



2009

Principal Evaluation and Student Achievement: A Study of Public Elementary Schools in Dupage, Will, and Lake Counties, Illinois

Edward John Condon
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss



Part of the [Educational Administration and Supervision Commons](#)

Recommended Citation

Condon, Edward John, "Principal Evaluation and Student Achievement: A Study of Public Elementary Schools in Dupage, Will, and Lake Counties, Illinois" (2009). *Dissertations*. 110.
https://ecommons.luc.edu/luc_diss/110

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).
Copyright © 2009 Edward John Condon

LOYOLA UNIVERSITY OF CHICAGO

PRINCIPAL EVALUATION AND STUDENT ACHIEVEMENT:
A STUDY OF PUBLIC ELEMENTARY SCHOOLS IN DUPAGE,
WILL, AND LAKE COUNTIES, ILLINOIS

A DISSERTATION SUBMITTED TO THE
FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN EDUCATIONAL ADMINISTRATION AND SUPERVISION

BY

EDWARD J. CONDON, III

CHICAGO, ILLINOIS

DECEMBER 2009

Copyright by EDWARD J. CONDON, III, 2009
All rights reserved

ACKNOWLEDGEMENTS

I would like to express deep gratitude to a number of individuals who had great impact upon the completion of this research study.

For her tremendous support and guidance during the entire dissertation process, Dr. Martha Ellen Wynne deserves a great deal of thanks. She was quick to share her knowledge and subtle humor in a genuine and patient manner, and her forthright approach was critical in ensuring that the project always continued to move forward. Her experience and insight are highly respected and valued.

Throughout all of the stages that led to the completion of this project, Dr. Beverly Kasper consistently served as a thoughtful, supportive, and reflective advisor. Over the years she has acted as an advocate and mentor, and her guidance has been very special.

It is also important to thank Dr. Marla Israel for the time she spent reading and commenting on the manuscript. These were helpful and vital contributions to the project.

Donna Hattendorf was instrumental, as well. Her friendship and assistance meant a great deal. Likewise, I would like to extend my gratitude to Dr. Sandra Martin for her understanding and flexibility over the period of the dissertation project.

Finally, I would like to thank and acknowledge my supportive family and parents. Most importantly, it was the particularly loving encouragement and patience from Gina, Eddie, and Emily, that made “the big D” an achievable endeavor.

To Virginia

“By the work one knows the workmen.”

- Jean De La Fontaine (1621–1695)

PREFACE

Within each profession, standards exist that identify the boundaries of excellent performance. Whether these standards are clear and uniform, or implicit within the results of that which has been accomplished, they should be used as part of the performance evaluation process to measure both the effectiveness of the individual as well as the quality of work that has been achieved.

In any given field, true professionals value the use of standards and evaluation for the purposes of achieving desired outcomes, advancing effective communication with others, and improving and understanding their own professional practice.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
PREFACE	vi
LIST OF TABLES	ix
ABSTRACT	x
CHAPTER ONE: Introduction	1
Critical Roles of Elementary School Principals	4
Purposes of Principal Evaluation	7
Methods of Principal Evaluation in Common Use	11
Statement of the Problem	15
Significance of the Problem	18
Research Design	21
Limitations of the Study	23
Bias of the Researcher	24
Definition of Terms	24
Summary	26
CHAPTER TWO: Literature Review	27
Introduction	27
Leadership and Principal Evaluation	28
Supervision and the Principal	35
Problems with Principal Evaluation Systems	40
Improvement of Principal Evaluation Systems	46
Development of Effective Principal Evaluation Systems	55
ISAT, NCLB, and School Accountability	63
Sanctions, Rewards, and Evaluation	68
Principal Perceptions of Evaluation	71
Constituency Perceptions of Principal Performance	77
Principal Evaluation and Pupil Achievement	82
Summary	90
CHAPTER THREE: Methodology	94
Introduction	94
Research Design	96
Research Site	98
Participants	99
Procedures for Data Collection	100
Data Analysis	102

Ethical Considerations	103
Prevention of Bias	104
Validity and Limitations	105
Summary	106
CHAPTER FOUR: Analysis of Data	108
Accuracy of Findings	109
Description of Survey	110
Description of Sample	111
Descriptive Data – Demographics	112
Data Reduction	119
Data Analysis	124
CHAPTER FIVE: Discussion	137
Review of the Study	137
Conclusions	145
Recommendations for Future Research	155
APPENDIX A: Survey Instrument	158
REFERENCES	164
VITA	171

LIST OF TABLES

Table	Page
1. Matrix Display of Sample Size	111
2. Matrix Display of Response Rates	112
3. Descriptive Statistics, Principal Years at Current School	113
4. Descriptive Statistics, Total Years of Principal Experience	114
5. Matrix Display of Years of Principal Experience	115
6. Matrix Display of School Structure of Principals' Employment	116
7. Matrix Display of Principal Evaluation Frequency	117
8. Matrix Display of Evaluator Title Frequency	118
9. Matrix Display of Schools in District	118
10. Matrix Display of Evaluation Method Intensity	125
11. "Evaluation Methods" Inter-Item Correlation Matrix	127
12. Correlation of Evaluation Methods and Qualitative Principal Perception Data	131
13. Matrix Display of Perceived Evaluation Objectives	134

ABSTRACT

The purpose of this study was to investigate the topic of the evaluation of school principals and the role it plays in fostering student achievement. The overarching research question for this study was the following: “Is there a relationship between the type and intensity of principal evaluation and pupil performance in DuPage, Will, and Lake County, Illinois public schools at the K-8 level?”

While there are a number of purposes for principal evaluation, the tools and methods used to evaluate principals vary tremendously by both state and school district. This study addressed the degree to which principal evaluation affects students’ achievement, as well as the principals’ perceptions about the impact of evaluation on student achievement and their own professional practice.

To supplement the primary research question, the kinds of evaluation that are identified were also correlated with the principals’ perceptions as they pertain to two specific topics. The first topic addressed the principals’ perceptions about the influence of principal evaluation on pupil performance. The second topic addressed the principals’ perceptions about the influence of principal evaluation on their own professional performance.

This study used quantitative research in order to accomplish the first objective of determining the type and intensity of the methods that are presently in use for the purpose of principal evaluation. An on-line survey instrument was used to collect the data. The second objective was to correlate the identified evaluative methods with pupil

performance, which were based on Illinois Standards Achievement Test (ISAT) school scores for the previous year. As this research study employed simultaneous mixed methods, open-ended questions were nested within the larger survey that allowed respondents to indicate their perceptions about the effect of the type and intensity of principal evaluation on student achievement, as well as their own professional performance. The data that were acquired from this additional component added richness and validity to the study.

The main constructs for the study included the following:

- The identification of the type/intensity of principal evaluation methods that are in predominant use in DuPage, Will, and Lake Counties
- The relationship between school performance level on the ISAT assessment and type/intensity of principal evaluation
- The relationship between principals' perceptions about the influence of principal evaluation on pupil performance and type/intensity of principal evaluation
- The relationship between principals' perceptions about the influence of principal evaluation on their own professional performance and type/intensity of principal evaluation

CHAPTER ONE INTRODUCTION

The component of personnel evaluation is an integral part of the landscape of contemporary educational practice. Although classroom educators have been systematically evaluated for years, “the last two decades have ushered in a dramatic change in the entire concept of evaluation” (Rebore, 2004, p. 191). Across the nation, educational faculty and staff members are being evaluated in much more depth, and with more consistency, than ever before. In 1996, Brown and Irby stated that “new challenges have driven educators to search for innovative and appropriate methods of evaluating professional performance” (p.3). Since then, the introduction of improved evaluation plans and research based personnel practices have markedly changed the manner in which most school employees have been evaluated, and this has changed the landscape of American education.

