instruction and learning outcomes for our students. A decade later, Clifford, et al. struck a similar note when they observed, “principal evaluation has long held promise for improving principal effectiveness, fostering learning and reflection, and increasing accountability for job performance” (p. 5). The policy logic behind the ‘promise’ of reformed principal evaluation programs is clear: (a) the role of the principal had changed significantly in recent years from a focus on being a ‘building manager’ to being a ‘learning leader’; (b) research based standards of practice– such as ISLLC – describe what effective instructional leadership looks like; (c) traditional principal evaluations did not focus on the right things and failed to provide meaningful feedback to principals; (d) an improved principal evaluation that focuses on the right things (e.g. instructional leadership) can provide principals with actionable insights to drive improvement in their leadership practices; and (e) improved leadership will improve teaching and school effectiveness, which will, in turn, improve student learning outcomes (Goldring, et al., 2007; Wallace Foundation, 2009). Wallace Foundation (2009) summarized this policy logic by observing:

If assessment can become an important means of illuminating and changing the performance of school leaders, especially in our lowest performing schools, we may

finally begin to make a serious dent in the unacceptable achievement gap that confronts this country. (p. 13)

This ‘promise’ of the new wave of principal evaluations has been reinforced in recent years by the articulation of benefits associated with fixing evaluations. First, improved evaluations can clearly communicate and align ‘learning centered’ leadership expectations so that all principals understand what is expected of them (Catano & Stronge, 2007; Wallace Foundation, 2009). Second, a standards-based assessment that aligns to important policy objectives (such as measurable student growth) can be used to hold principals accountable for meeting these objectives and help districts to make informed personnel decisions (Wallace Foundation, 2009). Third, improved evaluations can help support the ongoing professional development of principals by providing valuable feedback, identifying areas of strength and opportunity, and highlighting successes that build the leader’s confidence (Portin, et al., 2006; Wallace Foundation, 2009). This third benefit aligns with Idaho’s stated goal of “promot[ing] reflective practice and development of ongoing professional development plans” for Idaho principals (Idaho State Department of Education, 2013, p. 200).

Best practice recommendations

As interest in reforming principal evaluations increased with RTT and the ESEA Waiver program, commentators offered recommendations for how States and local districts should redesign their evaluation programs (Wallace Foundation, 2009; Clifford & Ross, 2011; Davis et al., 2011; Shelton, 2013). Although some variation could be found among the different lists of best practice, most commentators agreed that (a) evaluations should focus squarely on the most important ‘driver’ behaviors that improve instruction and are anchored

in accepted leader standards; (b) evaluations should use valid and reliable instruments and processes; (c) evaluation should be a cyclical process rather than a one-time event; (d) evaluation programs should be flexible enough to accommodate differences in local school context, job responsibilities and years of experience; (e) principals should be actively engaged in the process of establishing performance goals and objectives and in assessing their own performance; (f) balance should be struck between the formative and summative aspects of the evaluation, and results should inform meaningful professional development planning; (g) the evaluation should include multiple measures of performance, and feedback from a variety of stakeholders; (h) evaluations should include multiple rating categories to clearly differentiate performance; (i) training must be provided to those who conduct the evaluations to ensure a fair and consistent process for all principals (Wallace Foundation, 2009; Browns-Sims, 2010; Davis et al., 2011; Clifford & Ross, 2011).

Although commentators agree that evaluations should provide principals with useful, trustworthy and actionable feedback (Clifford & Ross, 2011), some researchers further specified that leadership development requires ‘frequent feedback’ (Lashway, 2003; Portin et al., 2006; Davis et al., 2011). This is consistent with the concept, popularized in recent years within corporate circles, of ‘deliberate practice’ (Colvin, 2008). Deliberate practice is an approach to skill building and professional development that depends on repetition of practice and constant feedback. This concept of ‘frequent feedback’ is noteworthy within the context of principal development because most principal evaluation programs – including IDAPA 121 – require significantly less frequent feedback. The requirements under IDAPA 121, for example, specify that evaluations be conducted once annually, no later than May.

Several authors encouraged states and districts to expand evaluations to include student outcomes and teacher outcome data (Kearney et al., 2011, p. 28). New Leaders for New Schools (2010) for example argued that “student achievement and teacher effectiveness should be at the center” of the evaluation process” (p.6) by weighing these factors heavily at 70% of the total evaluation score, with the remaining 30% focused on principal behaviors that have been shown to increase student achievement and teacher effectiveness. They argued this 70% should be further divided as follows: 50% to student outcomes such as growth and proficiency, and the remaining 20% allocated to outcome measures related to the principals impact on ‘teacher effectiveness’ as measured by student performance and the retention of effective teachers.

One final observation from the literature merits discussion, and that is the finding from Kimball, Milanowski, and McKinney (2009) and others that the ‘process’ or overall ‘experience’ of the evaluation may be as important – if not more important – than the actual content of the evaluation instruments themselves. This conclusion arose in the context of studying the implementation of a new, standards based evaluation program that was similar to the situation now presented with IDAPA 121. In Kimball et al. (2009), the researchers found that inconsistent and incomplete implementation of the new standards-based evaluation program compromised the degree to which principals found the new model helpful, and made it difficult to measure the effectiveness of this new approach in terms of leadership development. This led the researchers to conclude that ‘implementation’ trumps ‘instrumentation’, a conclusion shared by other researchers (Davis et al., 2011). This is consistent with Sun, Youngs, Yang, Chu and Zhao (2011) that found principals were more

likely to focus on learning centered leadership when the district emphasized the ‘purpose’ of the evaluation.

Criticisms in the literature

This section describes three primary criticisms of principal evaluation models such as IDAPA 121. The first two criticisms relate to the professional practice component and the third criticism relates to the use of student achievement data. Implications of these criticisms are noted where appropriate.

The first criticism declares that it is wrong to impose an inflexible, one size fits all model of leadership on school leaders because the specific leadership needs of any given school will vary from one school to the next, and may even vary within a single school over time. This criticism applies equally to the three Domains of the professional practice component of IDAPA 121, and has its roots in the beginning of the instructional leadership movement. Edmunds (1979), who is considered one of the leaders of that movement, prefaced his description of the effective school by first expressing his desire to “make clear at the outset that no one model explains school effectiveness for the poor or any other social class subset” (Edmunds, 1979, p. 22). Three years later, Bossert, Dwyer, Rowan, and Lee (1982) expressed a similar view:

No single style of management seems appropriate for all schools ... principals must find the style and structure most suited to their own local situation ... a careful examination of quantitative studies of effective schools ... suggests that certain principal behaviours have different effects in different organizational settings. (p. 38)

These observations are consistent with the descriptions of Jackson (2000) and Hallinger (2003) of school improvement as being akin to a journey. As Hallinger (2003)

observes, “the type of leadership that is suitable to a certain stage of the journey may well become a limiting or even counter-productive force as the school develops” (p. 345-346). This ‘contingent’ nature of leadership is further supported in the corporate leadership literature (Goleman, 2000).

Finally, Leithwood, Louis, Andersen, and Wahlstrom (2004) acknowledged this contextual focus of effective leadership as follows: “Impressive evidence suggests that individual leaders actually behave quite differently (and productively) depending on the circumstances they are facing and the people with whom they are working. This calls into question the common belief in habitual leadership “styles” and the search for a single best model or style” (Leithwood, et al., 2004, p. 10). Ironically, this same summary of the research (Leithwood et al., 2004) was cited by the authors of the 2008 revision to ISLLC (Council of Chief State School Officers, 2008), which in turn directly influenced the drafters of the Idaho Standards for Effective School Principals. Thus, we find that the research which argued against the search for one ‘single best model or style’ of school leadership would later be used to justify the imposition of one single best model of leadership for all principals in Idaho and many other states across the country.

A related criticism of best practice leadership models like IDAPA 121 asserts that asking school leaders to successfully master and enact the many elements of Domain 1, 2 and 3 of IDAPA 121 is highly problematic, unrealistic and neo-heroic in its expectations (Leithwood, Patten, & Jantzi, 2010). Robinson, et al. (2008) echoed this theme when they explained that “our findings should not be interpreted, therefore, as implying that any school leader should demonstrate high levels of capability on *all five* dimensions ... the more defensible implication of our finding is that what matters is the frequency of various

instructional leadership practices rather than the extent to which they are performed by a particular leadership role” (Robinson et al., 2008, p. 668, emphasis in original). Douglas Reeves of the Leadership and Learning Center raised a similar alarm by noting:

The acid test of the new wave of principal evaluations will be whether they use real time data to narrow the focus. If every conceivable idea that might make for better leadership is evaluated, then the folly of new teacher evaluation will be replicated, prodding an evaluation process that will become a paperwork drill rather than a meaningful evaluation. (Clifford & Ross, 2012 p. 9)

A similar criticism has been made of best practice leadership ‘lists’ within the corporate context (Buckingham & Coffman, 1999).

The third criticism, and perhaps the most controversial aspect of the new principal evaluation models such as IDAPA 121, relates to the use of value added measures (VAM) derived from student test scores and other assessments of student performance to measure principal quality and effectiveness. Although the use of student test performance data to evaluate educational quality is nothing new in the era of accountability ushered in by NCLB, its use to directly evaluate principal leadership is novel and not well supported by the research. As reported by Superville (2014), “the growth in principal evaluation policies (that feature student growth measures) has not been matched with corresponding study of their implementation, reliability, and effectiveness” (p. 1). To the extent that research has focused on the use of value added measures to evaluate educational quality, most of those studies “have been geared toward similar systems for teachers (Superville, 2014, p. 1) which have in turn been criticized.

Prominent researchers recently criticized the use of VAM for teacher evaluations

(Hallinger, Heck, & Murphy, 2013). They note a number of problems associated with current models of teacher evaluation, including questions about whether current statistical models can accurately and reliably isolate the impact of individual teachers on student learning outcomes (Hallinger et al., 2013). This criticism is noteworthy given that the use of VAM in principal evaluation programs is based, in large part on these same, potentially discredited, teacher evaluation models (Superville, 2014, p. 1). As Brian Gill of Mathematica Policy Research observed, “I think most of the policy makers overlooked the fact that analytically [principal evaluation is] a different problem than analyzing a teacher’s performance” (Zubrzycki, 2013). Perhaps it is no surprise therefore to find strong criticisms of the use of VAM in the principal evaluation context (Fuller and Hollingworth, 2013; Lipsomb, Chiang, & Gill, 2012). Fuller and Hollingworth (2013) conclude:

There are currently no strategies to estimate principal effectiveness that accurately capture the independent effects of principals on student test scores; thus, these current strategies send inaccurate signals to both principals and those who make employment decisions about principals ... indeed such statistical estimates should [not be used] for making judgments or decisions about principals. (p. 466)

Even assuming reliable models of VAM could be used for principal evaluation, Clifford et al. (2012) noted that evaluations based on student growth may provide limited actionable feedback for the principal because student growth – or lack thereof – does not, by itself, reveal what the principal did or did not do well because principals exert only an “indirect effect” on student achievement.

The effect of evaluation on principal leadership behaviors

Research on principal evaluation systems and policies is surprisingly sparse (Davis et al., 2011) and little research has been conducted on *the effects of* performance evaluation systems on principals, their behaviors or the schools they lead (Clifford & Ross, 2011; Clifford et al., 2014; Radinger, 2014). Thus, this section of the literature review describes the few studies that have investigated the effect of standards based evaluation on principal leadership behaviors. Although these studies help shed some light on the issue, gaps in the literature remain, giving rise to the need for additional research (Davis, et al., 2011).

Kimball, Milanowski, and McKinney (2007) compared the experiences of principals who received the traditional, 'ratings' based evaluation (where three levels of performance were noted: Highly Proficient, Proficient, or needing Targeted Growth) with principals who experienced the revised, standards-based evaluation and a four level descriptive rubric. They were interested to compare the perceptions of these two different groups in terms of performance expectations, quality of feedback, emphasis on job tasks and professional growth. Mixed results were reported. Regarding perception of performance expectations, the researchers reported no statistically significant difference between the two groups. On the issue of quality of feedback, statistically significant differences were found; principals who experienced the new evaluation reported receiving more specific and useful feedback than principals who experienced the traditional evaluation. Focusing on emphasis on job tasks, the researchers expected that principals in the new system would report spending more time and effort in leadership tasks that were more heavily emphasized in the new system. Although positive results were found, none were statistically significant. Regarding impact on professional development, principals noted positive differences in the new system. These

included the production of more authentic dialogue with their evaluator and the use of the new performance rubric for self-evaluation. Nonetheless, none of these differences were statistically significant, an outcome that the researchers partially attributed to other development programs outside of the evaluation process within the district. More significantly, the researchers believed their results were impacted by a weak and inconsistent implementation of the new system. They note that inconsistency in implementation is not a new problem in principal evaluations, but would be especially problematic in the context of high stakes evaluations such as those used to inform promotion and retention.

Sun and Youngs (2009) applied quantitative analysis to explore the relationship between evaluation purposes, focus, assessed leadership activities, and principal behavior in thirteen Michigan districts. They reported several positive findings. First, principals were more likely to engage in learning centered behaviors such as setting high performance goals for students, coordinating the curriculum, and supporting instruction when school districts used evaluations processes to hold principals accountable for achievement of district goals, implementation of restructuring, when the focus of evaluation was on instructional leadership, and when they encouraged principal development. Second, they found evaluation processes were more likely to encourage learner-centered behavior when the evaluation focused on knowledge, skills and behaviors, rather than personality traits.

In a subsequent study, Sun et al. (2011) used survey data on 88 K-12 principals in Michigan and a comparison sample of 90 K-12 principals from Beijing, China. Again they used quantitative methods to estimate the extent to which features of the district evaluation were associated with principal leadership practices. They found that principals were more likely to engage in learning centered leadership behaviors when they perceived a higher level

of emphasis on evaluation purposes, content, sources of evidence and specific leadership practices. They further reported that frequency of evaluation did not explain variance in leadership behaviors, nor did receiving one type of punishment or reward. Although noting that causal inferences cannot be made due to limitations in their data, they conclude their findings “support the policy assumption that district principal evaluation can be effectively used to communicate school leadership responsibilities and serve as a powerful policy instrument to influence principals’ behavior” (p. 210). Recognizing that “principal evaluation alone may not lead to desired learning-centered leadership activities”, the authors recommended connecting evaluation with professional development that prepares principals to have meaningful interactions with teachers (p. 210).

Parylo, Zepeda, and Bengtson (2012) applied a phenomenological approach to understand principals’ lived experiences of being evaluated within four school systems in the state of Georgia. Although the authors note differences existed between each school system’s model, detailed descriptions of the evaluation protocols were not provided. As a result, it was not possible to determine if any of the four evaluation models in Parylo, et al. (2012) were otherwise to IDAPA 121. Nonetheless, several aspects of this study are worth noting. First, the principals reported that they were able to improve areas of concern when their evaluation was “formative in nature, involved opportunities to collaborate, have open dialogue, and engage in reflective practice” (p. 234). Second, the authors recommended that future research “examine the principals’ views on what other aspect of their work should be included in their assessment” (p. 234).

The four studies described above suggest principal evaluations offer some potential to impact leadership development in positive ways. At the same time, the evidence is far from

conclusive, and suggests effective leadership development may depend on an evaluation process that is formative in its focus, as opposed to summative, supportive of open dialogue and reflective practice, and implemented cleanly and consistently. As Derrington and Sanders (2011) observed, a “principal’s development of new leadership skills is dependent on the willingness to embrace change, learn new strategies and take risks” (p. 34). This is not easy work, and depends on an environment that embodies trust, openness and caring (Derrington & Sanders, 2011). These findings are consistent with Bickman, Goldring, De Andrade, Breda, and Goff (2012) which found that professional development of principals is significantly enhanced when principals receive a combination of feedback and coaching to help interpret and internalize the feedback and set specific development goals.

Summary

This chapter explored five primary trends and themes in the research literature related to new models of principal evaluation such as IDAPA 121. These included problems with traditional evaluations, the promise of reformed evaluations, best practices to ensure that new evaluation models fulfilled their promise, and criticisms of the one size fits all standards based evaluations and the use of VAM to measure principal performance. Finally, it reviewed the sparse body of research on the relationship between new evaluation models and their impact on principals’ leadership behavior. The following summarizes the salient points from this chapter.

Prior to the ESEA Waiver program and enactment of IDAPA 121, performance evaluations were perceived as ineffective in a number of areas, including: (a) principals reported that evaluations were not helpful in shaping their professional development or promoting school effectiveness; (b) evaluations were not aligned to professional development

plans; (c) evaluations were not aligned with relevant leadership standards; (d) evaluations were not informed by timely, relevant or trustworthy input; (e) evaluations were plagued by inconsistencies across districts; (f) few districts examined the validity or reliability of their evaluation protocols.

As the accountability movement gained momentum through the 2000's, commentators expressed hope that performance evaluation of school leaders could be used as a strategic lever to improve leadership and school performance despite their historical shortcomings. Evaluation proponents asserted an improved model of principal evaluations offered the promise to (a) more clearly communicate expectations for principals; (b) increase school leader accountability; and (c) provide helpful feedback to support principal professional development.

States and school districts prepared to implement new principal evaluation systems in response to RTT and the ESEA Waiver program. Commentators offered a number of recommendations for improving the effectiveness of principal evaluation. These included (a) enabling flexibility within the evaluation system to accommodate different school contexts and levels of experience of individual principals; (b) actively involving the principals in goal setting related to the evaluation; (c) ensuring that principals received useful, relevant and more frequent feedback; and (d) the observation that the quality of the evaluation process was as important as the content of the evaluation tool.

In the wake of the movement to revitalize principal evaluations as a lever to improve school performance, some commentators expressed concerns and criticisms with the new model of evaluations. These included concerns the new evaluation programs were adopting a 'one size fits all' approach to evaluation that was not supported by the literature, and

concerns with the unreasonably long list of standards that principals would be measured against. Critics also argued against measuring principals on the basis of student achievement because there was no reliable way to measure the impact of an individual school principal on the academic performance of his/her students.

As the above summary reveals, principal evaluations have been both criticized as ineffectual and lauded as strategic levers with the promise to improve professional development and principal performance. Against that backdrop, a small number of studies investigated the extent to which performance evaluation impacted school leaders. While these studies suggested that revamped evaluations could positively impact school leadership, their findings were sparse and inconsistent. Whereas Sun and Youngs (2009) reported that an accountability focus coupled with professional development appeared to influence principal behavior, Parylo, Zepeda, and Benton (2012) favored evaluations with a formative, developmental focus that offered opportunities for collaboration and reflective practice. Moreover, the evidence from these studies was thin at best. As Davis et al. (2011) observed in its comprehensive review of the literature on principal evaluations, “the literature ... leaves open the question of what impact, if any, stronger principal evaluation systems and practices may have on increasing effective leadership, strengthening teaching, reaching school improvement goals, or enhancing student growth” (p. 36). This study sought to help answer the first part of this question by exploring what impact IDAPA 121 had on the reflective practice, professional development and leadership practices of its participants.

Chapter Three

Research Methods

Introduction

One of the primary goals of IDAPA 121 is to “promote reflective practice and the development of ongoing, personalized development plans leading to improved support for turning around low-performing schools and measurably increasing student achievement for all students” (Idaho State Department of Education, 2013, p. 200). Because little research has been conducted on performance evaluation models like IDAPA 121, the purpose of this study was to describe how well the “promise” of IDAPA 121 compared with its actual impact on participants’ reflective practice, professional development, and leadership practices. Specific research questions that guided this inquiry were:

1. How do principals in Idaho describe the effect of the ‘professional practice standards’ of IDAPA 121 on their reflective practice and professional development?
2. How do principals in Idaho describe the effect of the ‘student growth component’ of IDAPA 121 on their reflective practice and professional development?
3. How does the experience of being evaluated under IDAPA 121 effect the leadership behaviors or practices of principals in Idaho?
4. How does the experience of being evaluated under IDAPA 121 compare with the experience of being evaluated before IDAPA 121?
5. What circumstances or experiences contribute most significantly to reflective practice and professional development for principals in Idaho?

With the foregoing research purpose in mind, this chapter describes this study’s research

methodology and is organized into the following sections: (a) rationale for research design, (b) participant selection and recruitment, (c) summary of information needed, (d) overview of research design, (e) data collection methods, (f) data analysis methods, (g) validity, reliability and ethics, (h) limitations and delimitations, and (i) a summary.

Rationale for qualitative research design

As Maxwell (2005) observes, “quantitative and qualitative methods are not simply different ways of doing the same thing. Instead, they have different strengths and logics, and are often best used to address different kinds of questions and goals” (p. 22). One of the primary goals of qualitative research, which aligns with the purpose of the current study, is “understanding the meaning, for participants in the study, of the events, situations, experiences, and actions they are involved with or engage in” (Maxwell, 2005, p. 22). Qualitative research is “suited to promoting a deep understanding of a social setting or activity as viewed from the perspective of the research participants” (Bloomberg & Volpe, 2008 p. 7-8). Consistent with that understanding, Creswell (2007) explains qualitative research is appropriate when we want to know “why people responded as they did, the context in which they responded, and their deeper thoughts and behaviors that governed their responses” (p. 40).

With these goals in mind, qualitative research was chosen for the current study because it centered on understanding its participants’ ‘experience’ with performance evaluation under IDAPA 121, the ‘meaning’ of reflective practice and professional development, and if and ‘why’ IDAPA 121 contributed to these in meaningful ways. Additionally, the study was interested in participants’ ‘deeper thoughts and behaviors’ related to specific aspects of IDAPA 121 such as its standards of professional practice and the one-

third of their evaluation that would be based on student performance. Because a qualitative approach is best suited for these goals, a qualitative approach was chosen for this study.

Rationale for basic qualitative research design

Before entering the field, I expected participants would report some impact from IDAPA 121 on their reflective practice, professional development, and school leadership. It was assumed that data concerning these impacts could be collected through participant interviews, documents, and the observation of events and activities related to these impacts. Given these assumptions, a case study design was originally contemplated for this study so that participants' experience with IDAPA 121 could be explored in depth from different angles and compared across multiple cases. In interviews, however, the first two participants reported that IDAPA 121 had no impact on their reflective practice, professional development, or leadership practices. Although the interviews were interesting, data-rich, and responsive to this study's research questions, there was nothing to be observed for these participants and little in the way of documents to review. This suggested that case study might not be the best design for the available data. A different research design was needed: one that could uncover answers to the main research questions while working primarily with interview data. Ultimately, the basic qualitative research design was selected as the best fit for both of these needs.

Authors use a variety of terms when describing a basic qualitative study. Thorne, Kirkham, and McDonald-Emes (1997) use the phrase "interpretive description." Sandelowski (2000) refers to a "basic or fundamental qualitative description". Merriam (2009) reports using a variety of terms over the years to describe such as study, including the terms generic, basic, and interpretive. However, "since all qualitative research is interpretive, I have come

around to preferring to label this type of study a *basic qualitative study* (Merriam, 2009, p. 22; emphasis in original). Following Merriam (2009), the term basic qualitative study is used in this study. Comprising the “most common type of qualitative research” in applied fields like education (Merriam, 2009, p. 22) the basic qualitative study can be summarized as follows:

Basic or generic qualitative studies have the essential characteristic of qualitative research (goal of eliciting understanding and meaning, researcher as the primary data collection and analysis instrument, use of fieldwork, inductive orientation to analysis, richly descriptive findings) but do *not* focus on culture, build grounded theory, or intensely study a single unit or bounded system. They “simply seek to discover and understand a phenomenon, a process, or the perspectives and worldviews of those involved” (Imel, Kerka, & Wonacott. 2002, p. 7, quoting Merriam, 1998, p. 11).

This focus on the ‘perspective’ and ‘worldview’ of research participants appeared a good fit for the research questions of this study and the interview-based data I now expected to collect. Moreover, as Merriam (2009) explains, “all qualitative research is interested in how meaning is constructed, how people make sense of their lives and their worlds. The *primary* goal of a basic qualitative study is to uncover and interpret these meanings” (p. 24; emphasis in original). That description accords well with the primary purpose in this study, which was to uncover and interpret the meaning that participants ascribed to their performance evaluation under IDAPA 121 and its relationship to their reflective practice, professional development, and leadership practices. For these reasons, a basic qualitative research design was adopted for this study.

Participant selection and recruitment

Selection. Because this study focused on the experience of Idaho public school principals with IDAPA 121, it was deemed my research participants would need to be Idaho public school principals. With more than 700 public K-12 schools in Idaho (EducationBug.org, 2016), and presumably an equally large number of public school principals to potentially interview for this study, a sound strategy to guide the selection of participants was needed. As Marshall and Rossman (2016) observe, “one cannot study the universe – every thing, every place, all the time. Instead, the researcher makes selections of sites and samples of times, places, people, and things to study” (p. 107). This section explains the logic and process of participation selection in this study.

A purposeful approach to participation selection is appropriate in qualitative studies (Creswell, 2007; Ravitch & Carl, 2016). A purposeful approach means “deliberately select[ing] individuals because of their unique ability to answer the research questions” (Ravitch & Carl, 2016, p. 128). Ultimately, the decision was made to work with elementary and middle school principals from two different Idaho school districts. The rationale for this approach is provided below. To maintain the confidentiality and anonymity of my participants, these districts are referred to as District Alpha and District Beta.

The first decision was to apply a mixed and stratified sampling approach (Ravitch & Carl, 2016, p. 136) by focusing on multiple school levels, rather than a single level such as elementary only. This approach facilitates comparisons across groups (Creswell, 2007, p. 127) and increases relevance and credibility of findings (Ravitch & Carl, 2016, p. 136). With a choice between two or three different school levels to work with, the decision was made to study elementary and middle school principals and to exclude high school principals because

it was believed that high school principals might be more difficult to recruit for the study. It was also thought that studying middle school principals and contrasting their experience with elementary principals could be interesting because middle school principals were less frequently studied than principals from the other levels.

The second choice was whether to study principals from one school district or multiple districts. It was decided to expand beyond a single district in order to provide opportunities for comparisons and contrasts across multiple districts. As with the benefit of working with principals from multiple school levels, working with principals from more than one district enables triangulation that increases relevance and credibility of the findings and allows comparisons across the districts (Creswell, 2007; Ravitch & Carl, 2016). If, for example, consistencies across two school levels and two separate districts were found, that would lend credibility to the weight of that finding. Conversely, if variation between the different school levels or school districts was found, that would indicate the need for additional research to explore those differences more fully. Having decided to work with principals from more than one district, the last issue to address was which of the eight possible districts should be studied?

Within qualitative research, “choices of participants ... should be driven by a conceptual question, not by a concern for representativeness” (Miles, et al., 2014, p. 33). In this study, the conceptual questions centered on the experience of principals with IDAPA 121 without regard to size of the districts where the principals worked or their districts’ rural or suburban character. Similarly, this study did not have a goal of comparing implementation of IDAPA 121 across multiple districts. This conceptual focus on principal experience afforded a wide latitude in choosing whichever districts would provide the greatest opportunity to

work with participants who met the primary selection criteria (Merriam, 2009). These criteria were: (a) public school principal in Idaho, (b) experienced performance evaluation under IDAPA 121 during the 2014-2015 school year, (c) experienced performance evaluation within their district prior to enactment of IDAPA 121, (d) willing and able to participate in this study. Guided by these criteria, I selected District Alpha and District Beta for this study.

Next, the question turned to “how many” principals to study (Marshall & Rossman, 2016, p. 108; Creswell, 2007, p. 126). As Merriam (2009) states:

Unfortunately ... there is no [universal] answer. It always depends on the questions being asked, the data being gathered, the analysis in progress, the resources you have to support the study. What is needed is an adequate number of participants ... to answer the questions posed at the beginning of the study. (p. 80)

In reviewing these factors and in consultation with my advisors, ten principals, five from each district, were chosen as the minimal sample size for this study. This was anticipated to be a sufficient number to explore the participants experience with IDAPA 121 in depth. This proved to be a valid assumption, as this sample size allowed me to reach the point of saturation or redundancy (Lincoln & Guba, 1985) where “no new information is forthcoming from new sampled units” (p. 202). By the tenth and final interview, it unfolded more like a ‘member check’ that validated the other participants’ experience and perspective about IDAPA 121. Although prepared to conduct additional interviews if necessary, it became clear through the course of the interviews and preliminary analysis that ten participants would be adequate to answer the questions posed by this study.

Recruitment. I started in District Alpha with the intent to apply the technique of snowball sampling (Ravitch & Carl, 2016) by asking the district’s administrators for referrals

to experienced and knowledgeable principals for this study. After further consultation with the district, this approach was abandoned in favor of a direct recruitment strategy. Needing five principals to interview from District Alpha, I randomly selected eight District Alpha principals - four from the elementary and four from the middle school level. Although it was not known what kind of response rate would be received, it was assumed that some number of the selected principals would decline participation or fail to respond. It was hoped that an initial contact list of eight principals would yield a reasonable combination of elementary and middle school principals. Before contacting these principals, I applied an additional level of criterion based sampling by reviewing their biographies on their school websites to confirm they met the key criteria of having been in the role for at least two years. This would ensure they had experienced with IDAPA 121 during the 2014-2015 school year and with evaluation prior to IDAPA 121.

I composed a brief email that explained the focus of the study and requested their participation in one or more interviews of approximately 60 – 90 minutes each. Within less than twenty four hours, one principal indicated that he would be happy to participate in the study, and the following day, a second principal responded to my email and agreed to participate as well. Within a week, and following additional recruiting efforts, three additional principals agreed to join the study.

As Table 1 (Participant Characteristics) shows, one District Alpha principal was from the elementary level and four were from the middle school level. Two of these principals were male and three were female. Although more balance between elementary and middle school teachers in the District Alpha sample would have been preferred, it was hoped the

imbalance could be corrected through the addition of the District Beta sample. Fortunately, this proved true as discussed below.

When recruiting participants in District Beta, a similar approach to that taken in District Alpha was followed. Initially, only two principals responded to the original request to participate. Following additional recruiting efforts, three additional principals agreed to participate. Fortunately, as shown in Table 1 (Participant Characteristics), one participant was a middle school principal and four were elementary level principals. When combined with the District Alpha principals who were oppositely composed, this resulted in a balance of five elementary and five middle school principals across both districts.

Table 1 (Participant Characteristics)

Participant Characteristic	District Alpha	District Beta	Total
Elementary level	1	4	5
Middle school level	4	1	5
Male	2	2	4
Female	3	3	6

Summary of information needed

Bloomberg and Volpe (2008) identify four types of information that are typically needed in qualitative research studies: contextual, perceptual, demographic, and theoretical. Because this study was focused on the experiences and perspectives of the participants with IDAPA 121 and related concepts of reflective practice and professional development, perceptual information was deemed the most relevant. As Bloomberg and Volpe (2008) explain, “perceptual information relies to a great extent on interviews to uncover participants’ descriptions of their experiences ...” (p. 70). Knowing that interviews would play a key role in my data collection, it was decided that interviews would be used to gather relevant demographic information from the participants.

Overview of research design

This section summarizes the major steps of this study. Following this brief overview, additional details are provided. I began by conducting a thorough review of the literature on major models of school leadership of the last four decades and their impact on school performance and student achievement. This included a heavy emphasis on the rise of instructional leadership as a policy focus, which in turn led me to the topic of performance evaluation of school principals and the discovery that Idaho was in the midst of implementing a state wide change in how its public school principals would be evaluated. This prompted me to re-start my research project and literature review with the new focus on principal evaluation as a means of developing and improving school leadership. This review suggested that performance evaluation might hold promise to positively impact school leadership, although significant gaps existed in our understanding if, or to what extent, this promise is real.

I obtained approval from the University of Idaho's Institutional Review Board (IRB) and District Alpha and District Beta to conduct this study. I contacted potential research participants by email and in some cases follow-up phone calls. Arrangements for in-person interviews were made with the principals who agreed to participate. Written informed consent to participate was obtained from all study participants in the form attached as Appendix B.

I conducted semi-structured interviews with my ten participants. All but two of these interviews were conducted in the participants' school office. For the convenience of the participants, the other two were conducted in a publicly accessible interview room at my place of employment. Interviews were recorded via audio-recorder and professionally

transcribed into electronic document form. Brief researcher memos were prepared at the conclusion of some interviews. Data analysis was conducted using printed and digital copies of the interview transcripts, researcher memos, and documents collected from participants.

Data collection methods. Qualitative data can be collected through a variety of sources including interviews, observations, review of documents, and audiovisual materials (Creswell, 2007). Merriam (2009) reports “in education, if not in most applied fields, interviewing is probably the most common form of data collection in qualitative studies. In some studies, it is the *only* source of data” (p. 86; emphasis in original). Although some data was collected through observation and documents, semi-structured interviews of the 10 participants provided most of the data collected in this study. Interviews were the most appropriate method to collect the perceptual information needed from participants because of this study’s focus on their experience and perspective (Bloomberg and Volpe, 2008; Marshall & Rossman, 2016, p. 151). A semi-structured approach to the interviews was chosen because it allows the researcher to solicit “specific information from all participants ... while responding to the situation at hand, to the emerging worldview of the respondent, and to new ideas on the topic” (Merriam, 2009, p. 90). Although a focus group could have been used, individual, face to face interviews were conducted because of the potentially sensitive nature of performance evaluations, and the potential of interviews to yield richer, more detailed information from each participant.

Pilot interviews and interview schedule. Following the recommendations of several prominent scholars (Creswell, 2007; Marshall & Rossman, 2016), preliminary pilot interviews were conducted with one former elementary principal from District Alpha and one elementary principal from District Beta before developing the interview protocol. Two

primary aims motivated these pilot conversations. The first was to confirm the overall viability of the proposed research topic. Both principals acknowledged it was an interesting and worthwhile topic with the potential to help school leaders. The second goal was to learn as much as possible about their experience with performance evaluation prior to implementation of IDAPA and about their opinions and feelings about the changes that IDAPA 121 would bring. The principals were asked about their recent experiences with evaluation, what they thought about reflective practice and being measured on student performance, what inspired and drove their own leadership style, and what informed their own professional development. The pilot conversations helped shape the content of this study's interview questions and the strategy for some of the topics. These pilot discussions enabled me to engage the study's participants more efficiently and effectively.

An initial interview protocol was developed on the basis of the literature review, the research questions, conceptual framework, and the pilot conversations. This featured open ended main questions (Rubin & Rubin, 2012) about the following: reflective practice, professional development, their leadership priorities, their experience being evaluated under IDAPA 121, comparisons with prior evaluation schemes, and being measured on student performance. Following a review and feedback from my advisor, the final list of interview questions was developed. A copy of this initial interview protocol is attached as Appendix C. After completing the first interview, an additional topic was added to the interview protocol that asked participants what changes they would recommend to improve performance evaluation and increase its impact on reflective practice, professional development and school leadership.

Interview process. Each interview was conducted in person. All but two interviews occurred at the participant's school office. I used the main questions from the interview protocol and augmented these with additional questions that arose spontaneously through the course of the interviews (Brinkman & Kvale, 2015, p. 161). The length of each interview ran between 45 and 80 minutes, with the average length being 54 minutes. All interviews were recorded via digital voice recorder. Those recordings were transcribed by a professional transcriptionist within 5 days of completion of each interview and converted to Microsoft Word files that were used in both digital and printed form throughout the research process.

The interviews were augmented by handwritten notes captured during the interview process and research memos drafted after completion of the interviews. Originally, I planned to conduct interviews of one or more staff members for each principal. The purpose of these additional interviews was to explore the relationship between the principals' experience with the performance evaluation and his/her leadership practices as observed by staff members who work closely with the principal. Because all participants reported there was no relationship between their performance evaluation and their leadership practices, the additional staff interviews were deemed unlikely to yield relevant information or additional insights. For this reason, no additional staff interviews were conducted.

Data analysis

Merriam (2009) describes data analysis as the "process of making sense out of the data" (p. 175). Creswell (2007) states that it consists of "preparing and organizing the data ... for analysis, then reducing the data into themes through a process of coding and condensing the codes, and finally representing the data in figures, tables, or discussion" (p.

148). This section describes the qualitative data analysis techniques and practices employed in this study.

As this was a basic qualitative study, I applied a “basic inductive and comparative analysis strategy” that is “suitable for analyzing data in most interpretive qualitative studies” (Merriam, 2009, p. 197). Specific analytic techniques included (a) simultaneous data collection and analysis, (b) open and axial coding, also known as first order and second order coding, (c) the creation and use of matrix displays to conduct comparative, (d) jotting and analytic memoing, and (e) the creation and testing of ‘assertions’ to derive this study’s final findings (Creswell, 2007; Maxwell, 2005; Miles, et al.; Yin, 2014; Merriam, 2009).

Additional details and examples of each technique are presented below.

Simultaneous data collection and analysis. Multiple authors recommend the simultaneous collection of data and analysis (Creswell, 2007; Maxwell, 2005). Creswell (2007) notes “the process of data collection, data analysis, and report writing are not distinct steps in the process – they are interrelated and often go on simultaneously in a research project” (p. 150). Maxwell (2005) explained this important aspect of qualitative research as follows, “the experienced qualitative researcher begins data collection immediately after finishing the first interview or observation, and continues analyzing the data as long as he or she is working on the research, stopping briefly to write reports and papers” (p. 95). Based on this advice, simultaneous data collection and analysis occurred throughout this study.

Immediately following the first interview, I spent time reflecting on what was learned and mentally comparing the participant’s experience with what the literature reported and what was learned in the two pilot conversations. I took additional notes to capture thoughts, and continued to reflect on the interview that morning and later that evening after work. Two

days later, the transcribed record of the first interview was received and printed. That night, I read through the transcript two times - once without taking notes, and a second time with a highlighter to identify potentially interesting parts of the dialogue. I assigned simple, descriptive words or phrases to the highlighted language in the margin of the paper, thus beginning the process of ‘open’ or ‘first cycle’ coding described more fully below. This early and active engagement with the transcript helped identify ways to improve the subsequent interviews. It also prompted me to add a new question to the interview protocol that asked participants what recommendations they might have to improve the evaluation process. This addition yielded the fourth major finding of this study as detailed in Chapter Four. With each subsequent interview, I followed a similar approach of immediate reflection and comparison across prior interviews, note taking, and the active reading and coding of the interview transcripts.

This study benefited from simultaneous data collection and analysis in a number of ways. On a practical level, it helped prevent the data analysis from becoming a “giant ... overwhelming task” (Miles, et al., p. 70). The simultaneous approach enabled me to quickly identify and remedy a potential data gap after the first interview as noted above. It also helped make analysis a “lively ongoing enterprise” (Miles, et al., p. 70), enabling and enriching many of the other analytic techniques used in this study.

Coding. Although the technique of coding was originally associated with grounded theory research (Corbin & Strauss, 1990), it is now widely used in qualitative research generally (Maxwell, 2005; Miles, et al., 2014) and is appropriate for basic qualitative research (Merriam, 2009). Saldaña (2013) defines a code as “most often a word or short phrase that ... assigns a summative, salient, essence-capturing, and/or evocative attribute or a

portion of language based ... data ... [such as] interview transcripts” (p. 3). Codes can be derived both inductively and deductively (Miles, Huberman, and Saldaña, 2014), and the process known as coding comprises at least two levels of activity that are commonly called ‘open’ or ‘first cycle’ coding and ‘axial’ or ‘second cycle’ coding (Creswell, 2007; Miles, et al.). This section describes the coding processes used in this study. The terminology ‘first cycle’ and ‘second cycle’ coding is employed hereafter for simplicity.

As described earlier first cycle coding was conducted on each interview transcript. This entailed (a) actively reading each transcript several times and highlighting key words, phrases and passages that appeared responsive to the underlying research questions that guided this study, and (b) ascribing short descriptive words or phrases to most of the highlighted sections in the right hand margin of the printed transcript. Some of these first cycle codes were ‘deductively derived’ ahead of time from the research literature and research questions (Miles, et al., p. 81). Three examples of deductively derived codes were “reflective practice (RP)”, “professional development (PD)”, and “student growth (SG)”. Other first cycle codes were derived “inductively”, meaning they “emerge[d] progressively during data collection” (Miles, et al., p. 81). Examples of inductively derived codes were ‘timing’ that related to when the participants were actively engaged in evaluation related activities and ‘what guided’ that related to the factors and circumstances that drove the participants’ reflective practice, professional development, and leadership focus. Whereas deductively derived codes have a strength in connecting one’s data back to the conceptual framework and underlying research questions, inductively derived codes demonstrate that “the researcher is open to what the site has to say rather than determined to force-fit the data into preexisting codes” (Miles, et al., p. 81).

As noted earlier, simultaneous data collection and analysis played a role in the first cycle coding as well. With each new transcript, the approach and decisions related to coding were informed by what had been learned in the earlier interviews. This is consistent with the following observation of Miles, Huberman, and Saldaña (2014): “codes will change and develop as field experiences continues ... [s]ome codes do not work; others decay ... others flourish, sometimes too much so” leading to the need for “breaking down codes into subcodes” (p. 82). Thus, the techniques of simultaneous data collection and analysis and first cycle coding were tightly interwoven in this study.

Second cycle coding. Whereas first cycle coding allows the researcher to “summarize segments of data”, second cycle coding “is a way of grouping those summaries into a smaller number of categories, themes or constructs ... that identif[ies] an emergent theme, configuration, or explanation” (Miles, et al., p. 86). These authors further explain that second cycle coding creates “a sort of meta-code” (p. 86). Consistent with the earlier discussion of simultaneous data collection and analysis, second cycle coding began early in the data collection process and continued as new information came in from additional interviews. Once all of the transcripts had been coded, additional rounds of second cycle coding were performed by going back through the first cycle codes and reviewing notes and memos that were created throughout the process. This process generated an initial list of 16 second cycle codes presented in a matrix display as discussed below. A complete list of the 16 second cycle codes can be found in Appendix D.

Matrix displays. Visual displays, including the creation of matrixes or tables of data, are a “primary way of analyzing data in fresh perspectives” (Miles, Huberman, & Saldaña, 2016, p. 91). The authors further explain that matrix displays help to “organize the vast

amount of condensed material into an ‘at-a-glance’ format for reflection, verification, conclusion drawing, and other analytic acts”, including cross-case analysis as discussed below (p. 91). Following this advice, I next created what would become the first of many matrix displays to facilitate deeper levels of analysis. This initial matrix display was printed horizontally on legal size paper, with five columns and six rows per page. The first column was populated with numbers representing each participant. The additional columns two – five on each page were populated with second cycle codes such as ‘RP’ (reflective practice) and PD (professional development). Ample space was provided in each cell of the matrix to allow handwritten notes from each transcript. Table 2 provides a partial example of one page from this matrix display.

Table 2 (Example of Matrix Display)

Participant	RP	PD	Challenges	Impact of Eval
1				
2				
3				

Once this matrix was created, I reviewed the transcripts and their first cycle codes again, extracted key data points, details and quotes that related to each of the second cycle themes in the matrix, and wrote these details in their corresponding cells. For example, in cell 1- RP, which corresponds to William and the second cycle code of reflective practice, I wrote “automatic, every day, natural response to events @ school, RP + observations → goals & PD”. In cell 2-RP, which corresponds to Nancy’s experience with reflective practice, I wrote “RP very important, no bearing on RP, necessary for growth.” Although this was a laborious and sometimes tedious process, the fully completed matrix provided a helpful ‘at a glance’

format for further analysis (Miles, et al., p. 91). The process of populating the matrix also had the side benefits of taking me back through the transcripts, an experience that would continue to yield occasional nuggets of insight through the balance of the process and enabling some aspects of case analysis.

Following the protocols of Yin (2014) and Miles, Huberman, and Saldaña (2014), the matrix was used for systematic analysis and comparison of each second cycle code across (a) all participants, (b) both school levels, and (c) both school districts. According to Yin (2014), analysis of this nature increases the quality of one's study (p. 170). Miles, Huberman, and Saldaña (2014) describe a similar strategy which uses a 'meta-matrix' like the one described above to enable "systematic comparison" (p. 103). This analysis revealed similar profiles across all factors for some codes, isolated differences between the two school levels, and small differences between the two school districts as reported in Chapter Four. Through the process of studying the matrix and 'jotting', described below, several additional 'meta-codes' or themes emerged. Three examples included 'conflict between formative and summative processes', 'non-evaluative support', and 'policy disconnect', underscoring the inherently iterative nature of qualitative analysis.

Jotting. Throughout the process of coding described above, I engaged in 'jotting' which is described as creating "analytic sticky note[s]" (Miles, et al., p. 93). The authors explain these jottings, "hold the researcher's fleeting and emergent reflections and commentary on issues that emerge during fieldwork and especially during analysis ... [and] strengthen coding by pointing to deeper or underlying issues that deserve analytic attention" (p. 94). As suggested by the authors, I jotted some of these notes in the margin of the interview transcripts while coding, and in separate notebook pages when analyzing the

completed matrix discussed above. Those jottings arose from my immersion in the data and cross-case analysis discussed earlier.

Analytic memoing. Miles, Huberman, and Saldaña (2014) define an analytic memo as “a brief or extended narrative that documents the researcher’s reflections and thinking processes about the data” (p. 95). The authors further explain “these are not just descriptive summaries of data but attempts to synthesize them into high level analytic meaning” (p. 95). Referring to analytic memos as “one of the most useful and powerful sense-making tools at hand”, the authors recommend the following: (a) memoing should be a top priority, and they urge the researcher to stop whatever is happening and write the memo as soon as an idea strikes, (b) memoing should begin as soon as data becomes available and continue throughout the report writing process, (c) memos should be about ideas and not simply summaries of the data (p. 98-99). Consistent with these recommendations, I wrote several analytic memos during the research process. One example of an analytic memo was written after the coding was completed and I had spent considerable time reviewing the growing pile of jottings and the completed matrix discussed earlier. This memo could be described as a composite portrait of the participants’ experiences with evaluation under IDAPA 121, reflective practice, professional development and leadership practices. This memo distilled the most significant points that had learned about participants’ experience with IDAPA 121. This memo covered eighteen points and served as the basis for a longer, type written summary that was prepared a few weeks later. A typed version of the original eighteen points is attached as Appendix E.

Assertions and propositions. As the final step in qualitative analysis, Miles, Huberman, and Saldaña (2014) recommend researchers generate ‘assertions’ and/or

‘propositions’ in order to “formalize and systematize the researcher’s thinking into a coherent set of explanations ... that reflect the findings and conclusions of the study” (p. 99). They define an assertion as “a declarative statement of summative synthesis, supported by confirming evidence from the data and revised when disconfirming evidence or discrepant cases require modification of the assertion” and define a proposition as “a statement that puts forth a condition event – an if-then or why-because proposal that gets closer to prediction or theory (Miles, et al., p. 99-100). Following creation of the final, ‘composite portrait’ analytic memo described above, I sensed that it was time to move to this final stage of analysis.

Armed with the completed matrix, jottings, analytic memos, and an organic understanding of what was really going on in this study, I created the “bullet points of major patterns, themes, trends, and findings that [could be] confidently put forth about [this] study” (Miles, et al., p. 100). Some of these initial assertions survived scrutiny as they were cross checked for confirming and disconfirming evidence from the matrix and other analytic materials. Examples of assertions that survived include “evaluation had no impact on the participants’ reflective practice or professional development” and “participants believe policy makers are out of touch with what really happens in schools.” Several other assertions failed to survive this analysis against the evidence and were dropped. One example is the assertion “participants believe there is an inherent conflict between the formative goals of IDAPA 121 and the summative aspects of the evaluation process” which was found to be true for some participants and untrue for others. Another example of an assertion that was not supported by the evidence was an analogy to principals being like individuals strapped inside a rocket-ship for the duration of the school year, isolated from their peers and unable to focus on meaningful development because of the frenzied pace of the school day and year. Although

there was strong evidence to support this from a few of the participants, the evidence was not pervasive across the participants, so it too was dropped from this study's findings. As I continued to cross check against the evidence and adjust the content and wording of the assertions, the study's final four findings emerged to tell a clear and well supported story of what was really happening (or more accurately, what was not happening) with evaluation under IDAPA 121.

One final point about the analysis process bears mentioning here. The foregoing sequential list of analytic techniques might give the impression these techniques were employed separately from one another, and independent of data collection. On the contrary, many occurred simultaneous with data collection and in overlapping and complementary ways. For example: (a) coding occurred while data collection was ongoing, (b) second cycle coding occurred at the same time as first cycle coding, and (c) jottings and analytic memoing occurred during coding and continued during and after creation of the matrix display as well.

Validity, reliability and ethics

Validity. Internal validity “deals with the question of how research findings match reality. How congruent are the findings with reality? Do the findings capture what is really there?” (Merriam, 2009, p. 213). To enhance the internal validity of this study, the techniques of rich data, member checking, triangulation, and reflexivity were employed (Merriam, 2009; Maxwell, 2005; Miles, et al.). Each of these is described below.

Rich data refers to the collection and presentation of “data that are detailed and varied enough that they provide a full and revealing picture of what is going on” (Maxwell, 2005, p. 110). In the present study, and as Maxwell (2005) advises, major findings and conclusions are supported by “verbatim transcripts of the interviews, and not just notes on what the

researcher felt was significant” (p. 110). Allowing the participants to speak for themselves, in their own words helps counter the threat of researcher bias and increases the internal validity of the reported findings (Maxwell, 2005; Miles, et al.).