However, the matter of school principal evaluation has been different. Even today, the tools and methods used to evaluate principals vary tremendously by both state and school district. In the State of Illinois, principal evaluation is mandated by the School Code, which states that “School boards shall ensure that their principals are evaluated on their instructional leadership ability and their ability to maintain a positive education and learning climate” (Illinois Association of School Boards, 2006, p. 257). Yet the manner in which the evaluation should occur, and even the individual(s) specified to conduct the evaluation, are considerations that are not directly addressed in the School

Code, are very ambiguous, and are highly inconsistent from one school district to another.

In some ways, it should not be surprising that school principals have received limited evaluative scrutiny for so long. As the individuals who have historically been accountable for implementing the increasingly elaborate faculty and staff appraisal systems, administrators have traditionally been shielded through their responsibilities as “evaluators.” Still, recent acknowledgment regarding the importance of principals to student academic performance helps explain the trend for increasing focus on principal evaluation.

McCleary (2001) states that “current theory holds that the primary purpose of evaluation is to establish a basis for change of individual behavior such that both personal satisfaction and organizational effectiveness is improved” (p. 46). Following this assertion, the main rationale for principal evaluation must then be two pronged: improvement of both the school and individual professional performance. This concept is similar to the dual impact that effective teacher evaluation has on both personal teaching practice, as well as the subsequent academic success of the students in the teacher’s classroom.

The advent and expansion of the trend toward increasingly common principal evaluation has not been without issue, however several unique challenges have surfaced. First, “principals and their evaluators tend to overlook the developmental nature of learning when they reflect on their own growth and development” (Connecticut State Dept. of Education, 1990, p. 7). This phenomenon accounts for a great deal of frustration when potential weaknesses are identified, and remedy is subsequently sought or expected

with a degree of immediacy. Another challenge stems from the fact that “evaluation of the principal in his job is intensely emotional and personal” (McCleary, 2001, p. 45).

This phenomenon is inevitable in occupations where individuals tend to become highly invested in their work, but the particularly public nature of the principal’s work magnifies flaws and personal shortcomings in the mind of those evaluated and adds a challenging aspect to the process.

Making matters more complicated is the observation that many principal evaluation systems do not successfully address the original purposes for which they were initially implemented. Anderson (1991, p. 77) states that “the evaluation methods used by many districts are not designed to enhance principal performance, but to satisfy accountability requirements that make principal evaluation mandatory.” In less-than-ideal circumstances, these perfunctory approaches to the important task of evaluation can breed resentment, frustration, confusion, and a lack of self-efficacy.

With a unique system of locally organized school divisions, the State of Illinois is an especially intriguing case. While the Illinois School Code and the Illinois State Board of Education provide considerable uniformity of educational practice, this consistency does not always trickle down to the local education agency (LEA). At present, the State of Illinois has 896 separate school divisions, each operating under their own set of local norms and cultures. When pondering the issue of principal evaluation under such atypical conditions as those in Illinois, important questions begin to form. In what manner are principals evaluated? With what intensity are principals evaluated? Is there a relationship between principal evaluation and pupil performance? What are principals’ perceptions about the influence of principal evaluation as it pertains to pupil

performance? What are principals' perceptions about the influence of principal evaluation on their own professional performance?

Critical Roles of Elementary School Principals

In order to consider issues relevant to the evaluation of principals, it is vital to identify the main roles that school principals are expected to play. As this study addressed principal evaluation through the lens of the K-8 elementary principalship, the primary roles for individuals working within this context must be delineated and explained. Though effective principals play a wide range of roles, many of which vary with the context of the specific nature of the school program, there are several vital roles that must be clearly identified in order to further understanding about the essential tasks on which a school administrator is evaluated.

In both theory and practice, the main role of the elementary principal is that of instructional leader. In the State of Illinois, the employment contracts for school principals are mandated to reflect this premise by requiring that the majority of a principal's time be spent in this area. Still, Catano and Stronge (2006) assert that "more often than not, the demands are different and may even be at odds with one another (p.222)." Though the activities that comprise the realm of instructional leadership are many and varied, essential tasks would include such things as curriculum management and development, evaluation of instructional personnel, and staff development. Fredericks and Brown state that "the instructional leadership role is one that provides a crucial link between the principal's activity and the effectiveness of his or her school" (1993, p. 11). Reeves (2004) builds upon this idea by emphasizing the need for principals to be given the authority to act responsibly as instructional leaders (p. 8).

Unfortunately, there is some role confusion for many principals about the degree to which they actually perform in an educational leadership capacity, versus the degree to which they should. When surveyed about their feelings on the matter principals shared frustration, stating that “the principals’ time was too heavily driven by job demands rather than educational goals, (and) that too much time was spent on student behavior and district office issues” (Fredericks & Brown, 1993, p. 10).

School principals have another vital role to play as organizational managers. While this role garners much less flash and attention than that of instructional leadership, the management component of a well-run school is critical. “Management, broadly defined, means the organization of people and processes to accomplish a goal” (Catano & Stronge, 2006, p. 225). In practical terms, the management role for an elementary principal can include anything from oversight of the facility, to coordinating events and activities, to ensuring school security and safety, to ordering necessary supplies and materials. Though management does not typically entail the tasks that are usually viewed as integral to the improvement of curriculum and instruction, no school can function effectively unless these tasks are performed. As “good management requires consistency and assurance that daily operations will be handled fairly and expeditiously” (Catano & Stronge, 2006, 225), management tasks often represent the bulk of that which has traditionally been appraised as part of most principal evaluation systems.

Another critical role for school principals is that of parent and community liaison. Traditional school organizations have evolved a great deal, and principals have an increased role in two-way communications with both of these important constituencies. This is particularly evident with parent groups. Rallis and Goldring emphasize that “the

boundaries of schools are very permeable, and principals, as boundary-spanning incumbents, are forced to link the internal functioning of the school and the environment” (1993, p. 7). In more contemporary work, they further develop this idea by advocating for principals to welcome and accept their changing roles (Rallis & Goldring, 2000). With increasingly open school borders, effective principals understand and embrace the role of liaison. However, they also understand the significant investment of time and energy that is necessary to play this role with competence.

Still another important role for elementary principals is as a provider of student needs. Student needs are highly diverse from one school setting to another, and from one child to another. Included within this role is the responsibility for principals to provide for appropriate student discipline, child-centered support programs, and extra-curricular opportunities. Yet the obligations for principals in meeting various student needs has become increasingly expansive. Stronge and Catano state that “public messages, illuminated by the media, indicate that schools should also pay attention to violence prevention, bullies, and the emotional needs of their students” (2006, p. 222). While these items are of clear significance, they are also evidence of the burgeoning manner in which principals have an essential role in providing for their students’ well being as individuals.

In a cogent discussion about principals’ roles, it is important to acknowledge the issue of role conflict. Role conflict can occur with any individual in almost any setting, and it is characterized by an inconsistency between the perceived role of the individual, and the role that is dictated by the social group (or organization). Just as with individuals in other professions, school principals can fall victim to role conflict quite easily.

However, the highly public nature of school administration can cause cases of role conflict to have far-reaching implications for both the individual and the larger organization. Duke and Iwanicki emphasize that “to the extent that role conflict cannot be minimized, there is likely to be a lack of fit between the principal and the setting in which the principal is working” (1992, p. 31).

Clearly, a “lack of fit” will have potential consequences of a measurable nature when considering the implications for performance evaluation. The problem for principals of course, is that they must be able to quickly and accurately discern the substantive expectations that others have for them in their role. “Effective principals are good at scanning the school environment and identifying what their constituencies really expect them to do. (But) these real job expectations often are not included in the principal’s job description” (Duke & Iwanicki, 1992, p. 31). This point is a powerful one when contemplating the issue of principal evaluation. Evaluation is generally considered effective when it is closely linked to an accurate description of the work that should be accomplished. When the job description is not consistent with job expectations, it becomes very difficult to accomplish authentic performance appraisal.

Purposes of Principal Evaluation

In order to establish uniformly high performance standards, the Interstate School Leaders Licensure Consortium (ISLLC) has established national standards for school leaders. These standards address the full range of administrative leadership skills that are necessary in an effective school setting. The ISLLC standards are highly useful in that they “provide a credible foundation for evaluation since they reflect a broad professional consensus on essential leadership skills” (Lashway, 2003, p. 4). In addition, “various

states (have) established performance standards for principals at the state level, and localities followed with attempts to standardize the work of principals through the development of evaluation instruments” (Catano & Stronge, 2006, p. 222). Though the ISLLC standards will be discussed in more depth in Chapter 2, their existence (as well as the need for state-level standards) helps remind us that performance appraisal is now being employed to achieve both individual and organizational purposes. Unless the link between the evaluation instrument and the purposes for it are clearly established, the evaluation process becomes inconsequential. As such, it is vital to identify the primary purposes that principal evaluation is intended to address.

Many purposes of principal evaluation exist. The most commonly identified purpose is to serve the objective of summative evaluation. Anderson (1991) states that “(summative) evaluation serves as an end, a judgment of performance on which to base principals’ retention, promotion, demotion, incentive pay rewards, and other personnel actions” (p. 79). While summative evaluation is used to ascertain the degree to which a principal has met the expected requirements of his or her position, “evaluations can be conducted for other summative purposes such as certification and licensing” (Thomas et al, 2000, p. 216). Properly implemented principal evaluations do not simply discern whether tasks or obligations have been fulfilled. They should also “serve to document the level of effectiveness with which principals execute their job responsibilities” (Catano & Stronge, 2006, p. 231). In these cases, the evaluations are being used to measure the essential concept of competence. Thomas et al (2006) also refer to the degree to which these evaluations can be used to further district objectives (p. 216).

Another facet of summative evaluation stems from the need for educational

organizations to maintain accountability. “In many jurisdictions, the primary purpose is to fulfill the need for the annual evaluation required by school boards” (Thomas et al, 2000, p. 216). Regrettably, the need for accountability (combined with time constraints) may leave missed opportunities by the wayside if evaluation does not provide for individual professional growth. Anderson (1991) states that “many principals contend the only feedback they receive is summative evaluation” (p.80). Still, summative evaluation is a necessary task. As Catano and Strong (2006) assert, “evaluation instruments are powerful tools for influencing the behavior of principals, reinforcing the adage that ‘what gets measured is what gets done’ ”(231).

The need for professional development reflects another purpose of evaluation: providing for formative growth. “Formative purposes relate to expected improvement of principals’ performance following evaluation and identification of ways in which principals can change their administrative style and improve their skills, attitudes, and knowledge” (Thomas et al, 2000, p. 216). In helping principals to develop professionally, it is hoped that formative evaluation can be useful in moving principals to a new role as school administrators. As “the new role of the principal appears to be one of facilitating, collaborating, and consensus building” (Green, 2001, p. 4), it is critical that formative tools are used to help move toward this goal.

In order for evaluation to be utilized in a formative manner, it is necessary that there be opportunities for dialogue between the principal and his or her supervisor. “The (formative) evaluation process requires on-going communication between superiors and principals, with the focus on improving not only principal performance, but the overall educational program as well” (Anderson, 1991, p. 79). This communication process can

also provide essential chances for the supervisor to get to know the leadership styles and habits of the principal, which is vital in determining the principal's appropriateness for a given school community. Thus, one of the additional purposes of using formative evaluation is when a principal's "suitability for a particular principalship needs to be ascertained" (Thomas et al, 2000, p. 317).

The purpose of leadership development is another formative goal of principal evaluation. The process of discourse between the principal and the supervisor is valuable for providing the key elements of growth and leadership development, which do not often come easily. In practice, "growth occurs through highly developed networks of relationships and communication – creative lateral connections with supportive individuals and groups, intelligent accountability in vertical relationships – that enhance the capacity of those in the system" (Canty et al, 2006, p. 10). As such, leadership development and growth can only be fostered and sustained in an environment where evaluation possesses a collaborative formative component. Andrews (1990) states that "administrators who do not regularly receive expert, constructive criticism and suggestions are denied important opportunities to expand their horizons, they are invited to stagnate" (p. 3).

The principal evaluation process has the additional purpose of helping to further the objectives of the school division, when the process and evaluation tools are well aligned with the district objectives. "Evaluation instruments . . . communicate the intentions and values of the school districts that author them" (Catano & Stronge, 2006, p. 231). Further, "one would expect the orientation of evaluation instruments to push principals' behavior toward the implicit values of the instrument" (Catano & Stronge,

2006, p. 227). If there is a regular procedure in place to modify and update the evaluation instrument, then the evaluation process may also be tied to school improvement objectives. The Connecticut State Department of Education (1990) asserts that one of their purposes for principal evaluation is “to promote the improvement of school programs and the enhancement of student learning” (p.5).

Finally, principal evaluation serves the purpose of clearly identifying the performance expectations that must be met. As “it is natural that principals would aspire to meet the job expectations framed within evaluation instruments” (Catano & Stronge, 2006, p. 227), the expectations that are outlined by the evaluation process should be clear and standards-based, and the instruments carefully designed. “Evaluation instruments are powerful communication tools that serve to articulate the responsibilities deemed important for principals to execute” (Catano & Strong, 2006, p. 231). When these responsibilities are worthy of merit, effective principal evaluation instruments also have a mechanism to document when “the principal has exhibited highly meritorious service that deserves special recognition and reinforcement” (Shufflebeam & Nevo, 1993, p. 25). In the end, the most important consideration is that “an evaluation process needs to be in place which will promote the priorities of the district and will model collaborative school improvement” (Stine, 2001, p. 3).

Methods of Principal Evaluation in Common Use

A wide variety of principal evaluation methods are presently in common use, although several general categories exist. Fortunately, “the desired qualities of a principal evaluation system have been expressed in the literature” (Amsterdam et al, 2003, p. 222). Though the most common methods will be described herein, one of the greatest problems

in the collection of evaluation data stems from the fact that “many principals report being unaware of what information is collected on their performance and how it is obtained” (Anderson, 1991, p.85). This problem is confounded by the frequent occurrence of supervisors providing principals with a perfunctory performance evaluation, a practice that Thomas, Holdaway, and Ward refer to as “ritualistic” evaluation (2000, p. 227).

One general category of principal evaluation methods is self-evaluation. Self-evaluation methods are based on the premise that the principal should be the one (at least in part) to collect the data. Lashway identifies these types of appraisals as “free-form” evaluations (2003, p. 2), but they vary widely in design. Portfolios are likely the most common type of self-evaluation, and they are popular because they “combine both formative and summative evaluation” (Brown & Irby, 1995, p. 1). Portfolios are “conceptual containers into which principals can place a wide variety of artifacts documenting their achievements (Lashway, 2003, p. 5). Maintained by the principal, the main benefit of portfolios is “that the self-assessment and reflection inherent in portfolio development promote administrator growth, which leads to improved performance and, ultimately, to improved schools and learning” (Brown & Irby, 2001, p. 1). Other types of self-evaluation include descriptive narrative (anecdotal), as well as the use of dossiers. Like portfolios, dossiers include “a broad range of sources that can be collected to evaluate principals: attendance and test records, committee reports, newsletters, clippings, and time logs” (Anderson, 1991, p. 4). Whatever type of self-evaluation method is used however, it is important that the evaluation information is discussed in tandem with the evaluator. “By going over the self assessment with others, the principal’s self-evaluation may be illuminated and improved” (Ediger, 2001, p. 4).