Member checking, also known as ‘respondent verification’ is defined as “systematically soliciting feedback about your data and conclusions from the people you are studying” (Maxwell, 2005, p. 111). Maxwell (2005) asserts that member checking is “the single most important way of ruling out the possibility of misinterpreting the meaning of what participants say and do and the perspective they have on what is going on” (p. 111). Member checking was performed in this study by emailing participants a copy of his/her ‘participant summary’ as found in Appendix F and a summary of the major findings in the form attached as Appendix G. Participants were asked to verify the accurateness of both documents and provide any needed revisions via return email. They were also invited to share any other feedback they had concerning the study or its findings. As shown in the Member Check Results detailed in Appendix G, six participants responded to the member check request. All six confirmed the accuracy of the major findings presented in Chapter Four.

Triangulation is another “well known strategy to shore up the internal validity of a study” (Merriam, 2009, p. 215; Creswell, 2007; Maxwell, 2005). Triangulation involves the use of multiple perspectives to validate and support a finding. These multiple perspectives can come from different data sources, different methods, different researchers, different theories or even different types of data (Ravitch & Carl, 2016, p. 195). Because the information needed in this study could only come from interviews of principals who had experienced IDAPA 121, triangulation of data sources was achieved by having multiple

participants (rather than a single point of view) from more than one school district and school level.

A final technique used in this study to increase validity is known as “reflexivity” (Ravitch & Carl, 2016; Maxwell, 2005) which seeks to minimize the impact of the researcher’s personal perspective on his/her participants and the researcher’s own interpretation of the data. In the present study, I endeavored to maintain a reflexive perspective throughout by declaring his own less than rewarding experience with performance evaluations in the proposal, and by avoiding the use of leading questions in the interviews. During data analysis, I ensured data supported emerging understanding of participants’ experience by repeatedly going back to the transcripts to ensure that solid evidence from multiple participants supported the findings and that no contradictory perspectives had been provided. As noted previously, several tentative findings were dropped because of contradictory or insufficient data in the data.

Reliability. Reliability, in the traditional sense, refers to “the extent that research findings can be replicated” (Merriam, 1998, p. 205). Within the context of qualitative research, however, the term has a different meaning. Lincoln and Guba (1985) suggest that instead the researcher should focus on the “dependability” or “consistency” of the results obtained from the data (p. 288). To enhance the reliability of the research findings, several techniques were employed including (a) triangulation of sources, mentioned above, as this strengthens internal validity and reliability, and (b) creation of a clear audit trail by describing “how data were collected, how categories were derived, and how decisions were made during the inquiry” (Merriam, 2009, p. 222).

External validity, also known as generalizability, “is concerned with the extent to which the finding of one study can be applied to other situations” (Merriam, 2009, p. 223). While external validity is one of the explicit goals in *quantitative* research, many qualitative researchers resist or outright reject the importance of external validity in their studies (Merriam, 2009; Miles et al., 2014). In this study, the relatively small sample of principals – all coming from large, suburban districts – could not be expected to speak for all other principals in Idaho when it comes to IDAPA 121. Yet, it is hoped that the research design and data analysis has yielded what Stake (1995) calls “naturalistic generalization”. This allows readers to “gain insight by reflecting on the details and descriptions presented in case studies ... and consider whether their situations are similar enough to warrant generalizations” (Mills, Durepos, & Wiebe, 2010). Merriam (2009) recommends three strategies to increase the likelihood of naturalistic generalization that are employed in this study. The first strategy is to provide rich, thick descriptions of participants’ experience so readers can determine if their own situation is similar to the participants’, and thereby determine if the findings and conclusions might apply to their own situation. The second strategy is called ‘typicality or modal category’ and entails describing how typical participants are compared with the broader class “so that users can make comparisons with their own situation” (Merriam, 2014, p. 228). The description of what participants experienced in their evaluation helps with the first and second strategies. The third strategy is the use of a multi—site design that increases the diversity of participant experiences and expands the potential range of transferability. In this study, diversity was achieved by working with principals from two districts and two school levels.

Ethics. Miles, et al. (2014) catalogue a number of specific ethical issues that frequently require attention before, during and after a qualitative study. As they explain, “[w]e cannot focus only on the quality of the knowledge that we are producing, as if its truth were all that counts. We must also consider the rightness or wrongness of our actions as qualitative researchers in relation to the people whose lives we are studying, to our colleagues, and to those who sponsor our work” (p. 288). With this admonition in mind, several ethical considerations were addressed during the planning and execution of this study, including obtaining informed consent of participants, anticipating and mitigating potential harm and risk, and maintaining privacy, anonymity, and confidentiality of participants. Each is discussed further below.

Informed consent. This is a cornerstone of ethical research, and requires that participants in my study fully understand what will be expected of them, any risks that may be involved, and voluntarily agree to participate. Following standard protocol, all participants were provided, and required to first sign, a copy of the informed consent form attached as Appendix B before participating. Importantly, this consent enabled them to withdraw from participation at any time and for any or no reason without repercussions.

Harm and risk. Miles, et al. posit that real or feared harm is inevitable in qualitative studies, and can come in a variety of forms. This can include “blows to self esteem or ‘looking bad’ to others, to threats to one’s interests [or] positions ...” (p. 61). It was understood this study’s findings could be embarrassing to the principals who participate or to the school district and/or the Idaho State Department of Education. For this reason the confidentiality and anonymity of participants was maintained throughout the research and writing process as outlined more fully below.

Privacy, anonymity and confidentiality. The participants rightfully expect this study to respect and maintain their privacy, anonymity and confidentiality. Not only is this right thing to do, it may also increase the quality of this study by helping to build and maintain trust. Although some important commitments on these issues are spelled out in Appendix C, active steps were taken to maintain these commitments throughout the research process. These included the standard operating procedures of masking the identity of participants in recorded interviews, notes and the final report, and keeping research materials in a safe and secure location.

Summary

This chapter provided a detailed explanation of the research method employed in this study. A basic qualitative research design was chosen to explore the meaning and experience of performance evaluation for Idaho principals under IDAPA 121, and to understand the relationship, in their experience, between their evaluation, reflective practice, professional development, and leadership practices. Ten active Idaho public school principals were purposefully selected for this study using a combination of criterion and network sampling. Data were collected primarily from semi-structured interviews of participants; interviews were recorded and transcribed. Credibility of findings was enhanced through selection of participants from two different school districts and two different school levels.

An extensive review of the literature on performance evaluation of school principals was conducted to inform the conceptual framework and research questions that guided this study. Well established techniques of data analysis were employed to derive the primary findings presented in Chapter Four. These included: (a) simultaneous data collection and analysis, (b) first order and second order coding, (c) the creation and use of matrix displays to

conduct within-case and cross-case analysis, (d) jotting and analytic memoing, and (e) the creation and testing of ‘assertions’.

Measures were taken throughout design and implementation of this study to enhance the internal validity, reliability, and external validity of the results and to minimize potential risks for its participants. Internal validity was addressed through a combination of rich data, member checks, triangulation of sources, and researcher reflexivity. The reliability of findings was enhanced through triangulation of sources and presentation of a clear audit trail. External validity was promoted through the use of rich data and a diverse research sample. Risks to participants were minimized through the collection of informed consent and multiple measures to maintain their privacy and anonymity.

Chapter Four

Participants' Experience with IDAPA 121

Introduction

The purpose of this study was to describe how well the “promise” of IDAPA 121 compared with its actual impact on participants’ reflective practice, professional development, and leadership practices. This chapter presents the key findings obtained from interviews of ten elementary and middle school principals regarding their experience with IDAPA 121 and its impact on their own reflective practice, professional development and leadership practices. These major findings are:

1. Participants experienced evaluation under IDAPA 121 as a perfunctory, isolated end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice.
2. The student growth component of IDAPA 121, and the prospect of being evaluated partly on student growth had no impact on the participants’ leadership focus or practices.
3. Participants believe that policy makers and school policy related to their performance evaluation are out of touch with what happens in schools.
4. Participants had clear ideas for improving the system so principals would experience more meaningful reflective practice and professional development.

This chapter begins with a description of the setting of this study. Thereafter, the four major findings of this study are detailed. These findings are revealed through the descriptions and actual words of participants taken verbatim from the interview transcripts. This provides “thick description” (Lincoln & Guba, 1985) that conveys the reality of participants’

experience with performance evaluation under IDAPA 121 while providing a solid foundation for the interpretive analysis in Chapter Five.

Setting

This study was conducted in Idaho within the public school districts Alpha and Beta. District Alpha and District Beta are suburban and semi-rural school districts in south-western Idaho. Data were conducted primarily through interviews of the study's ten participants who are public school principals in District Alpha and District Beta. Table 1 (Participant Overview) identifies the name, school district and school level for each participant.

Table 3. Participant Overview

Participant Name	District	School Level	No. of Years as principal
William	Alpha	Middle-School	~5 years
Nancy	Alpha	Elementary	> 5 years
James	Alpha	Middle-School	> 5 years
Elizabeth	Alpha	Middle-School	> 5 years
Charlotte	Alpha	Middle-School	> 5 years
Robert	Beta	Elementary	> 10 years
Mary	Beta	Elementary	> 5 years
Edward	Beta	Middle-School	< 5 years
Scott	Beta	Elementary	> 5 years
Henry	Beta	Elementary	> 15 years

Major Findings

This section details the four major findings of this study as told through the actual words and descriptions of the study's participants.

Finding 1. Participants experienced evaluation under IDAPA 121 as a perfunctory, isolated end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice. Contrary to the promise of IDAPA 121, a primary finding of this study is that participants experienced performance evaluation under IDAPA 121 as a perfunctory and isolated end of year event without meaningful impact on their reflective practice, professional development, or leadership practice.

This finding is reported in four sections: (a) the perfunctory, end of year nature of their evaluation experience; (b) participants' perspective and experiences about their reflective practice; (c) professional development and its non-existent relationship to their evaluation; (d) the disconnect between participants' performance evaluation under IDAPA 121 and their leadership practices.

Perfunctory and isolated end of year event. All participants described their performance evaluation occurring at the end of the 2014 – 2015 school year with little to no discussions about performance expectations or the evaluation process at the beginning of the year or during the year. For example, Robert reported that “we met, I want to say the first week of May”. Nancy reported a similar experience, “I had my performance evaluation a couple of weeks ago (middle to late May). I was not very familiar with the performance evaluation form before the evaluation.” When asked to describe the process by which she reviewed her performance with her evaluator during the year, Nancy reported she did not discuss her performance or the evaluation during the school year. She explained “my evaluator had been in my building three times this year. Twice to run – to see how lockdown

drills were run. And once to talk to me about a student we would be receiving next year. And then, I guess, a fourth time he came in was to do my principal evaluation.”

Consistent with the other participants, Edward met with his supervisor in late May. When Edward was asked if he had any discussions with his evaluator during the school year about his performance expectations or the evaluation prior to his evaluation meeting, he replied, “No. I honestly couldn’t even tell you the categories as they sit now in my evaluation. Last year it was the same process really.” Edward reported a similar experience. “About two weeks ago (late May), I met with my supervisor who is in charge of my evaluation.” When asked if he had any discussions with his supervisor at the beginning of the year or during the year about his evaluation, he answered, “No. We never talked specifically about the evaluation.” Edward characterized her evaluation experience as “mostly a summative, end of the year event.” When asked if she knew what performance criteria would be used to evaluate her performance that school year, she replied “yes, because they had sent it out ... I think they had made changes to it ... but did I pay attention to it? No.” As these examples illustrate, participants in both districts experienced their evaluation as a single, end of year event. None of the participants discussed performance expectations with their supervisors at the beginning of the year or during the course of the school year.

Regarding the substance of the end-of-year discussion, the participants from District Alpha reported a more detailed evaluative discussion when compared with the principals from District Beta. Here is how Nancy from District Alpha described her evaluation discussion:

We talked through it together. It was a conference. He didn't come in with anything pre-populated or pre-done. He had my goals listed that I set for the year, so

we talked through what I had done, and what he had seen, and what he had heard, and we decided as a team where I would fall in the category areas, and then I provided evidence of what I had done. And then we just talked about my movement toward meeting my goals on the form.

Elizabeth from District Alpha reported a similar experience, with her evaluator asking her “where I thought I fell on the evaluation tool.” As she further described, “I emailed him back and I told him the areas I felt like I was proficient and the areas I felt like I was distinguished and that I would provide evidence for him on those distinguished areas, but if he had specific questions for me in the proficient areas to please email me those questions.”

Compared with the experiences of the District Alpha participants, participants from District Beta reported having shorter discussions about their performance with little or no focus on performance rating or the evaluation form itself. Robert’s experience was typical:

We spent a long time talking about [my career goals]. We talked honestly ...it was not a formal look at a document and check things off. It was a conversation about my leadership style. And honestly, the evaluation part was about a minute, and I never saw the document, and he was looking at his computer screen. I'm not even positive that the document was there. And he told me, you know, I'm going to type this thing up and ...I'll send it to you ... I'll get it to you sometime this summer. I haven't seen it yet (as of June 8, the date of our interview).

Scott from District Beta reported a similar experience as follows:

This year ... there was no evaluative discussion ... We talked about the year in review, but more as a how did things go, as opposed to discussing goals and objectives or whether they were met or not. The first few years as an administrator,

we used to actually sit down and ... the form would actually be out, and we would go through line item by line item and talk about ratings relative to the individual objectives. But for the last several years that hasn't been the case. We haven't spoken to specific objectives, and the form hasn't even been present during the last several -- they've been more like discussions and then the form was completed after the fact and sent to us electronically.

As Mary from District Beta describes below, the evaluation form was more present in her meeting than was the case with Robert or Scott, although it was not reviewed or discussed in any great detail. She described her evaluation experience as follows:

My evaluation was pretty basic. My evaluator asked me a lot of questions about how I felt the year went, what I needed help with, where I saw myself going in the future as a leader. I did most of the talking and [my evaluator] asked some questions ...

Q. Okay. Right. And did your evaluator use a form?

There is a form, yes. We didn't go over that form piece by piece. She asked some general questions that related to that form and then put that into the actual form and sent that to me.

As these interview excerpts reveal, the participants experienced performance evaluation under IDAPA 121 as an isolated, and relatively light, end of the year event. In reflecting on his experience and the quality of his end of year conversation, Robert lamented the absence of meaningful discussion of his professional goals, obstacles or how he was working through them. He noted that such conversations, if they were to occur, "is where we're going to see real growth." Unfortunately, as he experienced the evaluation conversation, he noted that "it feels like a postmortem ... it has no room for the ongoing

reflection.” Nancy summarized well the substance and overall feeling of her evaluation experience as follows: “It’s just an hour, come in and let’s get it done, and put it in the file and then do my real job.”

Reflective practice. Participants characterized reflective practice as critical to their success as school leaders, and an essential part of their professional development and leadership practice. When asked to describe what reflective practice meant to him, William offered the following:

So, for me it just simply means ...being able [to] identify what is productive, what is efficient, and what is not ... to set aside any pride, being willing to admit when things are not working ... whether it's ... a pedagogical approach or maybe a system structure within the school, or whatever the case maybe, but just looking back on ...what things that are being done, things that are being said ...and determining what's working, what isn't.

As Nancy stated, “I believe to be an effective school administrator you have to be reflective in what you do.” William concurred, observing that “I don’t think a person can reach their maximum ... abilities, without reflective practice.” Other participants conveyed the value of reflective practice by describing it as “very important” (Elizabeth), “super-important” (Edward), and “necessary for growth” (Robert). William explained his own reflective practice served as a natural precursor to goal setting and planning.

Participants reported engaging in reflective practice regularly, on their own or with their leadership team if they have one, and expressed the belief that reflective practice comes naturally for any principal who is motivated to help her school, students and staff. William reported engaging in reflective practice “every single day” and described his reflective

practice as “an automatic thing.” Edward described the regular nature of her reflective practice this way:

I am a person who – like I said – after every event we’ve done, try to get together the people who were involved and say ‘how do we make it better?’ ... Whether it’s registration, whether it’s Back to School night, whether it’s testing, it doesn’t matter. How will we make it better?

Nancy shared a similar experience,

At the end of the year I’ve always debriefed with office staff, because they are the only people still here after the last day of school. We ask ourselves, what went well this year? What didn’t? What do we want to change for next year? How can we make things smoother as an office and an administrative team?

For these participants, reflective practice occurs daily, if not multiple times a day, in response to events and circumstances within the school and can occur spontaneously in the moment or in a more formal way. Participants also reported engaging in reflective practice to drive strategic planning for their school, as the above quote from Nancy shows. Although the triggering event or focus of reflective practice varies across situations, it usually involves some version of the following questions: “Did I do the right thing? Should I have handled that differently? Was this good for our students?” Nancy provided a clear example of this when reflecting on the end of the year debrief with her leadership team:

This year, because there were leadership premium positions for teachers, I was able to pull my leadership team, which is 15 staff members in the day after school was out for a four-hour meeting to debrief. That had nothing to do with the evaluation tool, but that was the first time I’d ever been able to do that. To actually ask teachers,

okay, let's break it down into all the different areas we want to and talk about it.

What did we like? What don't we like? What do we want to change? So I received lots of feedback from the staff as a whole, both positive and negative, and we reflected as a group. And that was really healthy for me and very helpful in changes we need to make for next year, just to build our community and make our system better.

When asked to describe the relationship between their performance evaluation and their reflective practice, all ten participants reported that there was no meaningful relationship between them. As Nancy concisely explained, “the principal evaluation this year had no bearing on my reflective practice.” She further explained:

I know where I'm doing a good job and where I'm not, and someone coming in to evaluate me who's been in my building an hour out of the year really doesn't know what I do. But I do, and so I am a better evaluator, because I'm a little Type A of my own abilities, and strengths, and weaknesses probably, than anybody else ever could be.

Edward described the relationship between his performance evaluation and reflective practice as “completely disconnected.” Mary shared the same experience. When asked to describe the relationship between her evaluation and her reflective practice, she responded this way:

When I was reflective was I thinking, I better look at this because it's on my evaluation? No, I didn't. Looking at my evaluation was never what drove anything I did. Meaning that, oh, it's on my evaluation, I better look at it. I would look at things that I thought were important. I also am a firm believer that people who want to do

better are doing a reflective process anyway and they're not waiting for a major evaluation.

Elizabeth echoed the experience of reflecting on matters that she deemed important rather than being guided by the evaluation process. As an example, she described how she worked with her leadership team to create their professional development plan for the coming year:

So, I think my teachers are doing a lot of inquiry strategies in the classroom, but not everything is as purposeful as I need it to be. And I have reflected that we really need that backwards design to make it more purposeful. And so, as we met as a leadership team we developed our professional development plan for next year, and that we will do a review of inquiry, formative assessment, guiding and essential questions for new staff members, and then we're going to take off and run with backwards planning and beginning training our staff in how to do that right.

In describing her reasons for focusing her reflective practice in this way, she explained that her focus came not from her “evaluation or anything that the district was expecting of [her] formally”, but rather her own sense of what is “effective ... and good for kids.”

Underscoring the disconnection between his evaluation and his reflective practice, William acknowledged he “had not given his performance evaluation a single thought all year” until he was approached to participate in this study in May of 2015.

Professional development. On the question of professional development, all participants explained they valued it, and that it can help them to be a more effective leader. As William explained, “I always have two or three specific [professional development] goals that I share with my assistant principals because when you share a goal it becomes more

powerful.” Scott reflected on leadership development that he received with a group of other principals earlier in his career, observing, “I think I benefited from it personally and I still reflect on some of the lessons that we learned in that, as well as some of the materials that we received.”

Although they all find professional development valuable, they reported a general perception that formal development programs for principals are less prevalent now than in the past. For example, participants from both districts (Nancy and Edward) reported receiving stipends in prior years to attend leadership development conferences, but no longer receive these. However, as Edward explained, “Well, then, obviously, with the recession, that went away ... and it hasn’t been readdressed.”

Most professional development for participants currently takes the form of self-directed learning such as reading, taking graduate school courses, or by working through novel or challenging situations on the job. William, for example, described his professional development program this way: “in the last couple years I've been working towards an advanced degree, which I completed; this year it's more about diving into one particular book on a very detailed basis.” In reflecting on his professional development this part year, Edward emphasized ‘on the job learning’:

My growth this year really has been around experiences at my building that I hadn’t experienced before. So my own growth has been in the trenches, not anything formal. I was faced with several opportunities, having to walk through some interesting situations with parents in the building who were at odds with each other. Those have been my growth opportunities this year really.

Regarding the impact of the 2014-2015 performance evaluation on participants' professional development, they unanimously reported that it was a "non-issue" and non-factor. Nancy agreed "the current evaluation system experienced this year had no bearing, positive or negative, on [my] professional development." Edward concurred, explaining that "to my development as a principal or my own goals that I have for my leadership, it [my performance evaluation] is a non-issue." He further elaborated that professional development, although important to him, was 'completely disconnected from the evaluation system' and that 'whatever goals I set for myself I just set them and my evaluator has no idea what they are really.' As Edward explained, "even informal conversations with my evaluator or bosses about 'what are the areas you want to grow in, how can we help you, how can we support you, those I haven't experienced either.'" Nancy reported a similar experience, noting that her performance evaluation "really doesn't speak to the things she was working on [in terms of her professional development] and she has her own independent means of identifying her professional development goals outside of this [evaluation] process." Nancy summarized this finding well by declaring the evaluation a "non-factor" and a "check the box exercise" and by surmising that "next year, the same would be true."

To summarize, participants valued reflective practice and professional development, and believed both were essential to their continued growth and efficacy as school leaders. At the same time, they reported IDAPA 121 and their evaluation experience during the 2014-2015 school year had no bearing or impact on their reflective practice or professional development. Perhaps unsurprising, they also reported IDAPA 121 had no bearing on their leadership practice as described more fully below.

Leadership practice. Regarding the impact of their performance evaluation process on their leadership priorities and practices, all participants from both districts reported the evaluation did not drive their practice. William from District Alpha put it this way: "... honestly, there has not been one day this entire year where I've thought about doing something because of that evaluation piece. Not one day. Not one minute." Edward from District Beta shared the same experience, "Looking at my evaluation was never what drove anything that I did. Meaning that, oh, it's on my evaluation, I better look at it. I would look at things that I thought were important.

Instead, their leadership goals and practice were driven by their own understanding of what their school, its students and teachers needed them to focus on. Edward explained that her leadership actions were:

Guided more by her understanding of what the school needed and what she needed to be doing differently with her leadership team based on data and other events that came to her, as opposed to being connected to what she knew the district valued her to be doing.

Charlotte's experience was nearly identical, as she explained:

Did the evaluation form and related expectations inform my own personal goals? No, they did not, because we work on those based on our student data. While it's an important document, it does not drive what I do. What drives my daily practice is who are the students on my campus, who are the parents on my campus, what are the needs of our teachers, and what are our building goals and where are we in the midst of all of these stakeholders in getting to the place that we need to be.

James similarly explained that his awareness of the performance expectations from the district is not “anything that’s going to guide my practice . . . for me it really comes down to what I and my staff feel like the students’ greatest needs are.” He further explained that what drives his leadership practice is his “own understanding of what my school needs, what our kids need, what my teachers need, and what should be the strategic focus for the specific needs of the school at that time.” Robert similarly acknowledged that his leadership priorities were “really self-driven, as opposed to originating out of the performance evaluation.”

To summarize this finding, participants experienced performance evaluation under IDAPA 121 as an isolated and perfunctory end of the year event as opposed to a year-long process. Their evaluation event was disconnected from anything else they experienced or focused on during the course of the year. Contrary to the promise of IDAPA 121, their performance evaluation had no meaningful impact on their reflective practice, professional development or leadership practices. In the words of Edward, his evaluation under IDAPA 121 was both “inconsequential” and a “non-issue”, descriptors that could have been articulated by any of this study’s participants.

Finding 2: The student growth component of IDAPA 121, and the prospect of being evaluated partly on student growth had no impact on participants’ leadership focus or practices. Perhaps the biggest difference between IDAPA 121 and previous approaches to principal evaluation is that one-third of the principal’s evaluation would now be based on some measure of student performance. One of the primary goals of this study was to explore the effect of this change on the leadership focus and behaviors of Idaho principals. Would being measured on student growth motivate them to drive learning and student achievement, or cause them to work harder or differently in some specific way? A

second major finding of this study was the prospect of being measured in part on student performance had no impact on what participants cared about, what they focused on, or how they led their school. Three aspects of this finding are detailed below: (a) how the student growth component did not motivate or influence any aspect of the participants' leadership; (b) participants' concerns about the theoretical and procedural validity of the student growth component; (c) data availability challenges experienced by both districts that made it practically impossible for them to actually implement the student component during the 2014-2015 evaluation cycle.