The most prominent category of principal evaluation is defined by rating or checklist methods. The specific instruments themselves vary tremendously, but “by far the greatest number of assessment instruments are home-grown, typically taking the form of checklists” (Lashway, 2003, p. 5). These instruments are used by the evaluator to identify the degree to which certain pre-identified performance indicators are present in the principal’s performance. “Often, such forms are made up of items relating to what are viewed to be important behaviors . . .” (McCleary, 2001 p. 48). A useful illustration of a checklist system can be found in a representative instrument created by the State of Oklahoma, titled “Oklahoma Criteria for Effective Administrative Performance.” (Oklahoma State Department of Education, 1999, p. 21). This instrument is applied to rate principal performance in three distinct areas: Administrator Management Indicators, Instructional Leadership Indicators, and Administrator Product Indicators (Oklahoma State Department of Education, 1999, p. 21).

The data used to complete these ratings can be collected in several ways; one way is by “recording specific statements and actions made during onsite visits where (supervisors) shadow the principal for extended periods” (Anderson, 1991, p. 86). Other methods of data collection include observation, review of student achievement performance, and critique of communication documents. However, when checklists and rating systems are applied for the purpose of principal evaluation, “the interpretation and use of such information is the critical factor” (McCleary, 2001, p. 48) toward their effectiveness. Unfortunately, “many superintendents and supervisors do not directly observe the performance of principals” (Anderson, 1991, p. 86) before completing the evaluation instrument. The result is that “most evidence for principal evaluation derives

only from superiors' perceptions of how principals perform rather than from direct observation" (Anderson, 1991, p. 86).

Another general category of principal evaluation methods is based on client-centered feedback. Clients such as parents, teachers, and students can be culled for information about their perceptions of the principal's performance. In 1991, Anderson stated that "fewer than 1 percent (of districts) reported that they systematically collect teacher, parent, or student perceptions of principal performance" (p. 87). However, this approach is being advocated more often today. Ediger asserts that "school principals periodically should be evaluated by teachers," and that questions about effectiveness "need to be answered by those involved under the principal's leadership" (1998, p. 546). Surveys or questionnaires are generally utilized for the purpose of gathering this information from the respective constituent groups. "Surveys of teachers, support staff, students, and parents can provide quantifiable evidence for key aspects of the principal's job" (Anderson, 1991, p. 4). Because of its limitations, client-centered feedback is typically used in concert with other evaluation methods. Murphy and Pimental (1996) describe an evaluation process that uses a point system in which "data on academic outcomes and results from teacher, parent, and student surveys regarding the principal's performance make up the bulk of the points" (p. 76).

Finally, a less common type of principal evaluation is peer supervision. In methods of peer supervision and review, principals rely on their colleagues to provide them with feedback relative to their performance. Though occasionally used in tandem with other evaluation methods, "peer supervision/evaluation is an innovation that has not yet achieved wide acceptance" (Anderson, 1991, p. 87).

In most cases, the individuals performing the evaluation are the principal's supervisors, who are almost always central office personnel. Sometimes, the evaluation is completed by more than one supervisor (by two different assistant superintendents, for example). Gil states that this kind of situation can be challenging however, because multiple evaluators may have "differing standards and expectations" (1998, p. 28) for principal performance.

Statement of the Problem

The research question in this study, directed at K-8 principals in a three-county area outside of Chicago, Illinois, primarily addresses the issues of principal evaluation and student achievement. The primary research question was as follows: "Is there a relationship between the type and intensity of principal evaluation and pupil performance in DuPage County, Lake County, and Will County, Illinois public schools?" Clearly, the manner in which principals are actually evaluated is a vital consideration in the investigation of this matter.

Several secondary questions also exist in this study. The first of these addressed the relationship between school performance level on the 2008 ISAT assessment and the type and intensity of principal evaluation. The second question addressed the relationship between principals' perceptions about the influence of principal evaluation on pupil performance and the type and intensity of principal evaluation. The third question investigated the relationship between principals' perceptions about the influence of principal evaluation on their own professional performance and the type and intensity of principal evaluation.

Unfortunately, the current research indicates great variability on the type and

intensity of principal evaluation in contemporary use. Davis & Hensley (2000) assert that “the literature suggests that principal evaluation processes are anything but clear, systematic, or purposeful” (p. 388). Marcoux et al (2003) state that “evaluation techniques of principals have not changed much during the last decade; conventional procedures are still widely used and evaluations are expressed in the form of checklists, scales, and descriptive assessments” (p.4). They further attest that “despite recent advances in the quality of teacher evaluation, performance evaluation for principals remains poorly thought out and largely ineffective” (Marcoux, Brown, Irby & Lara-Alecio, 2003, p. 4).

The problem of poorly used evaluation systems is exacerbated by the unique and political nature of the principalship. Davis & Hensley (2000) emphasize that “it is not uncommon for principals to find themselves caught between issues and situations in which demands from constituents are incongruent with those of the district office” (p. 386). This scenario can be highly challenging when the principal must meet practical objectives that are inconsistent with those of the supervisor (who typically serves as evaluator). Lashway emphasizes that “many evaluation instruments treat leadership skills as “binary” traits that either exist or do not exist, whereas in reality many skills fall along a continuum” (2003, p. 3). Principal evaluation instruments and systems must also consider the complexity of the principalship when appraising performance. While some evaluation systems attempt to distill performance into a simplified checklist, “this implies that an employee can be effectively evaluated only within the context of attaining certain pre-established objectives” (Rebore, 2004, p. 192).

Additionally, some evaluation systems have not been properly designed to reflect

the goals and priorities of the larger organization. Cases exist in which school divisions have elected to employ principal evaluation systems that do not reflect alignment with the district's objectives. In these cases, "evaluation systems void of clearly articulated and communicated purposes are meaningless and contribute minimally to the accomplishment of the goals of the organization" (Davis & Hensley, 2000, p. 388).

Lastly, some evaluation systems inspire doubt as to whether the ethical issues inherent to the consideration of human dignity are receiving proper acknowledgment. As both a transactional and human process, personnel evaluation must be contemplated through an ethical lens. After all, ". . . the importance of human dignity implies that ethical analyses apply to both substantive work rules and outcomes on the one hand, and the governance and formative processes by which they are developed on the other" (Budd & Scoville, 2005, p. 9). Any evaluation system that does not reflect an accurate (and thus holistic) view of the individual being appraised is questionably ethical, as it violates the primacy of human dignity.

These issues help to define the nature of the problem with the present state of contemporary school principal evaluation. As to whether there may be a relationship between the type and intensity of principal evaluation and pupil performance has still to be fully determined. Yet, the fact remains that there exists significant variability in both the type and intensity of principal evaluation, and this fact make it quite difficult to ascertain the effectiveness of principal evaluation. In order to be most effective, "principal evaluation should reflect the proficiencies necessary to measure the competencies of the principal" (Marcoux, Brown, Irby & Lara-Alecio, 2003, p. 2) in a consistent manner, and within the contextual dynamics of the role. Still, the question

remains: “Is there a relationship between the type and intensity of principal evaluation and pupil performance in DuPage, Lake, and Will Counties, Illinois public schools?”

Significance of the Problem

The research question is significant for several reasons. Schools have functioned with personnel evaluation systems for many years. However, the focus of these formal evaluation systems has traditionally been on faculty and staff members. School leaders have often been omitted from these evaluation systems, presumably because they were the ones responsible for the implementation of the evaluation program in the first place. In addition, “administrative appraisal practices are often ad hoc in nature and suffer from a lack of consistency” (McAdams & Barilla, 2003, p. 19). In a study conducted by Langer and Boris-Schacter (2003), principals “indicated they were frustrated by the discrepancy between what they think their role should be and the reality of their work” (p.16). For some, a lack of thoughtful and appropriate evaluation may contribute to this role confusion. Many human resource specialists assert that school principals should be evaluated in the same depth as faculty and staff members, both for reasons of professional growth and less often for possible remediation/dismissal. The issue of professional growth is of particular importance, as it “enables principals to refine leadership practices and increase school effectiveness” (Brown & Irby, 2001, p. 4). This change in evaluation focus represents a positive advancement in the profession of school administration, as well as great progress toward improving the overall quality of individual principals in our schools.

Unfortunately, the actual act of providing school administrator evaluation is highly inconsistent at best. Reeves (2004) states that “more than 18 percent of leaders we

studied had never received an evaluation in their current position” (p. 2). In addition, many leaders who are provided with some type of performance evaluation are not always able to extrapolate meaning for themselves, their schools, and their careers. This issue is of particular concern in the case of ineffective principals. In such cases, performance evaluations are often created and utilized to identify mistakes that have been made.

These instruments do not effectively address what may actually be a larger problem for ineffective leaders: lack of action. Interestingly, Davis (1998) states that it “appears that principals more often lose their jobs for reasons of omission rather than commission” (p.66). This creates a compelling case against the problem of poorly instituted administrative evaluation procedure. Even in situations where evaluation is standardized and objective driven, the process may not be robust enough to help an administrator improve his or her practice. In addition, the degree to which poorly instituted evaluation procedure might assist principals to increase the effectiveness of school performance would be limited at best. For these reasons, Brown and Irby (2001) assert that “current evaluation systems generally do not result in significant growth for the principal or the campus” (p.5), and that a new vision of evaluation is needed.

The research question is also significant in that no uniform method has yet been established for the task of evaluating school principals. It is important therefore to discern both the type and intensity of the evaluation procedures in current use. Not surprisingly, there does seem to be commonality about the emphasis for most principal evaluation systems; “they highlight congruence and conformity – loyalty to superiors and personal appearance” (Hart, 1992, p. 41). Yet the question must be asked about whether these concerns should be the critical focus for tools that measure the effectiveness of

educational leaders. A further problem is identified in the fact that the competencies measured in most methods of principal evaluation “tend to be descriptive and perceptually-based” (Hart, 1992, p.38). This allows for a phenomenon that Hart (1992) describes as the “fudge factor,” in which latitude is provided for the evaluator stemming from a lack of outcomes-based accountability measures. When it comes to the subject of principal evaluation, it is accurate to state that “the primary problems are poorly defined standards of leadership and undefined standards of performance” (Reeves, 2004, p. 4).

Unlike the realm of teacher evaluation, in which standardized methods are highly consistent from one school district to another, principal evaluation varies substantially in method. Those principals who are routinely assessed on effectiveness and job performance may be evaluated through one of many different methods, including anecdotal evaluation, performance checklist, observation of practice, self-evaluation, portfolio evaluation, or another method. Yet, regardless of method, many principals feel that evaluation systems are “... oriented to obsolete procedural checklists, are inconsistent and informal, and inhibit open communication and dialogue between evaluators and principals” (Brown & Irby, 2001, p. 5). Clearly, the intensity of principal evaluations varies greatly. In some cases, the evaluation instruments are extensive and utilized with a high degree of frequency. Other circumstances occur where evaluation is almost a superfluous (or non-existent) task. More information is needed regarding principals’ perceptions about the influence that evaluation has on their own professional performance; it is certainly clear that in those cases where “feedback is late, and evaluation is destructive, no national study is required to suggest that change is necessary” (Reeves, 2004, p.3).

Finally, the research question is important because it addresses whether a relationship exists between the type and intensity of principal evaluation and student achievement. Reeves (2004) emphasizes that “a balanced approach to leadership and test scores has proven increasingly elusive among educational leaders” (p.29). Although there must be no assumption of causality, the correlation or lack of correlation between student performance on a consistently administered standardized assessment and the type and intensity of principal evaluation is important information. Likewise, the principals’ perceptions about the influence of principal evaluation on pupil performance are also important. Through the use of simultaneous (concurrent) mixed method, both quantitative and qualitative data were used to determine the degree to which correlation exists, if any.

Research Design

This study utilized a survey research methodological approach employing simultaneous mixed methods. The data pertaining to the type and intensity with which each principal is evaluated was self-reported through the use of an on-line survey instrument. This information was then correlated to the 2008 ISAT (Illinois Standards Achievement Test) performance level of each individual school. This was accomplished by identifying the individual school ISAT performance levels for each school, then separating the schools into performance tiers. Surveys were then administered to the principals by group, depending upon the performance tier of the school in which they are employed.

As the priority within the research study was focused on the survey data, the predominant research method was quantitative. In addition, the survey contained open-

ended questions in which respondents were asked about their perceptions regarding the influence of principal evaluation on both pupil performance and their own professional performance. Babbie (1990) states that “social researchers who limit themselves to a single method, survey or other, severely limit their ultimate ability to understand the world around them” (p. 27). In order to counteract this concern, the qualitative component of the survey provided depth and richness to the results, and the qualitative data was embedded within the larger study. In terms of integration, the data collected was mixed during the analysis phase. Creswell (2003, p. 218) describes this type of research approach as a “Concurrent Nested Research Strategy.”

After carefully crafting and operationalizing the constructs, the survey was created through the development of questions. The questions were worded very deliberately so as to minimize bias as much as possible. The questions utilized in the quantitative portion of the survey were closed-ended, so it was critical to provide an exhaustive list of responses and make sure that the answer categories were mutually exclusive (Babbie, 1990). As this survey was administered online, special attention was also given to formatting considerations. A seven-point matrix was a central component of the survey, as well as questions about demographic and other related data. The qualitative component of the survey was addressed through open-ended response questions. These questions dealt with the degree to which the respondents perceive a relationship between the type and intensity of principal evaluation and pupil performance. The researcher then used data transformation, a process of “creating codes and themes qualitatively, then counting the number of times they occur in the text data” (Creswell, 2003, p. 221) in order to quantify the results. The survey was administered in strict

adherence to the requirements of Loyola's Institutional Review Board for the Protection of Human Subjects (IRB), and the researcher solicited the input of other researchers throughout the development process as to the viability and appropriateness of the survey questions.

The rationale for the use of the survey approach stems from the need that participants be assured confidentiality. As the survey asked participants about information pertinent to their employment and perceived job performance, the respondents' confidentiality had to be protected. In addition, the research question necessitated a frequency count of principal evaluation methodology. Lastly, as "the explanatory analyses in survey research aim at the development of generalized propositions" (Babbie, 1990, p. 42), it is hoped that some degree of understanding about the larger phenomenon of principal evaluation and student performance was achieved. Because of these considerations, the best way to have collected the desired data in light of this research question was through the use of a properly constructed survey instrument.

Limitations of the Study

As with any research, certain constraints provided limitation on this study. These constraints were largely the result of research design and access factors, and they included the following:

1. Although the survey was made available to all K-8 public school administrators in DuPage County, Lake County, and Will County who lead in schools that administer ISAT, the results of this study are not generalizable to any other population outside of this limited three-county region.
2. Although all eligible subjects were invited to participate in the survey, the

researcher had no control over the individuals who elected to respond. This factor imposed a limitation on the study that had particular importance to the interpretation of the results.