Motivational influence or impact. Participants reported that being measured on student performance had no impact on their leadership focus or practice. William's perspective on this dynamic was representative of other participants. When asked if he had done anything different as a leader as a result of being partly measured on student achievement, he answered:

You know, I really don't think I have. I just have not worried about it. I just have continued to do what I think is the right thing to do. So I really haven't dwelled on it, and I haven't even thought about it.

Elizabeth reported the same. When she was asked if the student growth element of her evaluation had changed her leadership focus in any way she replied, "no, it's not changing the game at all. We've always been focused on performing well in the state assessments."

Likewise, Charlotte explained: "It didn't change anything for me on how I operate." She added, "wouldn't it be egocentric of me to say 'if my evaluation is going to reflect how my students do, I need to really push and drive' ... the motivation is not right."

Several participants emphasized that being evaluated on student performance had no impact on them because they were already motivated to help their students learn and perform. Nancy's explained her motivation:

We should be allowing students to make academic, social, emotional, and physical growth all year long, and I like doing that. But it's not because it's measured or the District tells me to -- that's an internal motivation I have and I think that's part of the reason why I like this job because I can help institute changes.

William explained his focus and thoughts about student achievement as follows:

I guess the bottom line for me is I don't focus on student achievement because I want a better evaluation. I focus on student achievement because it's the right thing to do for students. With that in mind, I feel like if I just do what's right every day -- everything else is going to take care of itself. So, if our achievement scores are not the highest then so be it. It won't be for lack of effort and it won't be for a lack of knowledge and understanding.

As the following exchange with Edward illustrates, student learning and achievement was already a central focus of her leadership. As such, being measured on these things had no impact on her or her leadership. She explained, "Everything we do tries to focus on improving learning." When asked if being graded on student growth had increased the pressure or focus for her in some way, she replied that it did not. "My ultimate goal is student achievement, that's what I'm in this job for ... and if that's not our focus, then I shouldn't be in that leadership role."

As these interview excerpts reveal, the prospect of being measured on student growth did not influence what these participants focused on or how they led their schools. In a

similar vein, not one of the ten participants indicated that any aspect of IDAPA 121, including the student growth component, had any impact on their school leadership. As William wryly observed, “student achievement has always been a priority for us ...the only difference now is it's a priority for my supervisor [chuckles].”

Concerns regarding validity of data. Participants expressed a number of concerns with using student test score data to measure the effectiveness of a school leader. These concerns fell into three primary categories. First, some participants believed that principal effectiveness was only one factor that impacted student learning and performance. They believed that important factors beyond the control of the school or its leader such as student motivation could impact student performance on standardized tests. Second, some participants were concerned with how student performance would be evaluated. They noted the difference between measuring students on ‘growth’ versus measuring students on ‘proficiency’, and that different approaches might be more or less fair depending on the demographics of a school’s students. Third, some participants questioned if the standardized tests were sufficiently stable to generate dependable and accurate results. These three categories are described more fully below.

William, Edward and Charlotte represent the first category of concern that questions if student performance on standardized tests can truly measure principal quality. As William explained:

You know, we can only influence student achievement to a point, because so much of it lies within the student, especially for students who don’t have much interest or desire in putting forth their best effort when it comes to assessment. Philosophically, I don’t know that I agree with using student achievement as an evaluating piece,

especially such a large percentage. Like I said, there is only so much that we can do to influence student achievement because it comes from within each student.

Edward raised similar questions and concerns as follows:

Looking at student achievement, do you look at that as a sign of a good leader? I don't, there's [sic] so many other factors involved that I don't really know what vehicle would be a good one ... because I look at all the things that occur that can impact student achievement and I think it's hard to measure that. How do you measure effective leadership? That's the ultimate question I think. I don't know.

Charlotte wondered the same question out loud as she was reviewing recently received student growth data on her students. She explained:

Did this particular grade level meet their growth target? No, not in this grade. Would I have liked them to? Well, clearly, yes. And should that be reflected on my evaluation? If they need it to be then do it, you know, but would it be a fair presentation of my leadership ability? That's the question I think that we need to decide.

Several participants expressed concern with how student performance was to be measured under IDAPA 121, noting the difference between measures of "student growth" and "proficiency." One participant, for example, who led a Title 1 school in District Beta, expressed a preference for using student growth to measure school principals. In the following excerpt he discussed the merits of a growth model versus the proficiency model that he describes below as a 'summative assessment.'

If it is clearly on student growth, I think there's some merit to it. The summative assessment that has typically been used in the past -- you know, way back when it

was the ITBS, and then the ISAT, and now the SBAC, if you're just focusing on a singular point in time, I think that's unfair, especially to Title I schools just because there's so many variables that play into our students at Title I schools. We're supposed to show that students are learning and if we can show growth, even if their end result does not meet the minimum proficiency standards, if we're showing growth, then it's got to be seen as a positive. I think that's fair, and I think most educators would agree that if given the opportunity to show growth, they'd much rather take the growth option over a summative option.

Another principal expressed a similar preference for student growth as opposed to proficiency. He explained, “Well, for us, student growth is what we're all about, so to me that's great. If you said we're basing it on student proficiency or something, our school [has a high percentage of at-risk students] so we pride ourselves on the growth that our students can show.”

Contrary to the perspective noted above, two other participants expressed concerns with the use of growth measures because their students were already performing at high levels on the State and district tests. Because of that, they preferred the use of a proficiency model to evaluate the effectiveness of their school and their leadership. One principal explained her position this way, “Growth data for [my] school is hard to show because when you're already [above 95%] proficient on the MAP for reading and [nearly at the same level] proficient on the MAP for math, when you're already that high it's really hard to show good growth for your kids ... I would like to see a proficiency mark.” A second principal described the same challenge as follows:

Can you continue to show a good amount of growth over, and over, and over? Or is there a ceiling? We think there might be a ceiling, particularly with kids at the upper levels. So when you have [high performing students] like we do, the higher those kids start the less growth that they can truly achieve. So it's a challenge for us.

The third category of concerns regarding the use of student performance data related to questions about the stability and reliability of the standardized test(s) themselves. The perspectives of Scott, Edward, and James illustrate these concerns. Scott explained the following:

There are an awful lot of question marks in my mind, and actually in most educators' minds in terms of the relevance or the validity of this new test. It will take us, in my mind three to five years before we could actually say okay, now we have some valuable data that we can build on.

John expressed similar concerns, raising a number of questions about the reliability and suitability of the standardized testing data for evaluating purposes. She shared the following:

I think looking at data is super important. Do I think it needs to be a piece of that? Sure. How much? I think there's too many factors involved. We have a new measurement tool, yet all the bugs haven't been worked out and what the outcome looks like, we don't know. We all know that the outcome doesn't look very good so far, the scores are pretty low. Is that a reflection as me as a leader? I would hope not, so it's hard to say. We're measuring different students each year and I think that's the trickiest part of all of this is, you're not comparing those students. Do I take a seventh grader and to hope that they have improved by ninth grade? That's a fair assessment. But to take just all new students coming in, that's not fair because much of their

performance has to do with their elementary school teachers. So, I think that looking at data should be a piece, but I don't know how much it should weigh on that because there's so many other factors involved.

James expressed a similar point of view, reflecting on the historical instability of the standardized tests and questioning their suitability to inform high stakes personnel decisions about school leaders. He asserted:

So if they're going to take -- if they want to base salary on how we can grow on this particular test, then they're going to need to get the test solid enough where they're not making changes to it where we can actually start understanding the test, and then setting goals towards it. I think that it's a very scary model for me to have that kind of pressure put on tests, because I feel like that people make bad decisions under pressure.”

Data availability challenges. Under IDAPA 121, the 2014-2015 school year marked the first time that District Alpha and District Beta would be evaluating their principals on measures of student performance. As the participants revealed however, none of them were actually measured on student performance because the standardized test data upon which they would be measured was not available by the time of their evaluations. Charlotte’s description of the situation in her District Alpha explains it well:

We have to have our teacher evaluations done, signed, ready to turn in on May 1st. Principal evaluations, administrator evaluations are, I think, two or three weeks after that. During that window of time we still don't have any growth measure back, we have no student data back to even add into a teacher's evaluation or to a principal's evaluation.

When Nancy, also from District Alpha, was asked if she was evaluated on the basis of student performance, she reported the same challenge as Charlotte:

That's hard because I'm still getting student data. I mean, my evaluation is final, but my data isn't. I just received my math data today, you know, long after my evaluation was done and ISAT data still isn't out. The National Consortium is way behind on scoring and so that probably will be done by the end of June, we're being told. So, if they want that summative data included, they need to decide how they're going to get us that data, in time for the evaluation process.

The same problem with data availability occurred in District Beta, as Robert explained: Well, we take the tests in April and May, and as of yet -- and then now I'm talking about the summative assessments ... we haven't seen the results yet, and we're in summer. Teachers are gone.”

Robert further elaborated, “So, yeah, I think that's probably the big flaw, we don't have that summative information quickly.” Scott agreed, explaining that:

The new standardized tests we're taking, the SBAC, ISAT 2.0, we don't have results out yet. And in fact we were supposed to be able to send those out the last week of school, and the state is having significant problems with Smarter Balanced getting those results in. They're still not completed, and so here we are in the middle of June and we still don't have student performance data.”

Given this challenge, a representative from District Beta informed me that their district had decided to deem all of its principals ‘proficient’ on this aspect of the evaluation for the 2014-2015 school year unless there was compelling evidence that would suggest otherwise. The district representative did not explain what forms such evidence might take. District Alpha

appears to have taken a similar approach in bypassing the student performance component of the evaluation. When James was asked if he had been evaluated on student performance, he replied, "...I wasn't at all because ... the data wasn't there yet ... we won't even see those numbers until the fall; that is what the state has said."

To summarize this second finding, using student test data did not motivate or influence higher levels of commitment and performance by these participants.

Philosophically, participants reported that evaluating them on student performance had no impact on their own motivation or focus because they were already fully committed to student learning and performance. Operationally, participants expressed serious concerns with the validity of the current testing data and if a 'one size fits all' approach to measuring student growth was appropriate and fair, especially for higher performing schools.

Practically speaking, even assuming the test data was valid and reliable, problems with data availability led to a systemic breakdown in this element of IDAPA 121 such that principals were not actually measured on student performance in the 2014-2015 school year.

Finding 3: Participants believe that policy makers and school policy related to their performance evaluation are out of touch with what happens in schools. When reflecting on the validity and usefulness of the new evaluation model embodied in IDAPA 121, all participants expressed the opinion that the people responsible for the development of public school policy such as IDAPA 121 are out of touch with the reality of what happens in the school building. As Elizabeth commented:

I think there's a real disconnect in what actually happens in schools and what legislatures believe happens in schools, so I don't know that we'll ever make that connection to be honest with you. And until we do we won't have clear targets as

administrators because there's not the knowledge of what's really going on in the buildings. I don't know if it's possible. I think legislatures need to come and spend a day in the shoes of principals to see what really happens within the building.

Scott shared a similar observation:

It's fascinating to me how these [policy] decisions are made in the absence of -- I mean they're outside of the building. In fact, in most cases, they're not even involved with the school district. They're legislators, so they're business owners, or in our state, the vast majority of them are farmers. It fascinates me how people outside of the industry are so inclined to make judgments about how we should be measured. Oftentimes without seeking feedback from those people.

Other participants described policy makers as viewing schools through the lens of “factory” or “business”. Robert, for example, characterized them as wrongly viewing school personnel as “robots” and learning as a factory process that can be measured and scored with precision. He explained:

And my opinion is, and I think this might be the opinion of some of our District leaders is that some of those folks are far enough removed that they don't understand all of the variables that play into what's going on in a school every day. They see it as some kind of factory or a robotic system, that hey, if we just do “this”, then this will spit out the right numbers for us. They are missing that human factor. For example, they have no idea how many times I had to call the Department of Health and Welfare about a concern we have about a student.

Charlotte concurred with Robert's assessment, using the business language of sales, earnings and quarterly dividends that are easier to measure than what happens in a school. She explained:

That's the other thing we do in education, is we want that business model which is, you know, we have to have our quarterly dividends in, and they have to look good. Our sales and earnings have to look really, really good. And with people, we don't have quarterly dividends, and good earnings, and sales. It sometimes takes a little longer. I mean we're turning around learning deficits, and turning around lack of motivation, turning around a lack of support and finances.

Other participants questioned the policy assumptions that appeared to underlie specific details of IDAPA 121. For example, and as described earlier, several participants questioned the assumption that evaluating them on student growth would cause him to focus on teaching or student learning in a new and more impactful way. These school leaders did not focus on test scores because they wanted a better evaluation; they focused on student achievement because it was the right thing to do. Edward questioned policy makers' assumptions with even stronger language, as follows:

What's the real intent of the evaluations? What's the belief behind it? That we (principals) suck and that we need help doing what we're doing because we're failing the children of our great state? Because we're failing the children of our communities? I think even some of the assumptions that these evaluation tools are built on are really off-base; it's hard to build a good tool off of an assumption that's really not true that principals really don't care or teachers maybe don't care if students achieve.

James added to these criticisms of IDAPA 121 and its underlying assumptions with an eloquent critique of the underlying model and process of the current evaluation system. He explained:

Rather than starting with the relationship between an evaluator and a person, and how to effectively help that person grow, rather than saying what is that person doing, or should be doing, and then making the evaluation tool. I think that we went around -- about this process all wrong, and I think that's why it's so huge, and that's why it's not very effective and I think, that we're ignoring the human aspect of this, and we're trying to get it into some kind of a cold, hard fact that then will be able to stand up to scrutiny of having it linked to pay, and all the problems that go with that. And we're losing the things that really make a difference, which is a connection between people and support for getting better at what we do.

Finding 4: Participants had clear ideas for improving the system so that principals could experience more meaningful reflective practice and professional development. Research question number five sought to understand what circumstances or experiences would contribute most significantly to participants' reflective practice and professional development. After completing two pilot interviews and the first regular interview with William, however, it was appearing that evaluation under IDAPA 121 might have little impact on reflective practice and professional development.. Upon consulting with my advisor, and based on his recommendation, the interview protocol was expanded to include a new interview question that asked "if the current system is not working well, what would work better?" What poured forth from these conversations was a rich and lively exploration of what might work better and be more effective for school leaders, both

experienced and new to the role. The participants seemed genuinely pleased that someone wanted to know their thoughts on this, and for the most part, they had clear and ready answers to my inquiries. This final finding describes their perspectives, opinions and recommendations on how to improve the school leader's experience with evaluation, reflective practice and professional development. Section one focuses on how to improve implementation of IDAPA 121 and section two focuses on how to improve professional development for school leaders.

Implementation improvements. The participants offered numerous recommendations for how to improve the implementation of principal evaluation under a system like IDAPA 121. Nearly all of these suggestions focused on the professional standards component of this evaluation model and few suggestions addressed the use of student test data beyond the earlier described concerns with the reliability and validity of the current tests in Finding # 2. Their suggestions can be organized into three main categories: (a) recommendation for principals to meet at start of the year to review performance expectations, set goals, and drive more relevant reflective practice; (b) collection of genuine performance data that is based on actual and more frequent observations; (c) the establishment of a safe and trusting relationship between principal and evaluator.

Start of year goal setting. Participants focused on the fact that under the current system, most principals do not review or even think about the standards of effective school principals until the school year is over and they are preparing for their end of year appraisal meeting. William for example, admitted to not thinking about his evaluation until the day in May that he was recruited to participate in this study. Nancy similarly admitted "I hadn't even looked at it until my conference." To remedy this, participants recommended that

principals and their evaluator meet at the beginning of the year, in a pre-conference, to establish what Robert called a “clear orientation to the standards.” Robert explained the rationale for this recommendation as follows:

If anybody in a learning situation is made aware of what the standards are ... what the expectations are, then you've got something you can work with. It would be unfair, I think, even though somebody could say well, the person could read what their evaluation tool looks like and they can then know what's expected. I think that, I mean ultimately in a legal sense somebody could argue that and say nothing else needs to be done, but in a practical sense, I think a good strong program is going to be one where ... new administrators are oriented to what the evaluation tool is, and what the expectations would be ... If there's a plan, and new administrators are – and veteran administrators are oriented to, uh, the tool, and what the purpose and process -- processes connected to it are all about. I think that's a good thing.

Elizabeth also recommended up-front alignment at the start of the year so that expectations were clearly understood and to facilitate effective goal setting. When asked to describe what an ideal process might look like she shared the following:

[I]f I were a director and I were evaluating administrators I would set some clear targets for my expectations on what I felt was proficient, and then I would set some clear targets on what I wanted to see if they were distinguished. And I would probably link them with the language here, but put those targets in more layman's terms that reflected actual practices that I know the administrators are going to be putting into place. And I would set those goals up front and ask for that data along the way rather than just at the end.

Edward also emphasized the benefits of effective goal setting at the beginning of the year, with opportunities for reflective practice, a mid-year review and end of the year review. He explained:

I would include, at least as a launching point, some type of goal setting and reflection for principals, because the growth mindset and the growth model are so important to me and I think it's important to all of us. I think if you had a conversation at the beginning, a middle, and an end check in I think that would be reasonable given the other things we're trying to do.

James also saw potential in clarifying expectations at the start of the year to ensure everyone was clear on how the process would operate and to facilitate meaningful self-reflection. He explained:

I think that if we can train evaluators to work as a coach, have people set up their goals, show them how they're going to evaluate them on their goals so that everybody's clear in the process, everybody's clear on what does distinguished mean, what does proficient mean, and what does basic mean, and then have them look at themselves, I think you're going to get much better growth, and you're not going to get a horse and pony show which, you know, is what this evaluation can tend to be.

More meaningful performance data and feedback. Many participants expressed a desire to receive more feedback on their performance than they currently receive, and they emphasized that for this information to be credible and useful it must come from the first hand observations of people who actually know them and their practice. There was a general feeling among participants that in the current system, their evaluator lacks the time to observe their leadership practice directly or get to know them on a close level.

Scott emphasized these points, and the value of more frequent and first hand feedback, by contrasting his own experience as a school leader with that of classroom teachers. He explained:

Well, I think, the biggest difference between how we evaluate our teachers versus the way that we are evaluated (as school leaders) is the level of presence, and the fact that I try to get into every single classroom at least once per week if not twice per week. And so not only do I have a much better picture of what's happening in that classroom, but I'm also able to make incremental changes or suggestions as trends start to appear; a lot of those things can be midcourse corrections that don't even require a tool of any kind. They can be discussions in the faculty room. They can be informal conversations in the classroom after the kids have gone home. In a perfect world that's what it would look like for administrators, but I don't know how we would function to make that happen.

Robert shared his own perspective, explaining that an evaluator should have 'boots on the ground' information, and noting that his own evaluator was not able to obtain that in his case.

I haven't seen this in practice very often, and again I might be an idealist -- an idealist on this in that I mean for a supervisor, it's really, you know, be objective in their evaluation, they've got to have boots on the ground information, right? And it may just be because I'm a veteran administrator. My supervisor does not come to my building. He doesn't participate in anything I do here. I've run I don't know how many IEP meetings this year, I probably was a part of 60 to maybe up to 80 IEP meetings this year. A staff meeting once a month, problem solving team meetings weekly, all the evaluations that I do with teachers, all the parent meetings that I'm a

part of. And I know he's got a big area, but it would seem to me that to really be authentic, the supervisor would need to be in the building and have some sort of a method to evaluate the administrator.

Nancy shared a similar perspective, and summarized this desire for more informed feedback by someone who really knows her as follows:

They have to know me, as a person, and they have to know what I do professionally, to be able to help me. Maybe it's harder at the elementary level because we are the only administrator within the building. Middle schools, high schools have teams of administration where you see each other working on a daily basis. And so maybe you can reflect with your team because they know you and they know your practice as an administrator. At elementary, there's no one that knows us as an administrator within the building that's in that administrative team. And so I think you have to do your own reflection and seek input from your staff to help. I don't fault our directors for not being in our building more, because they are so overwhelmed and overworked at their jobs just like we are. They don't have the time to really be in the buildings and do that and so I don't know how to fix the system. How to get somebody in the building to see what you're doing.

Mary offered a partial solution to this challenge, observing that the people in her school building who know her and her leadership practices are in a better position to provide meaningful feedback. She shared the following:

We're also evaluated by our staff too. We get access to that. They have an online survey they can fill out for us. And to me, I almost look at that as more valuable than what the District gives me, because it's specific to me at this school and these are the

people that I see 180 days of the year. The District -- I don't see them or talk to them that often and they don't know the ins and outs of everything going on like the teachers do.

Safe and trusting relationship. Several participants emphasized that having a safe, positive and trusting relationship between the principal and his/her evaluator was essential for the evaluation process to work properly. The key concept for them was 'relationship' which James described as "where the evaluations have their hope" adding, "evaluations don't have their hope in the document themselves." One participant, for example, described the 2014-2105 evaluation as a positive experience because the relationship with the evaluator was strong. This contrasted with a prior evaluation process that was not positive because the principal did not trust the motives of a different evaluator. James concurred, asserting a principal needs to feel safe within the relationship to speak candidly with his/her evaluator about challenges and development opportunities. He explained:

... it's going to take a much different approach, and we're going to have to give principals the support to actually be open and say, you know, this is an area that I'm really struggling right now and know that the evaluator is going to say okay, well let me see how I can help you grow doing this, not say, wow, you're a really bad principal and I'm going to do my best to get you fired.

Nancy shared a similar perspective. She observed, "if it's an open, honest conversation [with my evaluator] I'm much more open maybe about asking for feedback or asking for constructive criticism, but I don't want him to misinterpret something and think I'm not proficient in something because I'm asking." Nancy further observed that the current "process of evaluation itself doesn't provide enough opportunities for her evaluator to

provide really helpful input, and timely input, informed input, to help develop that trust relationship to make it more of a coaching environment.”

Professional development improvement. Participants shared many interesting ideas about how to improve the formal, district or state level approach to professional development of school leaders. This section describes the following three most frequently recurring recommendations: (a) enhanced and more meaningful opportunities for peer collaboration, (b) formal mentoring programs, and (c) a focused and high quality formal training.

Peer collaboration. Participants from both districts identified peer collaboration as one important component of an improved system of professional development for school leaders. Within District Alpha, participants reported that they currently have no formal means to discuss common problems or share solutions with school leaders from other schools within the district. Within District Beta, by contrast, school leaders are provided opportunities for peer collaboration. That said, Scott explained that the current model of peer collaboration within District Beta is not very effective, and he hoped it would be possible to return to the earlier model of peer collaboration that worked better. This section details participants’ recommendations for non-evaluative peer collaboration.

Nancy began by describing some positive aspects of district sponsored meetings that she does attend. She said:

Our District does a pretty good job of having monthly meetings where they focus on a current area. I went to human trafficking meeting this morning, which is an issue that we need to be aware of as educators because it's in the world and coming to our community whether we want it to or not. Our District does a pretty good job of providing point of need, small professional development like this.

As she reflected on the nature of these meetings however, and the absence of any real opportunity for discussion or dialogue amongst her peers, she quickly identified the value of having time for peer collaboration. She explained:

When we meet, in our District, it's all just, throw stuff at us and then we come back to our building with a pile we need to figure out and go through. It doesn't seem like anybody has the luxury of time to actually have conversations, or to have input, or to discuss things. I think that would help a lot if, at administrative meetings, you actually had time to speak to each other, and have input, and discuss things, and bounce ideas off each other. Because I really don't know what's going on in other buildings around me because I don't have time to be in their building or I don't make time to be in their building, I just know my building.

When asked, “When is your own collaboration time, as a leader”? Nancy responded, “There is none. So, it would be, you know, during those monthly meetings even if we could just have time to speak to each other it would be nice.” Nancy further reported this was a request that “had been made before, but never been acted on.”

Elizabeth echoed Nancy’s call for structured opportunities for peer collaboration as she reflected on the wealth of knowledge that experienced school leaders could be sharing with one another. She explained:

I really see value and opportunities for us to have structured collaboration more so than trainings per se. We know what's going on in the building and what the needs are, and so there is an opportunity here to model something off the ‘Train the Trainer’ programs. Teachers training teachers. Should be administrators training administrators.

When asked what this model of structured collaboration would look like, Elizabeth explained that it would be about having “given goals and . . . conversations about how to achieve those goals . . . like a professional learning community for school leaders.” Like Nancy, she recognized that she did interact with other school leaders at district wide meetings, but that those events did not provide true opportunities for discussion, and she lamented this lost opportunity to learn from her peers and vice-versa. She explained:

We have principal meetings now, but we don’t often get an opportunity to really collaborate with each other. And especially if we were given clear goals like if our goal was, oh, the integration of technology into the buildings. I mean that's our topic, what are we going to do as leaders to help make that happen, have a conversation that's guided. I think that would be very effective, because there's a lot of really amazing leaders out there that have great ideas and we never talk to each other, ever.

Based on these recommendations from Nancy and Elizabeth, Charlotte was asked if a better system of professional development “would include some type of peer collaboration?” Charlotte responded immediately and clearly, “Oh, yeah, it would have to”. In explaining how such collaboration might work, Charlotte emphasized the value of school leaders coming together to share information and ideas on a topic of common interest. She explained:

I am probably not the only one that would choose -- If I'm just randomly picking -- ‘collaborative leadership’, and those who have also selected this for their domain for the year, would work on that together. And that's where when we would come back together, I mean we have the benefit now of, you know, online chats, and Edmodo, and you know, sharing, and all of that. So, I think, that would be a big piece of it.

And then come back together, and ‘this is what I did, these are my results, this is what I did, this is what I struggled with, this is what I did and it was great. You know, this worked so well’ ... So, I think, that's the collaborative piece, just having that trust with your colleagues to say ‘what are you doing that I'm not doing, because yours sounds way better than mine did. I like your agenda better than mine’, whatever it is. When asked if structured opportunities for this type of collaboration occurred on a regular basis today, Charlotte answered, “No ... Not formally.” Although, she added, “It does happen informally.”

Scott, who worked in District Beta, shared the same high opinion of peer collaboration as expressed by Nancy, Elizabeth and Charlotte. As he explained it, peer collaboration might be a more powerful tool for leadership improvement than performance evaluations like IDAPA 121. He explained:

If our ultimate goal is improvement, I don't know that you need an evaluation per se to be able to do that. Maybe you work together collaboratively with a team of your colleagues and/or a supervisor so that you could continue to set your own goals and work toward achieving those. That would certainly be a lot more nimble than how we do it today.

Although Scott is a seasoned school leader, he “desires to continue to grow and develop” and would appreciate receiving “some support, someone to talk to, someone to bounce ideas off of.” As Scott reflected on this, he acknowledged that his district currently provides “scheduled meetings every other month for collaboration ... arranged in heterogeneous groupings of principals” that include “elementary schools that are Title I’s, some that are non-titled, some middle school principals, and then some high school.” In the same breath,

however, Scott quickly stated that the current peer collaboration model, while 'helpful', is not as effective as an earlier model that the district used three years earlier. He stated, "If I had one wish for the day, it would be to go back to that previous model." Here then, is Scott's description for how the ideal peer collaboration model would operate, based on the most important features of the prior model:

1. Homogenous grouping of participants: Scott understood why his district made the move towards heterogeneous principal groupings. He explained, "you can see how the decisions you're making in your building are affecting the kids as the move through the different levels." While this was helpful and made sense for an annual planning session, Scott said homogenous groupings would be better for the peer collaboration working sessions.
2. Monthly meetings: Under the prior model, the district's peer collaboration meetings occurred monthly. Currently, the meetings occur every other month and Scott believed monthly meetings were better and going back to monthly would be "part of his wish."
3. Principals only: Scott explained that another "piece that changed, in addition to heterogeneous grouping, now there's always one of our supervisors present [during the peer collaboration meeting]". While Scott understood the intent of this change was to allow more streamlined problem solving with a district leader present in the discussion, "the unintended consequences was it took out that anonymity and people weren't as free to speak sometimes for fear of exposing, their deficiencies, or trying to protect a teacher, or colleague; it's changed the flavor of those meetings an awful lot." For these reasons, Scott

would like to restore the original model wherein only principals were present in the meetings, and if they had an issue that required district input or support, they could “meet with the supervisors after the meeting.”

4. Set their own agenda: Scott observed that “lately, those [peer collaboration] meetings have come with a preset agenda. While I understand their intent at the time, looking back, when we set our own agenda that seemed to be the most effective for us in the sense that we knew what our issues were, we knew what challenges we had, what questions we had, and we were able to have open and confidential conversations.”
5. Officially sanctioned meeting times: After the district changed the model of the peer collaboration meetings, some principals asked the district to restore the prior model. Scott reported that district declined to make the changes, but did allow principals to “meet any time we wished with whomever we wished on our own time” as one way for the principals to recreate the experience of the prior model. Unfortunately, and although “a dozen of us said, well we’ve got the green light, let’s go ahead and start our own advisory groups”, these never got off the ground “because of conflicting schedules”. Scott explained, “like it or not when the district office says you will meet on this day at this time you can put that on your calendar and bank on it.” For this reason, it would be important for these peer collaboration meetings to occur at officially sanctioned and scheduled times.

Mentoring. In addition to peer collaboration, several participants highlighted

mentoring as an important element of an improved system of leadership development for school principals. Edward, for example, thought that having a structured mentoring program “would be great.” She observed that with a formal, structured program, “it’s easier for me to call and ask specific questions.” She admitted that in the absence of a formal relationship of this nature:

Sometimes you’ll look at something [that you have questions about or that you want to improve in yourself] and say, ‘yeah, ok, I’ll work on that’ but then it’s gone, forgotten. But if there was a formal process you had to go through, you would know that it’s valued and it’s what you’re supposed to be doing.

Nancy shared a similarly positive view on administrator mentoring, although she focused more on the benefits to a newer school leader. She explained:

I think if we assigned mentors to administrators, just like we do for teachers now again, it would be a wonderful relationship because you would have someone you could rely on, or call, even just sharing the forms you use. If I had a mentor I would have loved that and I think I would have been a better administrator sooner. But there is nothing out there that I know of, or at least in our district for new administrators. It’s just ‘here you go.’

Nancy further explained that a strong mentoring program would provide new principals a first year mentor that you could remain in relationship with “for the first three years”, and after that “you’d be helping mentor [other new administrators]. You would get better at your job because you were mentoring someone ... and they would help invigorate you also.”

Robert and Scott also identified mentoring as important for the professional development of school leaders. Robert explained, “I think a good strong program is going to

be one where there is some mentoring going on”. Scott agreed, identifying “certainly a mentor” as something that new principals would benefit from, and possibly also something that more seasoned administrators might benefit from as well. He observed:

In our district we have a lot of movement where even folks who have been around for a while jump systems. So they may go from high school down to elementary. And even though they’ve been around, you know, their network of support isn’t going to have a clue about what’s happening in that new environment.”

Scott further explained the benefits of a mentoring relationship this way:

I think, that helps a lot because then they can check in as they're going through. Being an experienced principal, you know what those milestones are throughout the year, things you need to be doing. And a quick phone call or an email to say hey, you know, I'm getting ready to do this, have you done yours yet? Do you want me to walk through it with you? A confidante that they can call if they've got some challenges, particularly with staff, because they need -- that one's a sticky one really, because until you've been in the District long enough and kind of know who it's safe to talk to, I think, it's easy to flounder around out there on your own and try to make those decisions in a vacuum without knowing who to turn to for help on a number of different issues.

In addition to these benefits, Scott also reflected that having a ‘non-evaluative’ mentor relationship that provided a safe and non-judgmental environment was important as well. This would provide the principal with “somebody that I can call if I get stuck, and I can be vulnerable, and I can say ... as opposed to calling my immediate supervisor and saying ... I

know you told me this last month, but I can't remember how I was supposed to do it. I think that would be helpful.”

Focused and high quality training. In addition to peer collaboration and mentoring, several participants identified formal training as another important element within an improved system of professional development for school leaders. As Edward explained it, formal, technical training would be especially valuable to him if delivered on topics that did not come as naturally to him or that were not as intrinsically motivating. He elaborated:

Trying to create a good healthy work place. That I can do on my own, but areas like to how use data to drive school improvement and instruction and some of those other areas that would really take some good training. I haven't experienced that anywhere in any of the districts where I've been an administrator.

Scott, who is from also from District Beta, noted that:

For a few years, under a different superintendent, we had quite a lot of directed professional development on leadership theory, and leadership styles, and a lot of team building activities that were done as an administrator group, but we haven't seen that now for a few years ... whether [newer principals] are getting something else from the District that I'm not aware of, I can't attest to, but I know we're not doing it as a whole group.”

Scott acknowledged that the recent fiscal crisis likely drove the elimination of this structured professional development as “professional development is often one of the first things that go when the district needs to trim some fat.”

Elizabeth, who is from District Alpha, shared her District Beta colleagues' perspective on both the value of structured training and the fact that it is currently lacking in her district. She explained:

I think a piece that we're lacking at the District office is really meaningful professional development. I can't tell you how many trainings I've been to on 'our computer system and I don't need another training on that. We've been using it for years. I would like to see our professional development plan differentiated based on your needs as a building, and I think a focus needs to be on administrator's instructional leadership because I don't think there's enough of that that happens. And I don't think there's any training at the District level on what instructional leadership looks like and what climate building in a building looks like.

Charlotte also emphasized the value of professional development through training and offered the most detailed prescription for how to improve training for school leaders. Her recommendation included the following elements:

1. Narrowly targeted and personalized focus: Although Charlotte believed it important for school leaders to be evaluated on all the aspects of their leadership, she "would like to be able to say 'choose any one of these broad domains, such as instructional leadership, and really focus on that this year'. What is it that you want to do based on your reflective practice, based on your data, based on the input that you received? Pick a domain and really focus on that, and then get as much information, as much training as you can in that particular area. Then the following year I would want them to select something else."

2. High quality training that is ready to go: Implicit in Charlotte's first point is the expectation that if "they have said 'ok, if you're going to really work on collaborative leadership this year', that they have training ready to go that represents current best practices, what's new and emerging in the field. It would be important to really share with them the best, latest information."

3. Learning checkpoints: As a final element of her model, Charlotte believed it would be important for school leaders to have "checkpoints along the way" that would define for leaders "here's where we are, this is the information, here is your task or your assignment, your focus for the next month or six weeks, whatever it is. We're going to come back together, find out where we are, make adjustments, go back and work on this next aspect of this for another whatever window of time that would be."

In summary, these proposals from Charlotte and earlier comments from other participants illustrate how a highly targeted and personalized professional development could operate, along with peer collaboration and mentoring, to improve the leadership capabilities of school leaders.

Summary

This chapter presented the four major findings of this study. The first finding was participants experienced evaluation under IDAPA 121 as a perfunctory, isolated end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice. The second major finding was the student growth component of IDAPA 121, and the prospect of being evaluated partly on student growth had no impact on the participants' leadership focus or practices. The third major finding was participants believe

that policy makers and school policy related to their performance evaluation are out of touch with what happens in schools. The fourth major finding was participants had clear ideas for improving the system so principals would experience more meaningful reflective practice and professional development. These findings were consistent across participants from both school levels and both districts and supported by direct quotations taken from participants' interview transcripts.

Chapter Five

Interpretation of Participants' Experience with IDAPA 121

Introduction

IDAPA 121 was designed, among other things, to enable Idaho to qualify for the U.S. Department of Education's ESEA Waiver Program which offered flexibility from certain provisions of NCLB. This program required Idaho to adopt a principal evaluation system that could "be used to inform personnel decisions", thus serving the 'summative function' of traditional evaluations (U.S. Department of Education, 2012; Portin et al., 2009, p. 28), while also serving the formative goal of "provid[ing] clear, timely and useful feedback ... that guides professional development" (U.S. Department of Education, 2012). Idaho's State Department of Education asserted IDAPA 121 would fulfil this second, formative objective by "promot[ing] reflective practice and the development of ongoing personalized development plans" (Idaho State Department of Education, 2013, p. 200). Additionally, the federal ESEA Waiver Program required Idaho to evaluate principal performance not only on "measures of professional practice" but also on "data on student growth for" that must be a "significant factor" in the evaluation (U.S. Department of Education, 2012). Measuring principals on student growth is a relatively recent phenomenon nationally, and IDAPA 121 marked the first time that Idaho would be measuring its principals in this way.

The purpose of this study was to describe how well the "promise" of IDAPA 121 compared with its actual impact on participants' reflective practice, professional development, and leadership practices. Research questions 1 and 3 were directed at understanding if IDAPA 121 achieved its formative objective of promoting reflective practice and professional development of Idaho's public school principals. Research

questions 2 and 3 sought to understand the impact of the “student growth component” of IDAPA 121 on principal reflective practice, professional development, and leadership practices. Data were collected from semi-structured interviews of ten active public school principals in District Alpha and District Beta. Data were analyzed using a variety of qualitative analytic techniques including simultaneous collection and analysis, coding, creation of matrix and cross-case analysis, jotting, and analytic memoing.

The first major finding of this study revealed that evaluation under IDAPA 121 was an isolated and perfunctory end of the year event that had no impact on the participants’ reflective practice, professional development or leadership practices. Similarly, the second finding showed that contrary to those who advocated for evaluating principals on measures of student performance, the prospect of being measured under IDAPA’s ‘student growth’ element had no impact on the leadership focus or practices of the research participants. The third finding revealed participants believe policy makers who developed IDAPA 121 are out of touch with what really happens in schools. The fourth finding showed that participants had clear ideas about how to improve the professional development of school leaders, and many of their ideas can operate independently of performance evaluation models like IDAPA 121. These findings may be surprising to those who advocated for revised models of performance evaluation like IDAPA 121. Because these findings run counter to current policy in Idaho and in many other states, further analysis, interpretation and synthesis of these results is warranted.

Chapter Four provided descriptive findings related to participants’ reported experience with IDAPA 121, reflective practice, and professional development. This chapter provides interpretation that brings meaning and coherence to those findings, “offering

explanations, drawing conclusions, extrapolating lessons, making inferences, and otherwise imposing order” (Marshall and Rossman, 2016, p, 228). As Miles, Huberman, and Saldaña (2014) remind us, “research never proves anything; at best, it *suggests* (p. 223, emphasis in original). What follows are merely *possible* explanations suggested by a synthesis of the research literature and the reported experiences of participants. This chapter’s interpretive analysis focuses on the following analytic themes:

1. The relationship between the ‘professional practice standards’ of IDAPA 121 and participants’ reflective practice and professional development (Research Question 1)
2. The relationship between IDAPA 121 and participants’ leadership practices (Research Question 3)
3. The relationship between the prospect of being measured on student growth, and participants reflective practice and professional development (Research Question 2)
4. Participants’ perceptions about how to improve professional development for principals (Research Question 5)

Analytic theme one: The relationship between the ‘professional practice standards’ of IDAPA 121 and participants’ reflective practice, professional development, and leadership practice

Prior to the advent of the modern, standards-based performance evaluation like IDAPA 121, principals viewed performance evaluations as “perfunctory, [and] having limited value for feedback, professional development, or accountability to school improvement” (Clifford & Ross, 2011, p. 2; Portin, et al., 2006). Advocates of standards-

based evaluation like IDAPA 121 asserted the evaluation could serve not only the traditional ‘summative’ accountability purpose of evaluations, but also a ‘formative’ purpose that informed professional development and improvement of leadership practices (Goldring, et al., 2007). Against that backdrop, this study’s first research question sought to understand if, and to what extent, the professional practice standards of IDAPA 121 delivered this formative impact on participants’ reflective practice and professional development. Despite the “great promise” (Goldring, et al., 2007, p. 1) and “great potential” (Clifford & Ross, 2011, p. 2) suggested by the literature, participants reported their evaluation and the professional practice standards of IDAPA 121 had no formative impact on their reflective practice or professional development.

Edward’s experience with reflective practice was representative. He explained: “[W]hen I was reflective was I thinking, I better ...look at this because it's on my evaluation? No, I didn't... Looking at my evaluation was never what drove anything I did. Meaning that, oh, it's on my evaluation, I better look at it. I would look at things that I thought were important.” Regarding professional development, Nancy reported “the current evaluation system experienced this year had no bearing, positive or negative, on [my] professional development.” Edward concurred, explaining that “To my development as a principal or my own goals that I have for my leadership, it [my performance evaluation] is a non-issue.” Contrary to the hype surrounding the potential for standards-based evaluations to positively impact professional development of school principals, participants continued to experience performance evaluation as perfunctory, and having limited value for feedback or professional development (Portin, et al., 2006).

It appeared, at least for participants, that truly nothing had changed compared with how evaluations were experienced prior to the advent of the modern, standards-based evaluation (Clifford & Ross, 2011). Several possible explanations can be proffered for IDAPA 121's failure to achieve its formative objective for participants. Several operational problems with IDAPA 121, as implemented in District Alpha and Beta, are suggested by participants' experiences and relevant research literature, and may account for the failure of IDAPA 121 to deliver a formative impact on participants' reflective practice and professional development. These are explored below, followed by an analysis of potentially deeper, substantive explanations.

Operational challenges. First, as implemented in District Alpha and District Beta, evaluation was not linked in any systematic way to participants' professional development. The Wallace Foundation (2009) report acknowledged that a modern, standards-based evaluation was "one element of a broad systems challenge that requires state and district policies that promote better training and overall support for leaders" (p. 11). In order to have the desired effect, evaluation processes must "lead to appropriate professional development that addresses any weaknesses or concerns identified by the assessment process" (Wallace Foundation, 2009, p. 8; Guilfoyle, 2013, p. 1). Notwithstanding these recommendations from the literature, IDAPA 121 as experienced by participants, was disconnected completely from any formal or informal professional development processes or programs. During the 2014-2015 school year, participants pursued their own, individually driven professional development plans that had nothing to do with the evaluation process. Participants did not focus on the leadership standards of IDAPA 121 until the end of the school year when preparing for their own end of year evaluation conversation. As such, their own professional

development, to the extent it was pursued during the year, was driven by their own interests and self-defined needs. This is represented by the experience of Edward who reported that his professional development was ‘completely disconnected from the evaluation system’ and that ‘whatever goals I set for myself I just set them and my evaluator has no idea what they are really.’ This lack of connection between evaluation and professional development continued through the end of the year when participants sat down and discussed their performance with their evaluator. Even then, the evaluation conversations did not lead to formal or informal development plans or goals.

Second, and closely related to the first explanation above, participants did not receive timely or meaningful performance feedback during the course of the year. Timely and trustworthy feedback is widely viewed as an essential precursor to reflective practice and professional development (Goleman, Boyatzis, & McKee, 2002; Clifford & Ross, 2011, p. 2). Feedback on practice may be especially important for public school principals because “they have few sources of trusted feedback on their practice and commonly feel isolated from colleagues due to the rigors of their position” (Clifford & Ross, 2011, p 2). Despite the importance of timely feedback, participants reported receiving little to no performance feedback from their supervisor during the course of the school year. Nancy’s experience in this regard was fairly typical. As she explained, her supervisor came to her building three times before the end of year evaluation meeting, but none of those meetings entailed observation or feedback. Although she did not fault her supervisor for this, she noted that “someone coming in to evaluate me who's been in my building an hour out of the year really doesn't know what I do”, and is not in a position to provide helpful feedback. Another participant from District Beta reported a similar experience

We are fortunate to have our supervisors drop in periodically, but they're really more about – I don't know, I hate to say it – but more about PR than anything else. They're driving by to be visible, to be seen, but it isn't so much about trying to affect the course or the direction of a school.

Because participants did not receive timely feedback during the course of the school year, it is hardly surprising that the evaluation process failed to deliver a formative impact to their reflective practice or professional development. As New Leaders (2012) explained, “spending more and better time evaluating and providing feedback to principals is one of the most important aspects” of an effective evaluation program (p. 4). For this reason, New Leaders (2012) asserted evaluators must “prioritize the evaluation process by spending more time in schools conducting high-quality observations and providing actionable feedback to principals” (p. 17). Failure to do so, posed the risk of evaluation devolving “into a compliance-driven process” (New Leaders, 2012, p. 17).

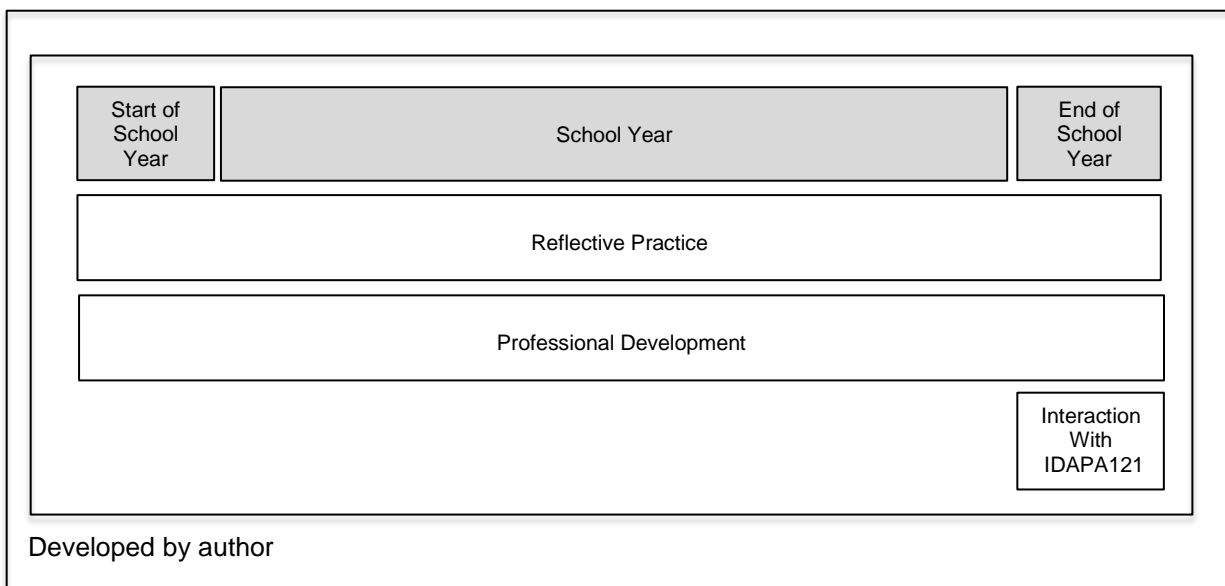
This study focused on the experiences and perspectives of the school principals only, and was not designed to explore IDAPA 121 from the vantage point of school district leaders or the supervisors who conducted the evaluations. As such, it is beyond the scope of this study to explore why supervisors spent so little time observing participants or providing feedback during the 2014-2015 school year. However, one policy related explanation might be suggested, and is derived from the above warning from New Leaders (2012) that a failure to invest in frequent and high quality observations and feedback posed the risk of evaluations devolving “into a compliance-driven process” (p.17). Might it be possible that the obverse of this statement is also true, i.e. that evaluations that are performed in order to “comply” with policy directives will result in infrequent observations and low quality performance

feedback? Although this is purely speculative, and deserving of further research, it is interesting to remember that IDAPA 121 was adopted by the State of Idaho to qualify for the ESEA Waiver Program discussed in Chapter One. Further, District Alpha and District Beta had no choice in the matter but were required to comply with the requirements of IDAPA 121. Thus we find two levels of compliance driven policy by the State and local Districts at the front end of this process, and a perfunctory implementation with infrequent observations and little meaningful feedback as outputs. As recommended in Chapter Six, additional research into the motivations and perspectives of State and District level would be worth exploring.

A third operational explanation is that IDAPA 121, as implemented in District Alpha and District Beta, did not require participants to engage with the professional practice standards until the end of the school year when they prepared for their performance evaluation conversation. By then it was too late for the standards to inform or influence participants' thinking or action during the school year. Henry's experience was typical. When asked to pinpoint when he first interacted with the evaluation process during the 2014-2015 school year, he answered, "Oh gosh, when I met with my supervisor to go over my evaluation ...at the end of the school year ... in [late] May or June." Elizabeth reported the same. When asked if she had any discussions with her supervisor during the school year about the professional practice standards, she explained she did not, and that her supervisor emailed her in early June detailing what information he wanted to see for her evaluation. Figure 2 illustrates the timing of participants' experience with reflective practice, professional development, and IDAPA 121 during the course of the 2014-2015 school year. As shown in Figure 2, participants actively engaged in reflective practice and different kinds

of professional development during the school year. However, the sequence and timing of events made it less likely for IDAPA 121 to influence their reflective practice or professional development during the 2014-2015 school year. This may be one reason why participants' described the evaluation process as 'disconnected' from their reflective practice, professional development and leadership practices.