3. Student achievement data has limited validity, as it was collected from one data point (2008 ISAT assessment scores). However, there is no other assessment tool that is administered with the same degree of continuity and uniformity across Illinois public elementary schools.

Bias of the Researcher

As the researcher has been actively employed as an elementary level principal for the last nine years, the possibility of bias must be carefully considered. As stated by Budd & Scoville (2005), “employment research is implicitly shaped by our underlying values” (p. 2). However, the researcher has been fortunate to be regularly evaluated in a manner consistent with currently recommended practice. As such, the researcher has some bias in favor of best practice.

If any member of the dissertation committee sensed the presence of this bias during the process of researching this study, the researcher would have established a personal reflective journal. This journal would have been shared with the dissertation director in order to ensure that personal bias did not affect the data collection, analysis, or integrity of the interpretation.

Definition of Terms

The following definitions are provided in order to assist in the interpretation of the research in this study:

Concurrent nested research strategy – a survey research strategy characterized by

the “use of one data collection phase, during which both quantitative and qualitative data are collected simultaneously . . . a nested approach has a predominant method that guides the project” (Creswell, 2003, p. 218).

Illinois Standards Achievement Test (ISAT) – the state-mandated achievement test in Illinois, it is administered to all public school students in grades 3-8 in subjects that vary by grade level (Writing, Reading, Math, and Science). This test evolved from an earlier state assessment program in the 1990’s, and is currently used to address the No Child Left Behind mandate requiring standardized achievement testing in order for states to receive federal education funds.

ISLLC standards – national-level professional performance standards developed for school leaders by the Interstate School Leaders Licensure Consortium, founded in best practice.

LEA – local education agency.

No Child Left Behind (NCLB) – federal legislation passed in 2001 that was intended to increase educational accountability for school districts, increased standards for pupil achievement, and institute improved equity in student access to quality instructional services.

Principal evaluation – the process by which school principals are provided with appraisal regarding the performance of their job-related responsibilities and professional obligations.

School ISAT performance level – a categorization created specifically for this research study, it represents the percentage of all students ‘meeting expectations’ or ‘exceeding expectations’ in each of the tested ISAT areas (by subject and grade level) in

a given school, averaged at the school level. The final result is the aggregate “school ISAT performance level.”

Summary

Though it is critical to emphasize that there is no assumption of causality, the results of this study are helpful in better understanding the nature of principal evaluation in DuPage, Will, and Lake Counties, and the relationship it has with ISAT standardized assessment performance. In addition, the study provides information about the degree to which principals perceive a relationship between the type and intensity of principal evaluation and pupil performance. By employing simultaneous mixed methods, the study yields valuable statistical information about this subject, as well as qualitative data that is rich, valid, and compelling.

CHAPTER TWO LITERATURE REVIEW

Introduction

This intent of this chapter is to provide a summary of the relevant literature pertaining to the topics of principal evaluation, perceptions of evaluation, and standardized achievement testing. In order to accomplish this goal, it is necessary to provide an overview of several additional related topics. Critical to a discussion of personnel evaluation are the subjects of supervision and leadership, for example. Likewise, it is vital to review the literature on the nature of evaluation if one is to sufficiently discuss its proper use and implementation. The issue of ISAT reform and school accountability is an important topic for review, as well as the implications of principal evaluation on student achievement. Lastly, it is essential that the issue of perceptions about evaluation be reviewed in depth, both from the perspective of the school principal and from that of other constituency groups.

It is unfortunate that McAdams and Barilla (2003) have found that “the relationship between research on administrative appraisal and actual practice is tenuous at best” (p.19). Yet, without familiarity as to the specific findings of the research, it is certainly difficult for one to implement best practices. This review of the literature addresses each of the issues identified above in as much detail as is relevant, and aids in establishing perspective for the research study by framing a context in which it can be

understood. Though the research on the topic of principal evaluation is relatively sparse, the literature is adequate to provide a clear background and research perspective on the topic.

Leadership and Principal Evaluation

Leadership is the very foundation of that which should be appraised through the use of principal evaluation. As such, it is important to review the scholarly literature on several relevant aspects of leadership, as they pertain to the appraisal of principal performance. An overview of the critical competencies that principals should possess can be found in the Interstate School Leaders Licensure Consortium (ISLLC) Standards for School Leaders. Originally established in 1996, “these research-based standards focus on indicators of knowledge, dispositions, and performances important to effective school leadership – learning, teaching, and the success of all students” (Kaplan, Owings, & Nunnery, 2005, p. 31). In order to consider the critical competencies upon which effective principal evaluation should be founded, it is helpful to look at several attributes of educational leadership through the lens of selected ISLLC Standards.

Although the subject of educational leadership is expansive, one of the most important considerations to the topic of principal evaluation is the notion that effective leadership is critical to organizational success. Reeves (2004) clearly asserts that “leadership matters. Even when other variables, including resources and personnel, are held constant, a single leader has an enormous impact on the entire organization” (p. xi). The need for principals to provide leadership for organizational success is emphasized in ISLLC Standard Three, which states that “a school administrator is an educational leader who promotes the success of all students by ensuring management of the organization,

operations, and resources for a safe, efficient, and effective learning environment” (Whaley, 2002, p. 166).

It is important that principals be evaluated in part on their ability to provide the leadership necessary to ensure organizational effectiveness. As highly complex organisms, schools operate in a fluid manner that is affected by many diverse stakeholder groups. Davis (1998) states that “the nature of work within most schools is generally characterized as segmented and unconstrained by rigidly applied systems of control” (p. 60). The challenge for the school administrator in these circumstances is to lead with appropriate context. This may even require that principals lead in a fashion that might seem counterintuitive, such as engaging in school politics. Bolman and Deal (1997) assert that politics can be a vehicle for achieving noble purposes, for example. Regardless, it is particularly critical that the objectives for organizational effectiveness are made clear, and that both internal and external stakeholder groups perceive effectiveness through a similar lens. In some educational settings, “when the external environment perceives that it is satisfied with what happens in the organization, then the organization is, in fact, effective” (Daresh, 2002, p. 87). Organizations with this limited degree of scope can make successful leadership very difficult to accomplish.

Still, a large part of the principal’s leadership role is to use his or her expertise in assessing the performance of the organization, and then making adjustments to the system as needed. When considering the subject of evaluation, this role must be weighed heavily. Hart (1992) states that “while principals and schools depend on each other and affect and shape each other, the relative influence exerted by a principal can provide a measure of success” (p. 52) that the organization is not able to provide. Clearly,

principals do have a vital responsibility in this realm. Likewise, “superintendents and other supervisors can examine principals’ use of organizational analysis techniques that can enhance their success as school leaders and provide opportunities to promote the instructional practices and goals valued by the school district” (Hart, 1992, p. 42). It is critically important to emphasize that “there is no single model or uniform prescription for leadership effectiveness. Effective leadership is a multifaceted process that is often defined through both subjective and objective measures of leader behavior and its effect on organizational processes and outcomes” (Davis, 1998, p. 59).

Yet, principals clearly do not have absolute control over the effectiveness of the larger school organization, and evaluation methods that reflect such an expectation are not realistic. Reeves (2004) states that “in the complex world of educational organizations, most results, both good and bad, are the result of the leader’s actions and a host of other influences within and outside the organization” (p. 20). In assessing the ability of the principal to lead the organization effectively, supervisors must remain even-handed. Reeves stresses that “a cardinal principle of leadership evaluation is that the evaluation system will have the greatest impact on improving individual and organizational performance when the evaluation is focused on those decisions and behaviors that are under the direct control of the leader” (p. 40). In other words, unless the principal is evaluated on those things within his or her span of control, the evaluation is limited in its ability to spur positive change.