Figure 2 (Timing of Events)



This conclusion is consistent with best practice recommendations from the research literature. The Wallace Foundation (2009) noted that “too often, leader assessment is seen as a single high stakes event – a form to be completed or an interview conducted – rather than an ongoing process connected to the goal of professional development and continuous improvement” (p. 4). Similarly, Davis et al. (2011) reported that “a common policy recommendation is that principal evaluations should be conducted on an ongoing, cyclical (rather than annual) basis” (p. 28). Clifford and Ross (2011), which favored a formative, developmental approach to evaluation recommended a process that begins with relevant standards, personalized goal setting, and multiple opportunities for practice and specific

feedback (p. 22). Each of these recommendations from the literature embodies the notion that evaluation should be an ongoing process, rather than a single, end of year event as was experienced by participants in this study.

This conclusion is also consistent with recommendations offered by several participants. They believed the evaluation process would be improved if evaluation was more of a yearlong process. They suggested an orientation to the leadership standards at the beginning of the year during a pre-conference, followed by more frequent observations and feedback, and a formal mid-year check in before the end of year discussion. Elizabeth, for example, explained that if she was evaluating principals, she would leverage the leadership standards and work with her principals to “set those goals up front and ask for that data along the way rather than just at the end.” Edward agreed, noting that “if you had a conversation at the beginning, a middle, and an end check in I think that would be reasonable and a reasonable expectation given the other things we're trying to do. “ Thus it appears likely IDAPA 121 and its performance standards were unlikely to influence participants’ reflective practice and professional development during the school year because participants were not formally required to review the standards until the end of the year. By then, it was too late for the performance standards to have any formative influence on participants. This is a purely practical explanation related to the timing of evaluation related events, as shown in Figure 2, and consistent with best practices in the research literature.

On the other hand, one could question why principals, who knew they were being evaluated at the end of the year on specific and publicly available standards, did not avail themselves of the opportunity to review the performance standards at the beginning of cycle and use that information to drive their own reflective practice and development. Should it

really be necessary to engineer a system that ensures principals sit down at the beginning of the year with the leadership standards to engage in standards driven reflection and goal setting? As this study found, eight of the ten participants did not review the standards until they were required to do so, strongly suggesting that a formal, front-end structure may be necessary. This conclusion is further supported by the perspectives of Robert and Henry who asserted a better program would systematically orient principals to the performance expectations and encourage relevant reflective practice and goal setting. Robert explained:

Somebody could say, well, the person could read what their evaluation tool looks like and they can then know what's expected. I think that, ultimately, in a legal sense somebody could argue that and say nothing else needs to be done. But in a practical sense, a good strong program is going to be one where the supervisors orient new principals to the [evaluation] tool itself, and talk about practical pieces of the tool, but also get the newer administrator to be immediately reflective of their own practice, and where they'd rate themselves as they look at the tool, and then set goals for themselves.

Regarding the need to have formal structures in place to ensure standards driven reflection and goal setting, Henry observed, "I think human beings are like water and they go to the place of least resistance if we can." When one considers the considerable demands on a principal's time and attention, and the number of 'required' tasks to be accomplished throughout the year, one can understand why participants largely ignored IDAPA 121's performance standards until formally required to focus on them by their districts at the end of the school year.

Substantive challenges. In addition to the operational issues discussed above, two potentially deeper substantive challenges are suggested by the findings and research literature and may further explain why IDAPA 121 failed to deliver the hoped for formative impact on participants' reflective practice and professional development. Although the operational challenges could conceivably be remedied, for example through implementation of a true evaluation "cycle" which provided more timely and relevant feedback, these substantive challenges may be more difficult to overcome. The first substantive explanation strikes at the heart of standards-based evaluations like IDAPA 121 that are based on comprehensive and one-size fits all "standards" like ISLLC '08. The second substantive challenge suggests that summative and formative evaluation purposes may be inherently incompatible, and therefore unlikely to deliver consistently positive impacts on the reflective practice and professional development of school leaders.

The first substantive challenge suggests standards based evaluation models like IDAPA 121 may fail to impact the reflective practice and professional development of participants because the "standards" upon which the evaluation is based are not relevant, engaging or helpful to school leaders. Support for this explanation is found in the experience of James and Charlotte who reviewed the performance standards when first made available, yet still reported no effect on their reflective practice, professional development or leadership practices. As James explained, his districts' evaluation form "is not anything that's going to guide my practice." Charlotte concurred, observing "while it's an important document ... this does not drive what I do. What drives my daily practice is who are the students on my campus ... what are the needs of my teachers, and what are our building goals." James reported a similar experience:

it really comes down to what my staff and I feel like the students' greatest needs are. So for example ... when you have students who are disconnected from school, how can we connect them to a school environment? That's a major focus. When you have students who are academically below grade level, how can you get them to grade level and that's a major focus.

These insights from Charlotte and James suggest that *timing* of the evaluation experience may not fully explain why the professional standards component of IDAPA 121 failed to influence participants' reflective practice and professional development. A deeper explanation may be that participants largely ignored IDAPA 121's professional standards because the standards lacked personal relevance and utility for these school leaders. As discussed below, this may have been true not only for Charlotte and James, but for the other eight participants as well.

Although somewhat speculative, and deserving of further investigation, this interpretation finds theoretical support in Knowles (1980) work on adult learning and in the reported experience of the participants. Knowles (1980) explains that adults "engage in learning largely in response to pressures they feel from their current life situation ... They tend, therefore, to enter an educational activity with a problem-centered or performance-centered frame of mind" (p. 53). This practical, problem solving orientation suggests school principals would focus their reflective practice and professional development on the specific real-life issues and challenges that they were facing in their school, and not on the general, and more abstract professional standards of IDAPA 121. This prediction is consistent with the experiences of James and Charlotte, noted above, in that both reported being focused on the specific needs of their students, teachers and schools. As James explained, "it really

comes down to what my staff and I feel like the students' greatest needs are." For Charlotte, the focus was her students, the needs of her staff, and achieving her school's goals.

Consistent with Knowles (1980), these two participants demonstrated "a problem-centered or performance-centered frame of mind" when it came to their own development and leadership practices. With these real-life issues and challenges as the priority focus, it is hardly surprising that they would choose to ignore the professional practice standards of IDAPA 121. Importantly, this interpretation is also consistent with the experience of the other participants who did not review the standards when first made available. In fact, this orientation to practical problem solving and performance may explain why the other eight participants completely bypassed the standards in the first place.

Nancy, for example, explained that she and her staff engaged in reflective practice and professional development in response to the perceived needs of their school. She observed:

The staff here is very experienced, honest, and good at identifying areas that they see as weak in the building. And it's not just for my development but also for theirs. If we, as a team are not doing well in an area, what are we going to seek out, or who are we going to bring in, or what are we going to strengthen. Sometimes it's just me, but sometimes it's a team of teachers or our entire staff.

Similarly, Edward, reported that his professional development occurred in the context of real life events and challenges in his school, noting "my growth this year really has been around experiences at my building" such as dealing with "parents in the building who were at odds with each". Mary reported that she was "guided more by [my] understanding of what the school needed" based on data and other events that came to her, as opposed to what she knew

the district valued. One final example comes from Henry, who explained that he is “constantly looking at everything that I’m doing to see that I’m being as effective as I can in meeting the needs of my parents, students and staff ... I’m looking at my students, and I’m thinking ‘am I meeting their needs ... are they getting the education they need’”.

From these and other participant examples a picture emerges of school principals who are focused on understanding and meeting the needs of their students, staff, and school communities through reflection, development and the exercise of school leadership. This corresponds well to the “problem-centered and performance-centered frame of mind” predicted by Knowles (1980) and may explain why they ignored the broad, generally applicable standards and focused their reflection and development instead on meeting the narrower, context specific needs of their schools.

This emphasis on the context-specific needs of a school finds additional support in the literature on contingent leadership (Bossert, et al., 1982; Hallinger, 2003; Leithwood, et al., 2004). Their work argued against the development and imposition of ‘one size fits all’ lists of desired leadership styles or behaviors such as those found in the leadership standards of IDAPA 121. As they explained, “no single style of management seems appropriate for all schools ... principals must find the style and structure most suited to their own local situation (Bossert, et al., 1982, p. 38). Leithwood, et al. (2004) stated:

Impressive evidence suggests that individual leaders actually behave quite differently (and productively) depending on the circumstances they are facing and the people with whom they are working. This calls into question ... the search for a single best [leadership] model or style (p. 10).

This ‘contingent’ theory of leadership is consistent with participants’ reported focus on the specific circumstances, data and needs of their schools. Understood in this way, participants could be viewed as demonstrating exemplary leadership by ignoring the broad and general prescriptions of IDAPA 121 when engaging in reflective practice and development. As Hallinger (2003) observes, “the type of leadership that is suitable to [one set of circumstances] may well become a limiting or counterproductive force [under different circumstances]” (p. 345-346).

A second substantive explanation for the failure of IDAPA 121 to impact reflective practice and professional development relates to potential conflicts between summative and formative purposes of evaluation which may undermine the hoped for formative effects. The literature recognizes “inherent tradeoffs and tensions among these [summative and formative] purposes” (Portin, et al., 2006, p. 40; Wallace Foundation, 2009, p. 7). As Wallace Foundation (2009) observes:

It’s generally easier to have a candid conversation about a principals’ strengths and weaknesses when the purpose is mainly to identify organizational or professional development needs than when a raise, a promotion, or possible termination are at stake (p. 7).

Portin et al. (2006) describe the summative and formative goals as “contrasting purposes for assessment ... [which] imply different approaches to assessment” (p. 40). Despite the recognition of the inherent differences and potential conflicts between the summative and formative purposes, proponents of evaluation reform argued that modern evaluation programs should be designed to achieve both objectives (Guilfoyle, 2013, p. 4). Importantly, the U.S. Department of Education’s ESEA Waiver Program, which was the impetus for the

State of Idaho's adoption of IDAPA 121, explicitly required states to adopt evaluation programs that accomplished both objectives.

However, the perceived conflict and risk associated with the combined summative / formative model of IDAPA 121 may have discouraged some participants from engaging in open reflective practice with their supervisors during their evaluation experience. Although this is a tentative, possible explanation only, and additional research is needed, the experience of Nancy and several other participants is interesting to consider. Nancy believed that the combined summative / formative model was "probably" a problem and "adversarial towards each other". She explained:

When I think about my conversations with my evaluator when he comes in, I'm more guarded in what I say to him when I know he's writing down what I say on my evaluation tool. If it's an open, honest conversation, I'm much more open about maybe asking for feedback or for constructive criticism, but I don't want him to misinterpret something and think I'm not proficient in something because I'm asking. So our conversations are different when I know he's writing on the evaluation tool versus when he's not. And I'm sure teachers are the same way with me.

Nancy's perspective was shared by other participants, including Scott and Mary. Scott observed that summative and formative purposes are inherently in conflict, and "you're not going to get all of the potential out of the developmental approach if the person you are going to for development support is the person judging you at the end of the year." Reflecting on the fact that classroom teachers have access to veteran teachers who can assist with development and who are in a completely non-evaluative role, he stated: "it would be great if we had something like that for school leaders." Mary agreed that summative and formative

approaches are “at cross-purposes to one another”, and their co-existence is a “tough challenge.” He explained that people cannot grow or develop professionally unless they feel safe, and are able to be open and vulnerable, but that “it doesn’t always feel safe to be open and vulnerable when you are being judged. Reflecting on evaluation of his teachers, where he needs to serve both formative and summative roles, he noted it has “been one of the biggest challenges of my tenure over the years”, even though “they taught me in principal school that you could be an instructional leader, a coach, and an evaluator all at the same time.

It is worth noting that not all participants believed that summative and formative purposes inherently conflict; some believed it depended on the particular principal. Charlotte, for example, reported: “I think it really depends on the person. With my supervisor, I feel like we have a great relationship, so when we sat down it was both.” Robert expressed the belief that “you’re going to have a wide range of folks on the spectrum”, with some people who are comfortable with the dual purpose and others who are not. He recognized the potential for disruptive conflict inherent in combining the two, and stated “it would definitely need to be presented well; you would need to have a good process, a good roll-out.”

Henry went even further by sharing the opinion that professional development should be completely separate from the summative evaluation process. He explained,

I’ll go back to what I said originally, the purpose of evaluation is to determine if a given principal or teacher is going to receive a contract for another year, and I’m actually good with that. I wish honestly that the entire evaluation process consisted of ‘yes, you have a job, or no, you don’t have a job’, and it could be completed with a simple click of a button because that is all that matters at the end of the day.

To be clear, Henry believed strongly in the importance of constructive feedback, professional development, and coaching for school leaders and teachers alike, noting that “if you didn’t do that then everything I said would be a disaster.” But in his view, there was no reason to combine the formative and summative objectives within a single, common process.

In summary, the professional practice standards did not contribute to participants’ reflective practice, professional development, or leadership practice, contrary to the claims of those who advocated standards based evaluations like of IDAPA 121. This result has possible operational explanations that theoretically could be addressed by changes to the evaluation process. These would include moving to more of a true evaluation cycle, rather than a single, end of year event, and the provision of more frequent and timely feedback by the evaluators. Assuming those challenges could be addressed, there may be potentially deeper substantive reasons that would be more difficult to remedy. These include (a) the possibility that participants might continue to disregard the one-size fits all leadership standards because they are perceived as not relevant or helpful, and (b) inherent conflicts between the combined summative and formative objectives that underlie the ESEA Waiver requirements and IDAPA 121.

Analytic theme two: The relationship between IDAPA 121 and participants’ leadership practices

Participants from both districts reported their experience with evaluation under IDAPA 121 did not drive their leadership practice. William put it this way: “Honestly, there has not been one day this entire year where I’ve thought about doing something because of that evaluation piece. Not one day. Not one minute.” This is hardly surprising given the finding that IDAPA 121 did not impact participants’ reflective practice or professional

development. In fact, several of the same explanations discussed above may also help shed light on why evaluation did not impact participants' leadership practice. As illustrated in Figure 2, one practical explanation may be that participants were not required, or even encouraged, to think about the contents of their evaluation until the end of their school year. Although policy makers may have wanted IDAPA 121 to serve as a map to guide the direction of these school leaders as they started their journey, it was actually experienced more like a rear-view mirror that participants briefly checked at the end of their drive. By the time of their formal evaluation, it would have been too late for IDAPA 121 to shape their thoughts or approaches to school leadership for the school year then ending.

As discussed above, two participants (Charlotte and James) did review their evaluation form during the school year, yet still it had no impact on their leadership practices. This may be explained by contingent leadership theory ((Leithwood, et al., 2004) which posits that effective leaders are attuned to the specific, unique needs of their schools rather than simply following a cookie-cutter approach to leadership. This explanation finds support in their description of what did drive their leadership practice. For Charlotte, it was “who are the students on my campus, who are the parents on my campus, what are the needs of our teachers, and what are our building goals.” For James, the focus was on “what my school needs, what our kids need, what my teachers need, and what should be the strategic focus for the specific needs of the school at that time.”

Other participants who did not review their evaluation during the school year described their leadership focus in similar ways. This suggests that IDAPA 121 may have had little effect on their leadership practice even if they had reviewed their evaluation forms earlier in the year like Charlotte and James. Mary, for example, explained that her leadership

was focused on “what the school needed and what she needed to be doing differently with her leadership team based on data and other events that came to her.” Robert similarly acknowledged that his leadership priorities were “really self-driven, as opposed to originating out of the performance evaluation.”

Analytic theme three: The relationship between the prospect of being measured on student growth and participants’ leadership focus and practices

One of the most significant differences between IDAPA 121 and prior evaluation models was the inclusion of the student growth component that would constitute one-third of participants’ total evaluation. Implicit in this change is the assumption that principals would be motivated, by a desire to obtain a favorable rating, to take additional or different leadership actions to elevate student performance on the relevant summative assessment. This study sought to explore the validity of that assumption by exploring the relationship between the student growth component of IDAPA 121 and participants’ leadership practice.

It is understandable that policy makers would be interested in the motivation of school leaders, and seek ways to ensure principals are motivated to improve student outcomes. Motivation is considered critical to strong performance, as shown by the following formula for performance: $\text{job performance} = \text{motivation} \times \text{ability}$ (Latham, 2012, p. 3). Under this formulation, motivation is seen as a key driver and multiplier of performance. This study revealed, however, that participants were not motivated or otherwise impacted in any way by the student growth component of IDAPA 121. As Elizabeth explained, being measured on student growth was “not changing the game at all. We've always been focused on performing well in the state assessments.” Other participants concurred, reporting that “it didn't change anything for me on how I operate” (Charlotte), and “student achievement has

always been a priority for us ...the only difference now is it's a priority for my supervisor (William).

Although the student growth component was not implemented at the end of the 2014-2015 school year because of data availability problems, the participants did not know whether this would be the case during the next school year. Thus, their perspectives on this element of IDAPA 121 may help shed light on the utility of using student performance outcomes to motivate school leaders within the evaluation context, and may have implications for other related policies like merit-based pay. This section of the study explores possible explanations for “why” the student growth component of IDAPA 121 did not influence participants’ leadership focus or practices.

One possible explanation is suggested by the participants themselves, who reported they were already focused on improving teaching and student learning. According to this explanation, they were already fully engaged and motivated to drive student learning in their schools; as a result, the hoped-for “incentive” effect of the student growth component failed to materialize. Nancy’s perspective supports this explanation, as she described her own internal motivation to help her students succeed:

We should be allowing students to make academic, social, emotional, physical growth, you know, all year long, and I like doing that. Um, but it's not because it's measured or the District tells me to -- that's an internal motivation I have and I think that's part of the reason why, um, I like this job because I can help institute changes. Nancy underscored the power of her own motivation by observing, “I do it because I care about the students in this building and I care about our community greatly. And if I didn't care, an evaluation tool isn't going to change that.” Other participants further supported the

idea that they were already motivated to drive teaching and learning. Mary, for example, reported that “everything we do tries to focus on improving learning.” As she added, “my ultimate goal is student achievement. That's what I'm in this job for, and if that's not our focus, then I shouldn't be in that leadership role.” This perspective belies the assumption that may be underlying policies like IDAPA 121 that principals are not focused on student outcomes, and that external pressure is therefore needed to help motivate their best efforts. Edward directly challenged this assumption, adding further support for the explanation that participants were not motivated by the student growth component because they were already fully motivated on their own. Edward asked:

What's the real intent of the evaluations? What's the belief behind it? I think even some of the assumptions that these evaluation tools are built on are really unfounded. It's hard to build a good tool off of an assumption that's really not true that principals really don't care or teachers maybe don't care if students achieve.

Charlotte further reinforced the subject explanation by observing, “wouldn't it be egocentric of me to say ‘if my evaluation is going to reflect how my students do, I need to really push and drive’ ... the motivation is not right.” These participant perspectives provide support for the explanation that the student growth component of IDAPA 121 failed to impact participants because they were already fully motivated to help their teachers and students succeed. At the same time, Charlotte's comment about the motivation being “not right” points to a second possible explanation, and one that finds support in the research literature on intrinsic and extrinsic motivation (Latham, 2012).

Intrinsic motivation is defined as “the doing of an activity for its inherent satisfactions rather than for some separable consequence” (Ryan & Deci, 2000, p. 56). Extrinsic

motivation, by contrast, “is a construct that pertains whenever an activity is done in order to attain some separable outcome” (Ryan & Deci, 2000, p. 60). Organismic Integration Theory (Deci & Ryan, 1985) posits a continuum of extrinsic motivation with varying loci of control and associated processes. At one end of the spectrum is ‘external regulation’ which has an external locus of control. Motivation that is driven by extrinsic rewards or punishments, such as receiving a positive performance evaluation and possible raise or promotion, or that is driven by fear of receiving a negative evaluation and potentially adverse employment consequences, would fall within this ‘external’ aspect of extrinsic motivation. At the other end of the spectrum of extrinsic motivation is what the authors call “integration” which has an internally focused locus of control and is associated with the synthesis or congruence of goals and personal values. If for example, one valued learning, and believed it was important for students to perform well on standardized tests because this represented something positive and good for the wellbeing of the students, then motivation that was directed towards helping students perform well on such tests would be an example of “integration” and fall within the “internal” aspect of extrinsic motivation. Although this might appear an example of intrinsic motivation, it is more properly an example of “internalized” extrinsic motivation because it is “an activity that is done in order to achieve a separable outcome” - in this case positive outcomes on a standardized exam.

As Ryan, Kuhl, and Deci (1997) found, “internalized” extrinsic motivation “yields manifold adaptive advantages, including more behavioral effectiveness due to lessened conflict and greater access to personal resources ... [a finding with] clear significance ... for behavioral ad performance outcomes” (Ryan and Deci, 2000, p. 64). This suggests, within the context of the student growth measure of IDAPA 121, that motivation to improve student

performance that was externally driven by a desire to receive a positive evaluation would be much weaker and less effective than a motivation that was internally driven by a sincere and personal desire to see students performing well. As long as the participant was internally motivated to help his or her students learn and succeed on measures of learning, Organismic Integration Theory suggests the additional “incentive” of receiving a positive performance evaluation would be unnecessary and add no further value. In the words of Charlotte, the “motivation” would simply “not be right.” As Nancy put it, when discussing her attitude about the student growth component of IDAPA 121: “I care about the students in this building, and I care about our community greatly. And if I didn’t care, an evaluation tool isn’t going to change that.”

Taken one step further, it might be possible to find truly intrinsic motivation in the example of a principal working with standardized test scores, where the principal engaged in instructional leadership activities because he or she found them inherently enjoyable. One possible example comes from the experience of Nancy who reported the following:

I love looking at the data and seeing are the students here at this school growing and disaggregating the data and asking, so what group grew the most, and which group didn’t grow? Whether it’s socioeconomic, or race or ethnicity, or just their baseline performance level when they came into grade level, what quartile were they in, what quartile are we growing the most. Trying to figure it out, taking the data down to the smallest possible chunk, so we can say ‘ok, here’s one area we have to work on this year ... I like doing that.

This excerpt reveals a school leader who experiences a powerful, internal motivation to do the work of data driven leadership. This is another instance where the additional “incentive”

of receiving a positive performance evaluation would be unnecessary and add no further value.

To summarize, participants may have experienced little connection between the student growth component of IDAPA 121 and their own leadership practices because, (a) practically speaking, they were already sufficiently motivated to improve teaching and learning, and (b) the prospect of being measured on student growth relied on the weakest form of externally focused extrinsic motivation that offered little in the way of additional motivational power when compared with their own internally driven motivation.

Analytic theme four: Meaningful professional development

As Portin, et al. (2006) observed, the primary aim of revised, standards-based performance evaluations like IDAPA 121 was “the improvement of leadership performance” (p. 3). Proponents believed effective evaluations could help promote the growth of leaders, and contribute to a “career-long professional development process” (p. 15). It was claimed that

Informal and formal assessment bring evidence to bear on leaders’ efforts to develop their skills (e.g. how to handle student data wisely). The leader as learner takes in feedback, imagines what it reveals about progress made or new areas for learning, and establishes different sets of learning aims. This feedback is not necessarily public but is part of the individuals’ efforts to make sense of their own learning (Portin et al., 2006, p. 15).