Another essential issue when considering the role of leadership in principal evaluation is ethical practice. ISLLC Standard Five states that “a school administrator is an educational leader who promotes the success of all students by acting with integrity,

fairness, and in an ethical manner” (Whaley, 2002, p. 168). Unless the school principal values and is committed to the idea that ethical behavior in leadership is a non-negotiable concept, the organization will suffer for lack of a moral compass.

The premise that ethical leadership is a critical element of the principalship must be a central component in the effective administrative evaluation instrument. The most significant reason for this is the obligation that principals have in providing for the moral development of students. At least two prominent researchers speak to this idea; Strike (2007) maintains that “the ethic of school leaders needs to be an ethic for educational institutions that teach children how to flourish in liberal democratic societies” (p. 11). Similarly, Starratt (2004) states that “moral educational leaders enact the foundational virtues of responsibility, authenticity, and presence – the same virtues that should characterize students’ learning” (p. 3).

Starratt (2004) carries the argument further by asserting that ethical leadership is vital for all stakeholder groups, and defines “the work of leadership as involving the cultivation of virtues that will ground the work of the school as well as guide a diffusion of leadership among all the constituencies of the school” (p. 8). Another compelling argument is that educational leaders do not only have the obligation to behave ethically, but also to seek out additional opportunities to utilize ethical decision-making within the school community. “It is the responsibility of each educational leader and the education community in the aggregate continually to search for what is ethically good in providing services for students and in supporting the activities of school-district employees” (Rebore, 2001, p. 45). The premise that ethical leadership is most essential because of its direct impact on others in the leader’s charge is a truly powerful theme in the literature

related to this topic.

Principal evaluation instruments can be useful tools in helping to inculcate the values that reflect the philosophies of an individual school district, and possibly the larger professional field of educational administration. As “values play an important part in constructing an administrator’s mindscape and in determining leadership practice” (Sergiovanni, 1992, p. 9), appraisal systems should be utilized that reflect key values and beliefs. If behavioral performance expectations are not spelled out explicitly, there is a risk that some administrators may leverage ambiguity so as to make questionable decisions. These administrators can get caught up in a mindset where “the result is an emphasis on doing things right, at the expense of doing the right things” (Sergiovanni, 1992, p. 4).

In the realm of principal evaluation, ethical leadership cannot be separated from technical leadership. “Technical expertise without a moral compass is inadequate for the task, as is a moral compass without technical expertise” (Starratt, 2004, p. 4). Nor can ethical behaviors be easily isolated from one’s general conduct. Rebore (2001) asserts that “in humans, conduct does not merely occur, but emanates from the totality of the person” (p. 5). These ideas underscore the necessity that ethical leadership and values must be considered a central part of effective principal evaluation.

The importance of developing and maintaining quality relationships with others is yet another key attribute of educational leadership with implications for principal evaluation. ISLLC Standard Four states that “a school administrator is an educational leader who promotes the success of all students by collaborating with families and community members, responding to diverse community interests and needs, and

mobilizing community resources” (Whaley, 2002, p. 167). It is clear that this standard deals with the manner in which principals are expected to serve those constituencies in their school communities, and how essential it is for them to establish effective relationships with others. Effective principal evaluation must include a component that addresses the administrator’s ability to develop and maintain relationships with others. “Any model for leading and managing people is only as effective as its foundation, and the foundation of any leadership effort requires the mastery of and the ability to demonstrate a specific set of basic interpersonal skills” (Smith, Montello, & White, 1992, p. 242).

There are several reasons why the task of maintaining high quality relationships is important in educational leadership. Elmore (2002) claims that relationships are integral in the educational setting because they define leadership. “Leaders, then, engage people in shaping the content and conditions of their own learning in organizationally coherent ways” (p. 3). Not surprisingly, relationship building also requires the task of establishing strong communication with others, which is a primary job of the effective principal. “This responsibility seems self-evident – good communication is a critical feature of any endeavor in which people work in close proximity for a common purpose” (Marzano, Waters, & McNulty, 2005, p. 46). The desire to establish and maintain relationships with others prompts the sanguine school leader to communicate more thoughtfully, which in turn contributes to even stronger relationships.

High quality relationships are also important for the purpose of promoting success for the larger organization. Reeves (2006) states that “there is an emerging school of research, known as Positive Organizational Scholarship . . . that explicitly values the

primacy of interpersonal relationships as a key to organizational effectiveness” (p. 21).

Stronger relationships can also lead to increased support for educational initiatives.

Lambert (2003) asserts that “we must enlarge the circle of community to be more inclusive than in the past if we are to develop reciprocal partnerships with parents and members of the broader community” (p. 68). Finally, quality relationships can assist in opening dialogue about the goals and mission of the school program and the needs of the students. This can lead to empowering outcomes indeed, in the way that “parents who participate in conversations about schooling develop a broad perspective that enables them to honor their own values, remain vigilant regarding their own children, and advocate for and help create successful schools for all” (Lambert, 2003, p. 69). For all of the reasons above, the argument is well made that effective principal evaluation must address the degree to which educational leaders build, develop, and maintain quality relationships with others in the school community.

Finally, while evaluating principal effectiveness no area is more important than ISLLC Standard One, which states that “a school administrator is an educational leader who promotes the success of all students by facilitating the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by the school community” (Whaley, 2002, p. 164). Principals must put forth the ideals of the vision of learning at all times, including making sure that all stakeholders support the highest academic expectations for students. The real leadership challenge is to encourage all members of the school community to continue to strive for constant growth and improvement, even in those settings where extraordinary accomplishments are already being achieved. At their most effective, “leaders inspire their followers to sacrifice their

selfish interests for a larger cause” (Hoy & Miskel, 2001, p. 393). This can be a critical skill, particularly when the principal’s vision of learning represents a difficult path to follow.

Educational leadership components must be considered a cornerstone of principal evaluation. If performance appraisal systems do not acknowledge this reality through their design, they are not a viable reflection of the authentic role of the principal.

Supervision and the Principal

Another primary responsibility for which principals must be evaluated is that of supervision. Supervisory tasks for principals include the oversight and evaluation of staff in their instructional duties, the provision of appropriate staff development and training, and the implementation of human resources-related policies and procedures. It should be stressed that “in successful schools the emphasis in supervision is on understanding and improving teaching and learning, not on sorting or grading teachers” (Sergiovanni, 1995, p. 281). Sergiovanni further states that in effective schools “the primary concern of principals is on building a viable, workable, and meaningful supervisory program” (p. 281). This is one of the most critical tasks for which an instructional leader should then be accountable. For this reason, it is important to review the role of the principal as supervisor in more detail.

ISLLC Standard Two states that “a school administrator is an educational leader who promotes the success of all students by advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth” (Whaley, 2002, p. 165). This standard hits at the very heart of school administration, and it is the basis for each of the critical supervisory

responsibilities.

The principal's most prominent supervisory duty is to conduct performance evaluations of staff members in a deliberate, reflective, and beneficial manner. The evaluations must adhere to a method and format that is consistent with district expectations, particularly in cases when the staff member is not performing well. "Thus, an administrator must have sufficient command of due process so that he or she can act fairly but also responsibly, decisively, and expeditiously" (Stufflebeam & Pullin, 1998, p. 219). It is hoped that in most circumstances, staff members are performing their instructional obligations successfully. Yet in order to keep the teacher evaluation processes from becoming a perfunctory task for these individuals, it is critical that the staff members who are being appraised receive feedback that is both valuable and growth-oriented. In these situations, the principal must behave such that "the focus of evaluation is less on measuring and more on describing and illuminating teaching and learning events as well as on identifying the array of meanings that these events have for different people" (Sergiovanni, 1995, p. 283).