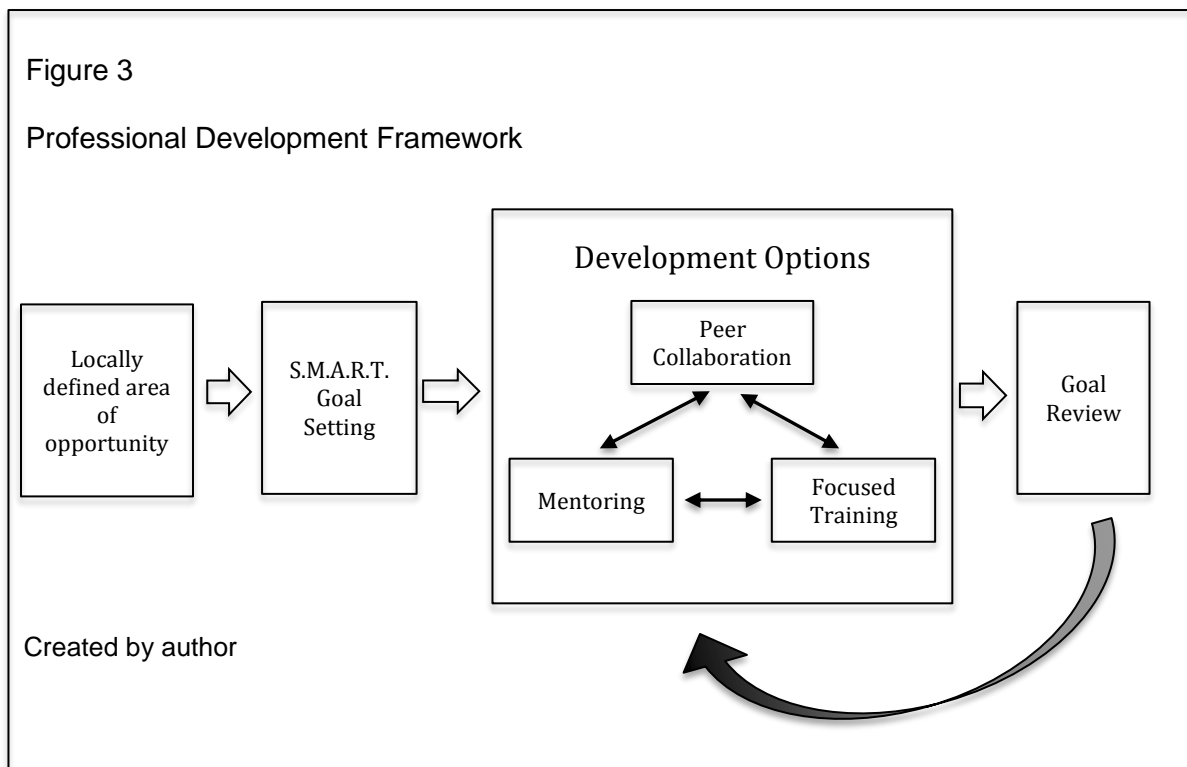
Against this promising context, IDAPA 121 was implemented across the State of Idaho, including in District Alpha and District Beta, with the hope that it would contribute to principals’ reflective practice and professional development. Despite these good intentions,

this study found no relationship between participants' experience with IDAPA 121 and their own reflective practice or professional development. This finding begs the question "if performance evaluation does not drive learning and development as hoped, what might work better?" One answer is suggested by the standard formula for improving performance: $\text{performance} = \text{motivation} \times \text{ability}$ (Latham, 2012, p. 3). Because this study suggests participants were highly motivated without IDAPA 121, one promising approach to improving participant performance would be to focus more explicitly on developing their abilities. Although "existing research provides little guidance on the types and forms of information that are helpful to principals" (Clifford & Ross, 2011, p. 7), participants offered clear ideas about what they would find helpful for their own professional development.

As reported in Chapter Four, participants recommended four broad categories of professional development related activities: (a) goal setting, (b) peer collaboration, (c) mentoring, and (d) high quality, focused training. Taken together, and viewed through the lenses of adult learning theory (Knowles, 1980) and Organismic Integration Theory (Deci & Ryan, 1985), these categories suggest the revised professional development framework shown in Figure 3 (Professional Development Framework).

The first element of the framework is the locally defined area of opportunity. Drawing from contingent leadership theory (Leithwood, et al., 2004), which asserts that different schools may require different forms of leadership at different times, the initial power of the framework arises from the opportunity for each school leader and leadership team to identify the specific area of opportunity within their own school to improve teaching and/or learning. Participants are likely to find this initial step deeply motivating and engaging, as they reported having a sharp, internal focus on the unique needs and challenges

of their school which coincides with internal orientation under Organismic Integration Theory (Deci & Ryan, 1985). As James explained, what drives his leadership practice is his “own understanding of what my school needs, what our kids need, what my teachers need, and what should be the strategic focus for the specific needs of the school at that time.”



Having identified the area of opportunity, the second step in this professional development framework is the creation of S.M.A.R.T. goals for the school and/or leader that relate to the chosen area of opportunity. S.M.A.R.T. is a well-known acronym in goal setting and communicates that effective goals are specific, measurable, achievable, relevant and time-bound (Massachusetts Institute of Technology, 2016). Goal setting has been shown to be a powerful driver of motivation and performance (Latham, 2012, p. 65), and helps set the stage for engaged learning in the next step of the framework.

Building on the recommendations of participants and adult learning theory, the third step of the framework offers participants a flexible choice among desired professional development activities that are best suited to the needs of the learner. In some cases, a combination of two or more of the different types of activities may be appropriate. For example, as was the case with Edward, a principal may desire to learn more about how to use data to drive more effective teaching. That learning objective might be satisfied through peer collaboration where principals share their own best practices, through targeted training, or mentoring. Alternatively, the principal might elect to pursue all three options, beginning first with targeted training, followed by peer collaboration and mentoring. Adult learning theory (Knowles, 1980) supports the synergy among the first three steps of this framework because it predicts that principals will “engage in learning largely in response to pressures they feel from their current life situation ... They tend, therefore, to enter an educational activity with a problem-centered or performance-centered frame of mind” (p. 53). By opening first with identification of an opportunity that the principal finds worthwhile, and then transitioning to concrete goal setting, the stage is set for engaged adult learning with this “problem-centered or performance-centered frame of mind.”

Based on the feedback of participants, peer collaboration and targeted training would be organized and delivered consistent with the recommendations outlined in Chapter Four. Formal, structured, peer collaboration, for example, should be conducted monthly, or more frequently if desired by participants, with homogenous groupings of principals, without supervisors present, and with a flexible agenda that accommodates the needs and interests of the participants. Such collaboration could occur among a large group of principals or as few as two. Formal training should feature a narrowly targeted, personalized focus on what the

leader needs to help the school leadership team achieve its goals. The training should be of high quality, and represent the most current thinking and practice. Learning checkpoints should be built into the timeline to ensure progress towards the learning goals is being achieved.

The next step in the framework is goal review, where the principal and his/her leadership team monitor progress against the organizational goals identified in step two. As indicated by the arrow pointed back from Goal Review to Professional Development Activities, this framework is envisioned as a cycle where principals would set goals, engage in learning, apply what they learned, measure progress against the goal, and possibly repeat some or all parts of the process as needed.

To summarize, the professional development framework introduced above appears to have several advantages over the standards-based evaluation model of IDAPA 121. Whereas IDAPA 121 began with a focus on the standards, and asked participants to engage in reflective practice and professional development independent of the actual priorities and needs of the school, this framework flips that approach 180 degrees and begins first with the perceived priorities and needs of the school and allows those to guide the professional development agenda. As discussed earlier, this approach is more consistent with adult learning theory (Knowles, 1980) and other relevant theories of motivation. Additionally, this framework is built on the three professional development activities recommended by participants – peer collaboration, formal training, and mentoring.

Summary

This chapter offered analysis and interpretation of the study's four major findings, in order to provide possible explanations for why IDAPA 121 did not impact participants'

reflective practice, professional development, or leadership practices. Interpretive analysis of four aspects of participants' experience with IDAPA 121 were explored: (a) the relationship between the 'professional practice standards' of IDAPA 121 and participants' reflective practice and professional development (Research Question 1); (b) the relationship between IDAPA 121 and participants' leadership practice (Research Question 3); (c) the relationship between the prospect of being measured in part on student growth, and participants reflective practice and professional development (Research Question 2); and (d) participants' perceptions about how to improve professional development for principals (Research Question 5).

Several operational, or procedural explanations were offered as to why IDAPA 121 did not deliver the promised impact. These were based not only on participants' reported experiences but also recommended best practices from the research literature. At a deeper, substantive level, other possible explanations were explored that find support in participants' experience and theories of adult learning, contingent leadership, and motivation. Together, these substantive explanations suggest that IDAPA 121 might still fail to deliver on its promise to drive reflective practice and professional development even if the operational challenges were resolved.

Chapter Six

Conclusions and Recommendations

Introduction

The purpose of this study was to describe how well the “promise” of IDAPA 121 compared with its actual impact on participants’ reflective practice, professional development, and leadership practices of Idaho principals. The conclusions from this study arise from the research questions and findings, and relate to the following three main areas: (a) the non-impact of IDAPA 121 as experienced by participants, (b) the student growth component of IDAPA 121, and (c) the value of a non-evaluative, developmental alternative to increase principal capability and performance. This chapter discusses the major findings and conclusions to be drawn from this study, and offers recommendations for further research and possible policy changes.

Non-impact of IDAPA 121

Before enactment of “improved”, standards-based evaluations like IDAPA 121, evaluation of school principals was found to have little impact on their professional development or leadership practice. Reformed models of evaluation like IDAPA 121 were supposed to change this, and usher in a new wave of standards-based reflection and professional development that would enhance leader capabilities and performance. Despite the promise, the first and third major findings of this study reveal that IDAPA 121 had no material impact on participants’ reflective practice, professional development, or leadership practice. This might be explained by problems with how IDAPA 121 was implemented as an isolated, end of year activity rather than a true cycle of reflection and feedback as recommended in the literature. Proponents of IDAPA 121 might assert, therefore, that the

promised benefits of IDAPA 121 can be realized through improvements to how it is implemented within the districts. However, deeper analysis suggests that substantive issues with the standards based, dual-purpose evaluation model itself might still render the evaluation model ineffective in its formative objectives. Although this is a tentative interpretation that warrants further research, several noteworthy conclusions can be offered. The first conclusion is that standards-based evaluation models may be ineffective at driving professional development and strong school leadership, contrary to promises in the literature. If true, this would call into question similar evaluation programs across Idaho and most of the U.S., and may have additional implications for how standards like ISLLC-2008 are used in other contexts like school leader certification programs. A second, possible conclusion is that dual purpose (summative and formative) evaluation schemes are ineffective at developing school leader capabilities. If true, this would undermine important federal policies (e.g. ESEA Waiver program), Idaho state policy under IDAPA 121, and similar state programs across the country. Hoped for improvements in school leadership may fail to materialize, as would follow-on gains in teaching, learning, and student achievement.

The student growth component of IDAPA 121

The study's second major finding was that the student growth component of IDAPA 121, and the prospect of being measured in part on student achievement, had no material impact on participants' leadership focus or practices. One conclusion to be drawn from this is that contrary to the expectations of policy advocates, incentive and consequence models like IDAPA 121 are misaligned to the needs and motivations of some school leaders. Participants did not appear lacking in deep, personal motivation to improve their schools, teaching or learning. This conclusion is further supported by the finding that participants believe policy

makers are out of touch with what really happens in public schools. If true, this would have implications for IDAPA 121 in other parts of Idaho, similar evaluation models in other states, and other incentive based programs like merit pay for school leaders and possibly for teachers as well.

Non-evaluative, developmental alternative

It is worth remembering that evaluations like IDAPA 121 were designed to have both a summative and formative effect, and it was hoped the evaluation experience would lead to meaningful professional development for school leaders. This study's conceptual framework (Figure 1), illustrates that IDAPA 121 was a "means" to the end of reflective practice and professional development, which in turn would drive more effective leadership practices.

If reflective practice and professional development were one important goal, then why did policy makers not simply focus on the delivering directly against that goal? This question assumes greater salience in light of this study's finding that participants had clear ideas about how to improve the reflective practice and professional development of school leaders. Importantly, none of these ideas depended upon an evaluation program like IDAPA 121, and several participants asserted that evaluation was not necessary for professional development. One conclusion to be drawn from this finding, and the finding that IDAPA 121 did not drive professional development, is that policy makers at the state and district level would be better served by delivering relatively low cost solutions like peer collaboration and mentoring as recommended by participants. A related conclusion is that if policy makers want to increase principals' capabilities in the three focus areas of IDAPA 121 (instructional leadership, collaborative leadership, and school climate), then some investment in high

quality direct instruction would be indicated, and much more likely to increase leader capabilities in these areas.

Recommendations for policy changes

Several policy changes are indicated by this study's findings, interpretive analysis, and relevant research literature. These changes would appear more likely, than a continuation of the *status quo* under IDAPA 121, to achieve the stated goals of increasing principal reflective practice and professional development. Four recommendations are offered. Three would likely require additional resources at the state or district level. The fourth recommendation would likely result in cost savings to the State of Idaho and its local school districts. These cost savings are unlikely to offset the resource requirements of the other recommendations, although no study was undertaken to estimate resource needs or cost savings. The four recommendations are detailed below.

First, to the extent possible under ESSA, the formative and summative purposes of IDAPA 121 should be decoupled, and professional development of Idaho public school principals managed separately from the traditional, summative goal of evaluation. Prior research showed that evaluation programs were ineffective at driving meaningful professional development, and the current study suggests this remains the case now under IDAPA 121.

Second, the participants expressed interest in meaningful professional development to increase their own leadership capability and that of other school leaders. Idaho school districts should take steps to implement the revised model of professional development outlined in Chapter Five. These could be implemented in pilot form. Some elements of the

model, including district sanctioned peer collaboration and mentoring, would be relatively inexpensive.

Third, there are clear opportunities to make the evaluation process more meaningful from a summative standpoint. Accordingly, Idaho school districts should take steps to ensure evaluation operates as a true cycle, and to explore ways to deliver more timely and relevant feedback to their school leaders. If the student growth component is retained within the evaluation process, the State of Idaho and local districts must ensure valid and reliable data are available before the evaluation cycle ends.

Fourth, to the extent possible under ESSA, Idaho should eliminate the student growth component from its formal evaluation process for public school principals. This recommendation flows from the finding that the student growth measure did not productively impact participants, their professional development or leadership practices. Additionally, participants expressed doubts about the validity of student test scores to measure their own leadership effectiveness because of the multiple factors outside their direct control and influence that impact student performance.

Recommendations for further research

I recommend additional research be conducted to validate and expand upon our understanding of the efficacy of evaluation models like IDAPA 121, their impact on reflective practice, professional development, and leadership practices of school leaders, and the follow-on effects on teaching and learning. In light of this, research into the following aspects of the subject phenomenon should be explored: (a) this study should be replicated with participants from smaller and rural Idaho school districts in order to compare and contrast their experiences with those from this study; (b) because other states have enacted similar

evaluation models, a comparative study across multiple states might be conducted to yield broader understanding; (c) a study should be undertaken that explores the efficacy of IDAPA 121 from the perspective of national, state, and district level officials, including district supervisors who perform the evaluations; (d) a study should be undertaken that further explores whether, and to what extent, principals find Idaho's standards of effective principals helpful outside the context of performance evaluations; (e) a study should be undertaken that further explores if, and to what extent, the dual purpose summative / formative evaluation model hinders or interferes with principals' professional development; (f) a study should be undertaken that explores the cost benefit analysis and efficacy of the professional development program recommended in Chapter 5.

Summary

This chapter presented conclusions and recommendations regarding performance evaluation under IDAPA 121 and the professional development of school principals. Conclusions focused on (a) the non-impact of IDAPA 121, (b) the student growth component of IDAPA 121, and (c) a non-evaluative, developmental alternative to drive more meaningful professional development. Recommendations for policy changes and further research closed the chapter.

References

- Bickman, L., Goldring, E., De Andrade, A.R., Breda, C., & Goff, P. (2012). Improving principal leadership through feedback and coaching. *Society for Research on Educational Effectiveness, Spring Conference*. Retrieved from <http://eric.ed.gov/?q=improving+principal+leadership+through+feedback&id=ED530123>.
- Bingham, D. (2013). A mixed methods study exploring the realities and perceptions of principal evaluation (Doctoral dissertation). Retrieved from <http://gradworks.umi.com/35/99/3599421.html>.
- Bloomberg, L.D. & Volpe, M. (2008). *Completing your qualitative dissertation: A roadmap from beginning to end*. Thousand Oaks: Sage Publications.
- Bossert, S., Dwyer, D. Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, 18(3), 34-64.
- Bright, B. (1996). Reflecting on “reflective practice”. *Studies in the Education of Adults*, 28(2), 162-184.
- Brinkman, S. & Kvale, S.(2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage.
- Browns-Sims, M. (2010). Evaluating School Principals. *National Comprehensive Center for Teacher Quality*. Retrieved from <http://eric.ed.gov/?q=evaluating+school+principals&id=ED543770>.
- Buckingham, M. & Coffman, C. (1999). *First, break all the rules: What the world's greatest managers do differently*. New York: Simon & Schuster.
- Catano, N., & Stronge, J.H. (2006). What are principals expected to do? Congruence between principal evaluation and performance standards. *NASSP Bulletin*, 90(30), 221-237.

- Clifford, M., Hansen, U.J., & Wraight, S. (2014). Practical guide to designing comprehensive principal evaluation systems, Revised Edition. *Center on Great Teachers & Leaders at American Institutes for Research*.
- Clifford, M. & Ross, S. (2011). Designing principal evaluation systems: Research to guide decision making: Executive summary of current research. *National Association of Elementary School Principals*. Retrieved from https://www.naesp.org/sites/default/files/PrincipalEvaluation_ExecutiveSummary.pdf.
- Clifford, M. & Ross, S. (2012). Rethinking principal evaluation: A new paradigm informed by research and practice. *National Association of Elementary School Principals and National Association of Secondary School Principals*. Retrieved from <http://www.naesp.org/rethinking-principal-evaluation>.
- Clifford, M., Behrstock-Sherratt, E., & Fetters, J. (2012). The ripple effect: A synthesis of research on principal influence to inform performance evaluation design. *American Institutes for Research*. Retrieved from http://www.air.org/sites/default/files/downloads/report/1707_The_Ripple_Effect_d8_Online_0.pdf.
- Colvin, G. (2010). *Talent Is Overrated: What Really Separates World-Class Performers from Everybody Else*. Penguin Books: London.
- Corbin, C. (2015) Ybarra, school groups welcome new federal education law. *IDEDNEWS.org*. December 16, 2015. Retrieved from <http://www.idahoednews.org/news/ybarra-school-groups-welcome-new-federal-education-law/#.Vovsbc4k-IU>
- Corbin, J. & Strauss, A. (1990). Grounded theory research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*. Volume 13:1. Retrieved from:

- <http://sites.duke.edu/niou/files/2014/07/W10-Corbin-and-Strauss-grounded-theory.pdf>
- Council of Chief State School Officers. (2008). Educational leadership policy standards: ISLLC 2008. *National Policy Board for Educational Administration*. Retrieved from http://www.ccsso.org/Resources/Publications/Educational_Leadership_Policy_Standards_ISLLC_2008_as_Adopted_by_the_National_Policy_Board_for_Educational_Administration.html.
- Creswell, J.W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007). Preparing school leaders for a changing world: Lessons from exemplary leadership development programs. *Stanford Educational Leadership Institute*. Stanford, CA: Stanford University.
- Davis, S. & Hensley, P.A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education*, 13(4), 283-403.
- Davis, S., Kearney, K., Sanders, N., Thomas, C., & Leon, R. (2011). The policies and practices of principal evaluation: A review of the literature. *WestEd*. San Francisco, CA.
- Deci, E. & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior* (1985). New York: Plenum
- Derrington, M.L. & Sanders, K. (2011). Conceptualizing a system for principal evaluation. *AASA Journal of Scholarship and Practice*, 7(4), 32-38. Retrieved from http://www.aasa.org/uploadedFiles/Publications/Newsletters/JSP_Winter2011.FINAL

- [.pdf.](#)
- Edmunds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-24.
- Fuller, E.J. & Hollingworth, L. (2013). A Bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*, 50(3), 466-400.
- Goldring, E., Cravens, X.C., Murphy, J., Porterm A.C., Elliot, S.N., Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *The Elementary School Journal*, 110(1) 19-39.
- Goldring, E., Porter, A., Murphy, J., Elliot, S.N., & Cravens, X. (2007). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Wallace Foundation*. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/Assessing-Learning-Centered-Leadership.pdf>.
- Goleman, D. (2000). Leadership that gets results. *Harvard Business Review*, March – April, 2000. Reprint R00204, 1-14.
- Goleman, D., Boyatzis, R., McKee, A. (2002). Primal leadership: realizing the power of emotional *intelligence*. Boston, MA. Harvard Business School Press
- Guilfoyle, C. (2013). Principal evaluation and professional growth. *Association for Supervision and Curriculum Development, Policy Priorities*, 19(2), 1-3, 5-7
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329-351.
- Halverson, R. & Clifford, M. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Educational Administrative Quarterly*, 42(4), 578-

619.

- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: Principals make sense of complex artifacts to shape local instructional practice. In W. K. Hoy & C. G. Miskel (Eds.), *Administration, Policy and Reform: Research and Measurement Research and Theory in Educational Administration*. Greenwich, CT: Information Age.
- Harrison, W.C. & Peterson, K.D. (1986). Pitfalls in the evaluation of principals. *The Urban Review* 18(4), 221-235.
- Idaho State Department of Education. (2012). ESEA Flexibility Request. Retrieved from <http://www2.ed.gov/policy/eseaflex/id.pdf>.
- Idaho State Department of Education. (2013). ESEA Flexibility Request (Current): Principle 3 Updated June 26, 2013. Retrieved from <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/id.html>.
- Imel, S., Kerkam S., Wonacott, M. (2002). Qualitative Research in Adult, Career, and Career-Technical Education. Practitioners File. Washington, DC: *ERIC Clearinghouse on Adult, Career, and Vocational Education*. (ERIC No. ED472366).
- Jackson, D. (2000). The school improvement journey: Perspectives on leadership. *School Effectiveness and School Improvement*, 7, 1-34.
- Jacques, C., Clifford, M., & Hurnung, K. (2012). State Policies on Principal Evaluation: Trends in a Changing Landscape. *National Comprehensive Center for Teacher Quality*. Washington, DC. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/StatePoliciesOnPrincipalEval.pdf>.
- Kimball, S.M., Milanowski, T., & McKinney, S.A. (2007). Implementation of standards-based principal evaluation in one school district: first year results from randomized

- trial. Paper presented at the annual conference of the American Educational Research Association. Retrieved from http://cpre.wceruw.org/publications/kimballmilanowski_mckinney.pdf.
- Kimball, S.M., and Milanowski, A.T., and McKinney, S.A. (2009). Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial. *Leadership and Policy in Schools*, 8 (3): 233–236. Retrieved from <http://eric.ed.gov/?id=EJ856582>.
- Knowles, M.S. (1980). *The Modern Practice of Adult Education*. New York: Cambridge University Press.
- Lashway, L. (2003). Improving principal evaluation. *ERIC Digest*, Number 172 (ED 482347).
- Latham, G.P. (2012). *Work Motivation: History, Theory, Research, and Practice*. Thousand Oaks: Sage Publications.
- Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lipsomb, S., Chiang, H., & Gill, B. (2012). Value-Added Estimates for Phase 1 of the Pennsylvania Teacher and Principal Evaluation Pilot. *Mathematica Policy Research*. Reference Number: 06815.300. Retrieved from http://www.mathematica-mpr.com/~media/publications/PDFs/education/value-added_PAteachprin_tchrpt.pdf.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a Conception of How School Leadership Influences Student Learning. *Educational Administration Quarterly*, 46(5), 671-706.
- Leithwood, K., Seashore Louis, K., Anderson, S., & Walstrom, K. (2004). Review of

- Research: How leadership influences student learning. *University of Minnesota, Center for Applied Research and Educational Improvement and University of Ontario, Ontario Institute for Studies in Education*. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf>.
- Macmillan (2014). *Macmillan Online Dictionary*, definition of “professional development” retrieved from <http://www.macmillandictionary.com/us/dictionary/american/professional-development>.
- Marshall, C. & Rossman, G. (2016). *Designing Qualitative Research* (6th ed.). Thousand Oaks, CA: Sage.
- Massachusetts Institute of Technology (2016). Definition of S.M.A.R.T. goal setting. *Massachusetts Institute of Technology*. Retrieved from <http://hrweb.mit.edu/performance-development/goal-setting-developmental-planning/smart-goals>.
- Maxwell, J. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.
- Merriam, S.B. (1998). *Qualitative research and case study applications in education* (Revised and expanded edition). San Francisco: Jossey-Bass Publishers.
- Merriam, S.B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass Publishers.
- McNeil, M. (2010). All but 10 states throw their hats into the Race to the Top ring. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2010/01/20/19rtt_ep.h29.html.

- Miles, MB, Huberman, A.M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Mills, A., Durepos, G., & Wiebe, E. (2010). *Encyclopedia of Case Study Research*. Sage Research Methods Online. Retrieved from <http://srmo.sagepub.com/view/encyc-of-case-study-research/n224.xml>.
- New Leaders (2012). Principal Evaluation Handbook. *New Leaders, Inc.* Retrieved from <http://www.newleaders.org/newsreports/publications/principal-evaluation-handbook/>.
- Parylo, O., Zepeda, S.J., & Bengston, E. (2012). Principals' Experience of Being Evaluated. *Educational Assessment, Evaluation and Accountability*, 24(3), 215-238.
- Portin, B.S., Feldman, S., & Knapp, M.S. (2006). Purposes, uses, and practices of leadership assessment in education. Seattle, WA: *Center for the Study of Teaching and Policy, University of Washington*. Retrieved from <https://depts.washington.edu/ctpmail/PDFs/LAssess-Oct25.pdf>.
- Ravitch, S. & Carl, N. (2016). *Qualitative Research: Bridging the Conceptual, Theoretical, and Methodological*. Thousand Oaks, CA: Sage.
- Radinger, T. (2014). School Leader Appraisal - A Tool to Strengthen School Leaders' Pedagogical Leadership and Skills for Teacher Management? *European Journal of Education*, 49(3), 378-394.
- Robinson, V.M.J., Lloyd, C.A. & Rowe, K.J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational*
- Reeves, D.B. (2005). *Assessing educational leaders: Evaluating performance for improved*

- individual and organizational results*. Thousand Oaks, CA: Corwin Press.
- Administration Quarterly*, 44(5), 635-674.
- Rubin, H. & Rubin, I. (2012). *Qualitative Interviewing: The Art of Hearing Data*. Thousand Oaks, CA: Sage.
- Ryan, R. & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54-67.
- Ryan, R., Kuhl, J., & Deci, E. (1997). Nature and autonomy: Organizational view of social and neurobiological aspects of self-regulation in behavior and development. *Development and Psychopathology*, 9: 701-728.
- Sandelowski, M. (2000). Focus on research methods: Whatever happened to qualitative description? *Research in Nursing and Health*, 23, 334-340.
- Shelton S. (2013). Evaluating School Principals: A Legislative Approach. *National Conference of State Legislatures*. Denver, CO. Retrieved from <http://www.ncsl.org/documents/educ/SchoolPrincipals.pdf>.
- Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Stine, D.O. (2001). Developing an evaluation system to improve principal performance and accountability. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA. Retrieved from <http://files.eric.ed.gov/fulltext/ED452278.pdf>.
- Sun, M. & Youngs, P. (2009). How does district principal evaluation affect learning-centered principal leadership? Evidence from Michigan school districts. *Leadership and Policy Studies in Schools*, 8(4), 411-445.
- Sun, M., Youngs, P., Yang, H., Chu, H., & Zhao, Q. (2011). Association of District Principal

- Evaluation with learning-centered leadership practice: Evidence from Michigan and Beijing. *Educational Assessment, Evaluation and Accountability*, 24(3), 189-213.
- Superville, D. (2014). States forge ahead on principal evaluation. *Education Week*. May 21, 2014. Retrieved from http://www.edweek.org/ew/articles/2014/05/21/32principals_ep.h33.html.
- Thorne, S. Kirkham, S.R., & MacDonald-Emes, J. (1997). Interpretive description: A noncategorical qualitative alternative for developing nursing knowledge. *Research in Nursing and Health*, 20, 169-177.
- U.S. Department of Education. (2009). Race to the Top Program Executive Summary. Washington D.C. Retrieved from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.
- U.S. Department of Education. (2012). ESEA Flexibility Policy Document. Washington D.C. Retrieved from <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>.
- U.S. Department of Education. (2013). Idaho Condition Principle Three Letter. Retrieved from <http://www2.ed.gov/policy/eseaflex/secretary-letters/id2.html>.
- U.S. Department of Education. (2014). Race to the Top Fund. Retrieved from <http://www2ed.gov/programs/racetothetop/index.html>.
- U.S. Department of Education. (2016). Transitioning to the Every Student Succeeds Act (ESSA): Frequently asked questions. Retrieved from <http://www2.ed.gov/policy/elsec/leg/essa/faq/essatransitionfaqs050316.pdf>
- Wallace Foundation. (2009). Assessing the effectiveness of school leaders: New directions and new processes. *Wallace Foundation*. Retrieved from

<http://www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/Assessing-the-Effectiveness-of-School-Leaders.pdf>.

Wallace Foundation. (2013). The school principal as leader: Guiding schools to better teaching and learning. *Wallace Foundation*. Retrieved from

<http://www.wallacefoundation.org/knowledge-center/school-leadership/effective-principal-leadership/Pages/The-School-Principal-as-Leader-Guiding-Schools-to-Better-Teaching-and-Learning.aspx>.

Yin, R.K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.

York-Barr, J., Sommers, W.A., Ghere G.S., & Montie, J. (2006). *Reflective practice to improve schools. An action guide for educators* (2nd Edition). Thousand Oaks, CA: Corwin Press.

Zubrzycki, J. (2013). Principal Evaluations: Striving for Fairness. *Education Week*, March 6, 2013. Retrieved from www.edweek.org.

Appendix A

Summary of IDAPA121

Requirements under IDAPA code

IDAPA 121.01(Standards) requires all districts to adopt a principal evaluation model that aligns with State minimum standards based on the Interstate School Leaders Licensure Consortium (ISLLC) standards, and which address the three domains of school climate, collaborative leadership, and instructional leadership along with additionally prescribed components of each domain. Additional details of these three domains and associated components are discussed in subsection two below. IDAPA 121.01 further requires district evaluation models to include proof of proficiency in conducting teacher evaluations using the Charlotte Danielson framework, which proficiency must be demonstrated by September 1, 2018 by passing a one-time proficiency assessment approved by the State Department of Education.

IDAPA 121.02 (Professional Practice) requires the district also to use at least one of the following data sources in the evaluation of professional practice: parent/guardian input, teacher input, student input, and/or portfolios.

IDAPA 121.03 (Student Achievement) requires district evaluation models to include “multiple objective measures of growth in student achievement” to account for thirty-three percent of the evaluation result. Additional details of this student growth element will be provided in subsection three below.

DAPA 121.04 (Evaluation Policy – Content) requires all local school districts to document their principal evaluation policies and include the following information:

- a. Purpose

- b. Evaluation criteria
- c. Evaluator
- d. Sources of Data
- e. Procedure
- f. Communication of results
- g. Personnel actions
- h. Appeal
- i. Remediation
- j. Monitoring and evaluation
- k. Professional development and training
- l. Funding
- m. Collecting and using data
- n. Individualizing principal evaluation rating system; and
- o. A plan to include stakeholders

IDAPA 121.05 (Evaluation Policy – Frequency of Evaluation) requires each district to evaluate its principals on a fair and consistent basis and at least once annually, no later than May 1 of each year.

IDAPA 121.06 (Evaluation Policy – Personnel Records) requires each district to maintain a permanent record of each evaluation in the principal's personnel file, and to report the rankings to the State Department of Education on an annual basis for Federal and State reporting purposes. Individual ranking information must be kept confidential pursuant to State and Federal privacy regulations. Accordingly, the State Department of Education will not release statistical data on evaluation rankings for districts with fewer than five teachers.

In summary, IDAPA 121 requires all districts to evaluate their principals in a fair and consistent manner, at least once annually, with thirty three percent of the evaluation based on multiple measures of student growth and sixty-seven percent of the evaluation based on three domains of professional practice that were derived in large part from the ISLLC standards.

Appendix B

Proposed Interview protocol

Title: Reflective practice, professional development and Idaho's new principal evaluation program; a multi-case study

Investigator: Stephen Robertson

Participants Name: _____ School/District: _____

Date: _____ Time: _____ Location: _____

Introduction:

Good morning / afternoon. Thank you for taking time from your busy schedule to speak with me today. My name is Stephen Robertson and I am conducting research for my doctoral dissertation on the evaluation of principals here in Idaho and its effect on their reflective practice, professional development and leadership practices. As a public school principal here in Idaho, you have valuable experiences, perspectives and opinions to share on these issues that will contribute significantly to this research. I also hope that you find this interview process interesting and meaningful to your own leadership practice.

Before we begin, please know that all information you share with me will be confidential and anonymous. Your identity and that of your institution will be protected through the use of pseudonyms.

Please interrupt me at any time if you would like clarification about my questions or you want to restate or elaborate on anything from earlier in the interview. If you prefer to decline comment on any question, for any reason, simply tell me that is the case and we will move on.

This interview is being recorded to ensure that your answers are captured and understood completely and accurately, and is organized into seven sections. These are: demographic information, reflective practice, professional development, leadership practices, prior experience with evaluations, experience with the new evaluation process, and closing comments and questions.

Do you have any questions or comments before we begin?

A. Demographic information

1. Please begin by telling me a little bit about yourself (age, educational level)

2. How long have you been principal of X school?
3. Briefly describe your work experience in education prior to becoming principal here at X school.

B. Reflective Practice

1. Please describe what the term ‘reflective practice’ (RP) means to you in the context of your leadership of this school.
2. With that definition in mind, how frequently have you engaged in RP this current school year?
3. Please describe one representative example of your RP this year (what was the issue, when was it, how did it resolve)
4. What events or circumstances contributed to your engagement in that specific instance of RP?
5. Where there any other contributing factors?
6. [Repeat cycle of 3 – 5 until all representative examples of RP have been explored]

C. Professional Development

1. Please describe what the term ‘professional development’ (PD) means to you, personally, in the context of your leadership of this school.
2. With that definition in mind, do you have a PD plan (PDP) for this year?
3. What events or circumstances contributed to the development of that PDP?
4. How frequently have you engaged in PD this current school year?
5. Please describe one representative example of your PD this year (what was the issue, when was it, what did you do, who else was involved).
6. What events or circumstances contributed to your engagement in that specific instance of PD?
7. Where there any other contributing factors?

[Repeat cycle of 5 – 7 until all representative examples of PD have been explored]

D. Leadership practices

1. Please describe your leadership priorities for the current school year.
2. [Taking one priority at a time] what events or circumstances contributed to your focus on this issue?
3. [Repeat 2 until identified priorities have been explored].

E. Experience with prior evaluations

1. Please describe the performance evaluation process that you experienced [last year / two years ago] before your district used (1) the Idaho Standards of Effective Principals and (2) Student growth measures to evaluate your performance.
2. In what ways did this prior evaluation process contribute to your reflective practice?
3. In what ways did this prior evaluation process contribute to your professional development?
4. In what ways did this prior evaluation process contribute to changes in your own leadership practices? [ask for specific examples, probe for detail as needed].

F. Current evaluation

1. In what ways has your experience with your own performance evaluation differed from the prior experience?
2. In what ways has your experience been the same or similar?
3. In what ways has the new performance evaluation process contributed to your reflective practice this year?
4. In what ways has the new performance evaluation process contributed to your professional development this year?
5. In what ways has the new performance evaluation process contributed to changes in your own leadership practices this year? [ask for specific examples, probe for detail as needed].
6. Focusing for a moment on the use of student achievement data in the evaluation of your performance this year, please describe how adding this to your evaluation has effected your reflective practice and professional development.
7. Please describe how adding student achievement data to your evaluation has effected your leadership practices this year [[ask for specific examples, probe for detail as needed].

G. Closing comments or questions

1. Is there anything else you would like to add regarding your experience with the new performance evaluation process and its relationship to your own reflective practice, professional development or leadership practices?
2. Do you have any questions of me?
3. Would it be ok to contact you for more information or clarification if needed?

Thank you again for your time and valuable insights!

Appendix C

Informed Consent

Title: Reflective practice, professional development and Idaho's new principal evaluation program; a multi-case study

Directions: Carefully review each statement below and initial, in the corresponding blank space, to indicate that you have read and understood each statement. Please sign where indicated at the end of this consent form if you understand and agree to its contents and consent to participate in this study.

1. _____ I understand that the University of Idaho Institutional Review Board has approved this research study.
2. _____ The purpose of this study is describe how the promise of Idaho's revised process of principal evaluation compares with the actual experience of Idaho principals in terms of their reflective practice, professional development and the follow-on effects on their own leadership behaviors and practices
3. _____ You will be asked to participate in a series of one to three interviews in a setting of your choice. Each interview will last approximately ninety minutes and will be audio-recorded. You will have the right to stop recording at any time. Once the audio-recordings have been transcribed, the audio-recording(s) will be permanently deleted.
4. _____ The researcher will also gather information for this study by observing you and/or adult members of your school community in activities that relate to this study's themes of performance evaluation for principals, reflective practice, professional development and leadership practices. The researcher will also gather information for this study by reviewing documents that pertain to these same themes and may request that you provide copies of relevant documents.
5. Your time as a professional is valued. The time spent in interviews will be arranged according to your schedule and will stay within the scheduled time frame. In case the researcher is unable to keep the assigned meeting time, all efforts will be made to contact you in advance.
6. _____ Your identity and that of your institution will be protected through the use of pseudonyms to maximize anonymity and confidentiality. Once the interviews are transcribed, you will have the opportunity to check the transcript for accuracy, and to ensure that any information that identifies you or your institution is flagged and

removed. All reasonable efforts will be made to protect your information and maintain confidentiality to the extent permitted by law.

7. _____ Interview questions will be related to your general background and your experience with reflective practice, professional development, leadership practice and the experience of being evaluated in the current school year and in prior years. Some questions may cause you to feel uncomfortable.
8. _____ It is hoped that the findings of this study will increase our understanding of how Idaho principals are experiencing Idaho's new evaluation model for principals, and its relationship to their reflective practice, professional development and leadership practices.
9. _____ Your participation in this study is voluntary and you may withdraw from this study at any time for any reason. If you choose to discontinue your participation, then information that pertains to you will not be used in the study.
10. _____ You may be asked to review preliminary and final findings and conclusions in order to ensure their validity and reliability.
11. _____ If you have questions or concerns at any time regarding the study or your participation in it please contact the researcher or his major professor through the contact information provided below:

Contact Information

Researcher

Stephen Robertson
708 W. Valentino St.
Meridian, ID 83646
Phone: 208-484-1069
StephenRobertson1986@gmail.com

Major Professor

Dr. Kathy Canfield-Davis
University of Idaho – CDA Center
1031 N. Academic Way, Suite 242
Coeur d' Alene, ID 83814
Phone: 208-667-2588
Canfield@uidaho.edu

I have reviewed this consent form and understand and agree to its contents. By completing and signing this consent form I agree to participate in this study.

Participant name: _____

Participant Signature: _____ Date: _____

Participant phone number: _____

[] Please send me a copy of final report:

Address: _____

Thank you, but I do not wish to receive a copy of this study.

Researcher name: Stephen Robertson

Researcher Signature: _____ Date: _____

Appendix D

Second Cycle Codes

1. Timing (which related to when the principal interacted with anything related to the performance evaluation).
2. Focus of Evaluation (which related to the substance of the evaluation related activity).
3. Instrument Use (which related to how the district provided evaluation tool was used or not used during the activity and/or the participant's perspective on the tool).
4. Ratings in tool (which related to the participant's experience or perspective on the various rating levels within the evaluation tool).
5. Reflective practice – RP (which captured the participant's perspective and/or activities regarding reflective practice).
6. Professional development – PD (which captured the participant's perspective and/or activities regarding professional development).
7. Challenges (which related to difficulties or obstacles within the evaluation process).
8. Impact of evaluation (which related to if or how the performance evaluation process impacted participant's reflective practice, professional development and leadership practices).
9. What did guide you (which focused on what other factors impacted their reflective practice, professional development and leadership practices).

10. Theory of evaluation (which related to the participant's opinions regarding the evaluation process).
11. Positive aspects (which related to any positive aspects of the evaluation experience).
12. Unique aspects (which related to individual or unique circumstances that pertained to the evaluation process or experience).
13. Magic wand (which related to how participants would change the evaluation process if they possessed a magic wand).
14. What works well (which related to aspects of the evaluation process that appeared to be having a positive impact on the participant).
15. What would be better (which related to improvements to the existing process).
16. Growth data (which related to the participant's perspective on the use of student growth data as one component of their evaluation).

Appendix E

Analytic Memo (Composite Portrait)

Although generalizability from my study to other situations is not necessarily the goal of qualitative research (Merriam, 1998; Miles & Huberman, 1994), all participants in this study reported powerfully consistent experiences with (1) their performance evaluation, (2) its complete and total non-impact on their own reflective practice, professional development and leadership practices, and (3) in their prescription for what would deliver a truly meaningful professional development experience for school principals. These powerfully consistent experiences allow me to offer what Stake (1995) calls “naturalistic generalization” so that readers may “gain insight by reflecting on the details and descriptions presented in case studies ... and consider whether their situations are similar enough to warrant generalizations” (Mills, Durepos, & Wiebe, 2010). With this goal of facilitating a naturalistic generalization in mind, I offer the following synthesized portrait of the principals in this study and their experience with their own performance evaluation and its impact on their own reflective practice and professional development. To maintain gender neutrality in this synthesized description, the writing alternates between male and female descriptors.

The school principal is a dedicated, hard-working and thoughtful leader who is focused on creating the best possible learning environment for her students and teachers. He is confident in his abilities and has a clear sense of who he is as a leader and what his strengths and areas of opportunity are.

Despite this strong sense of self-awareness, she welcomes feedback from those who know her and are familiar with her work and acknowledges feedback is important to her ongoing professional development and performance. He is an avid consumer of data and uses

data to fuel his reflective practice and inform his leadership priorities and professional development agenda.

Because school data is often discussed in the context of teacher and school leader accountability, our principal had clear opinions on accountability. First, she does not fear accountability – in fact she may welcome it – provided it is grounded in valid and reliable data.

Perhaps the biggest challenge to using student test data to ‘hold principals accountable’ arose from the fact that none of the high stakes test results were available at the time of the end of year principal evaluations

He believes that the people responsible for the development of public school policy such as IDAPA 121 are out of touch with the reality of what happens in the school building.

Regarding reflective practice, our principal believes this is an essential part of his school leadership and professional development (very important, super important, and necessary for growth). She engages in reflective practice regularly, on her own or with her leadership team if she has one, and believes that reflective practice comes naturally - automatically even - for any principal that is motivated to help her school, students and staff. Reflective practice occurs daily, if not multiple times a day, in response to events and circumstances within the school and can occur in the moment in immediate response to such events, or in a more formal way combined with additional data. The focus of reflective practice varies across situations but usually involves some version of the following questions: “did I do the right thing”, “should I have handled that differently”, or “was this good for our students”?

Without exception, the principal evaluation process during the 2014-2015 school year had no impact or bearing on the principal's reflective practice, either at the beginning of the year when goals were being solidified and communicated, or during the course of the school year itself.

For our principals, reflective practice is what happens day to day, within and in response to the daily and weekly flow of events that occur during the school year. Principal evaluation, by contrast, is an isolated and inconsequential 'check the box' event that occurs at the end of the school year, long after the instances of reflective practice and their triggering events have occurred.

Our principal values his own professional development and knows it can help him to be a more effective leader. Several years ago, the district offered more formal development programs for principals. Recently, including the 2014-2015 school year, the district has provided very little formal development for principals – presumably because of budget constraints – and most professional development now takes the form of self-directed learning such as reading or taking graduate school courses. She wishes there could be more formal development from the district both for herself and for newer principals who could really benefit from a more structured approach.

As with reflective practice, the 2014-2015 performance evaluation process had no impact or bearing on his professional development, whether self-directed, informal or formal in nature. The performance evaluation was a complete 'non-issue' and non-factor in so far as her professional development was concerned.

Regarding our principal's own performance evaluation, it consisted of a single meeting with his evaluator at the end of the 2014-2015 school year. Within the X district, the

evaluation form provided three levels of performance: basic, proficient and exemplary, and the evaluator asked the principal which level was the best description of his own performance during the past year. The evaluator asked the principal for some evidence to corroborate his own self-assessment. The principal and evaluator reviewed some parts of the evaluation form, collaboratively discussing the most appropriate rating for those items discussed. The evaluator shared the overall performance rating that he intended to give to the principal, and they may have spent a few minutes discussing the principal's goals for the year and engaging in general small talk. The conversation could be described as professional but ultimately perfunctory in nature as though the true aim was simply to get it done so that the principal and evaluator could return to their real jobs. Within the Y district, the evaluation conversation was shorter and featured a general discussion about the principal's goals and some small talk. The district's evaluation form was not discussed and was not present during the conversation. In neither district did the evaluator and principal discuss student performance, nor did they address the 1/3 of the principal's evaluation that was to be based on measures of student achievement.

When our principal reflected on the nature and extent of his own performance evaluation, he couldn't help but notice how different his own experience with evaluation is compared with the experience that he provides to his own teachers and staff. Whereas she works diligently to be in every classroom on a regular basis to observe her teachers in action, her own evaluator rarely visits her school, and when he does visit the conversations focus not on the principal and her leadership but on other unrelated topics such as the roll-out of a new district policy. Whereas the principal endeavors to provide meaningful and developmentally oriented feedback to her teachers during their evaluations, her own evaluation feels more like

a check the box exercise. The general feeling is that his evaluator really doesn't know him or his work, and how could he, given the very little time the evaluator spends with the principal during the school year.

Our principal doesn't blame or resent her evaluator for this dynamic; she recognizes that her evaluator has many other priorities to focus on within the district, and assumes that her evaluator is operating under the principle of 'no news is good news' as a way to rationalize the infrequent contact and interaction between principal and evaluator. At the same time, our principal can't help but feeling under-appreciated at times as a result of this dynamic, especially when he reflects on how much effort he puts into his interactions with his own teachers and staff. Nonetheless, our principal has come to accept this dynamic as the reality she must work within, and she doubts that the dynamic will change in the future given the other priorities for her evaluator and the financial pressures that our schools continue to experience.

Appendix F

Summary of Findings Submitted to Participants
Performance Evaluation of School Principals in Idaho
 Student researcher: Stephen Robertson, Ph.D. candidate, University of Idaho

Research purpose: The purpose of this study was to describe the experience of ten Idaho public school principals with their own performance evaluation during the 2014-2015 school year and to understand the relationship between this performance evaluation and their reflective practice, professional development and leadership practices.

Research findings: Four primary findings emerged from this study:

1. **Principals interviewed for this study experienced their own performance evaluation during the 2014-2015 school year as a perfunctory, isolated, and end of year event that had no meaningful impact on their reflective practice, professional development or leadership practice.**

Subordinate findings:

- a. Principals interviewed value reflective practice and professional development as important for their own professional growth and performance
 - b. Reflective practice and professional development are disconnected from the current performance evaluation process for principals
 - c. Reflective practice occurs in response to events that occur within the school itself, and not as a result of the formal evaluation process
 - d. For principals interviewed, their leadership practice, goals and behaviors are driven by (i) what the data tells them, and (ii) their own sense of what their school, students, teachers, staff and families need, rather than by the evaluation process
2. **Principals interviewed for this study reported that the prospect of being evaluated partly on student performance had no impact on their leadership focus or practices.**

Subordinate findings:

- a. Principals interviewed dispute the notion that measuring them on student performance will increase their motivation or effort because they are already fully committed to improving student learning and teaching
 - b. Principals interviewed are not afraid of accountability per se provided the measurements used to hold them accountable are fair and reasonable
 - c. Principals interviewed question the validity of using current standardized test scores to measure the effectiveness of their leadership and believe many factors that influence student performance are outside of their control
 - d. As a practical matter, principals interviewed were not actually measured on student performance during the 2014-2015 school year because the data was not available in time of their evaluation
- 3. Principals interviewed believe that policy makers and school policy related to their performance evaluation are out of touch with what happens in schools.**
- 4. Principals interviewed had clear ideas for improving the system so principals would experience more meaningful reflective practice and professional development. They recommend the following:**
- a. It would be helpful if principals were oriented towards the district's performance expectations at the beginning of the year as this could help facilitate reflective practice, goal setting and professional development plans that align with district expectations
 - b. Because the current performance standards are deep and wide, this goal setting and professional development should focus on one or two priority areas chosen by the principal rather than trying to address everything in the standards
 - c. This start of year goal setting should be followed-up by a mid-year and end-of year review of progress made against the goals
 - d. Meaningful performance evaluations require a trusting relationship and feedback from people who actually know you and what you are doing in your school; because it is unlikely that area supervisors can spend enough time with each principal to achieve this, meaningful feedback can be obtained from teachers and staff who work with the principal every day

- e. High quality training should be available that aligns to the professional development priorities chosen by the principals
- f. Because principals could learn a lot from one another (in addition to, or in conjunction with, formal training), structured peer collaboration time to share best practices with one another should be sponsored and supported by the district
- g. It would be best if these peer collaboration sessions occurred with a homogenous grouping of principals (e.g. elementary principals only, rather than elementary + middle school + high school)
- h. It would also be best if the area supervisors (evaluators of the principals) did not attend these peer collaboration sessions in order to increase open and potentially vulnerable conversations among the principals
- i. To support the attainment of goals and professional development plans, structured mentoring or coaching would be helpful. This would be especially helpful for new / newer principals although more experienced principals might benefit as well

Appendix G

Member Check Results

Participant	Is your participant overview accurate?	Are the primary findings accurate?	Other feedback
Willian	Y	Y	Looks accurate. Thanks.
Nancy	No response yet	No response	n/a
Robert	No response yet	No response	n/a
James	No response yet	No response	n/a
Edward	Y	Y	Accurate and represent the sentiments I communicated to you when we met.
Mary	Y	Y	I found it to be extremely interesting. All of the information is accurate
Elizabeth	Y	Y	I think it looks great.
Scott	No response yet	No response	n/a
Charlotte	Y	Y	I fully agree with the contents of the principal profile. The summary findings don't surprise me, and they confirm my own position. How can this information be shared with policy makers?
Henry	Y	Y	You have captured this well