The principal must address all of these considerations in a way that is respectful to the context of the supervisor/subordinate relationship. Principals must be sure to establish procedures that are manageable for the task at hand when evaluating others. For example, Danielson and McGreal (2000) emphasize that "thoughtful evaluation requires a significant investment of time, and a single individual can devote that time to only a limited number of people" (p. 57). It is also essential that principals recall the unique nature of the adult learner/teacher relationship when providing feedback in an evaluative manner. When supervisors and subordinates are adult peers, "there is a widely held view

that the relationship . . . should be participative and democratic and characterized by openness, mutual respect, and equality” (Tennant & Pogson, 1995, p. 171). Lastly, the role of the principals is to show consistent judgment from one situation (or one staff member), to the next. For those individuals being evaluated, “this consistency of judgment on the part of trained evaluators is an essential guarantee of the reliability of the system as a whole” (Danielson & McGreal, 2000, p. 23).

Principals also have the supervisory responsibility of providing for the staff development and training that is appropriate for the professional needs of the individuals and the school program. This responsibility is vital in ensuring that staff members have access to the most applicable and up-to-date core knowledge about best practice in instruction. Glatthorn (1997, p. 20) insists that principals implement staff development programs that are reflective, relevant, collaborative, intensive and on-going, and connected with other instructional strategies. Staff members should understand very clearly what principals expect from them as participants in the staff development activities. Sergiovanni and Starratt (1993) emphasize that “. . . those engaged in the supervisory process need to have some sense (by no means ever complete) of the substance of the supervision” (p. 183). Finally, it is critical that principals participate in the staff development programs and activities as well, and that the staff members also perceive the principal as well-versed in best practice.

Under ideal circumstances, principals should work collaboratively with staff members and district office colleagues in order to identify the needs that must be addressed through staff development programming. In this type of climate, “teachers and supervisors share responsibility for the planning, development, and provision of staff

development activities, and the focus is much less on training than on puzzling, inquiring, and solving problems” (Sergiovanni & Starratt, 1993, p. 267). Still, the effective principal should continue in his or her supervisor’s role with oversight, as recommended by Glatthorn (1997). He states that in an ideal school setting “the principal works with a leadership team and a central decision-making body, (but) the principal maintains an active role as instructional leader” (p. 13).

Finally, principals should strive to coordinate staff development so that it is of value to all staff members, regardless of their level of expertise. “Research in acknowledging and developing teacher expertise has provided increased understanding about both the natural and planned movement of teachers from the novice to the expert stage of development” (Danielson & McGreal, 2000, p. 16). Since staff members in every school will fall along the continuum of experience and professional competency, the effective principal will attempt to provide varying levels and types of supervision for each staff member when possible. Danielson and McGreal stress “the importance of building evaluation and staff development programs that allow and encourage this necessary differentiation” (p. 16).

When assessing principal effectiveness in the realm of supervision, a third major responsibility is that of human resource planning and the implementation of human resource policy. While certain elements of this responsibility carry management overtones, the need for the principal to enact human resource-based skills is a key element of school supervision. Human resource planning deals with the manner in which staff are utilized, both for maximum efficiency and for knowledge or skill-based performance. Smith (2001) defines human resource planning as “an organized process of

enabling an organization to achieve its mission by the effective anticipation of human resource needs” (p. 18). Principals who play this role with success have good knowledge of the school district’s objectives and resource availability. They also have a strong awareness of the professional skill sets possessed by the staff members at the school. As instructional leaders with a long-range focus, this role is an important one for principals to play because “human resource planning, as a process, ensures the smooth development of an organization” (Rebore, 2004, p. 38).

Principals must also be knowledgeable about the human resource policies and procedures to which they must adhere. Though some might argue that this responsibility is quite distant from the principal’s role as instructional leader, fair and appropriate application of human resource policies is essential to supervision. Human resource policies and procedures “provide the internal structure necessary to accomplish the school district’s primary mandate, to educate children” (Rebore, 2004, p. 321).

Unfortunately, situations involving human resource issues are often characterized by conflicting interests with a great deal at stake. In this particularly challenging supervisory role, principals must not only be evaluated on the degree to which they interpret and apply human resource policies, but also on the ethics of their behavior. Budd and Scoville (2005) remark that “in the end, the burden is on each individual HR manager to be reflective, always alert to the potential that what appears to be a routine decision may actually be a chance to do right” (p. 198).

The numerous tasks involved in school supervision are varied, yet together they represent the crux of the principal’s job. In evaluating the effectiveness of the school principal, it is necessary to consider the amount of success with which he or she

accomplishes these tasks, as they are foundational to both the staff members and the school alike.

Problems with Principal Evaluation Systems

In those instances when principals are evaluated through the use of performance appraisals, several notable problems often occur. These problems have been documented in the relevant research, and will be noted shortly. Yet the greatest concern is that they stem primarily from the beliefs and experiences of the principals themselves. “Principals have expressed concerns about the failure of evaluation systems to assess the complex nature and comprehensive scope of their positions adequately” (Brown, Irby, & Neumeyer, 1998, p. 1). For this reason, it is vital to identify and define the major problems that are characterized in many contemporary principal evaluation systems.

From the start, it is important to acknowledge the presence of principal-based problems, which can manifest themselves in a number of ways. Principal perceptions about the evaluative process are one common example. Peterson (1991) states that “superintendents often perceive the evaluations as being more thorough than the principals do” (p. 2). Whether or not this is true, it is an indication that many principals have a lesser opinion than superintendents do about the degree of thoroughness.

Principals may also have anxiety about the evaluation process and the implications for continued success. “Principals who have enjoyed high reputations often resist a more systematic program of evaluation, apparently fearing that they have little to gain and much to lose by the process” (Peterson, 1991, p. 3). To the principal who is being evaluated, there is also the risk that an evaluator may be off the mark with his or her appraisal. This is a particular concern when the principal is operating under the belief

that the evaluator is not well versed in the culture of the school. This phenomenon can cause significant problems for both principal and evaluator alike. After all, “principals know their school cultures better than do external evaluators, so opposition to ill-informed evaluations of their performance, which ignore or wrongfully interpret their culture, is bound to occur” (Thomas, Holdaway, & Ward, 2000, p. 220).

Instrument-based problems also undermine the efficacy of the principal evaluation process. Of particular concern is the fact that principal evaluation instruments do not always reflect appropriate evaluative criteria. Hart (1994) asserts that “rather than relying on research about principals, teachers, and students interacting in schools, principal evaluation systems traditionally have drawn their criteria from accreditation organizations” (p. 18). These criteria may or may not be based in best practice, however. In other instances the evaluative criteria are not clear. In their review of related literature, Amsterdam, Johnson, Monrad, and Tonnsen (2005) state that “researchers frequently cited vague evaluative criteria as one of the most numerous concerns about the validity of current principal evaluation systems” (p. 222). Heck and Marcoulides (1992) state that “there is a lack of theory guiding the development of instruments to measure leadership effectiveness, as well as demonstrated psychometric qualities of existing instruments” (p. 129). Further, they share that “in other cases, instruments do not allow the observer to discriminate on the basis of actual observation” (Heck & Marcoulides, 1992, p. 130).

Principal evaluation instruments may also cause problems in the measurement of desired behavior. Thomas et al. (2000) assert that “evaluation models often impose a restrictive view of how effective principals should act” (p. 220). Evaluation instruments that do not allow for liberal or creative practice are highly limiting, and may not provide

an accurate reflection of the principal's work. Administrative evaluations "often assess trivial principal behaviors employing methods and instruments that frequently lack even the rudiments of sound practice" (Snyder & Ebmeier, 1992, p. 76). Often, simple principal behaviors can be easily counted and measured. Yet, evaluative criteria that are based in these types of behaviors lack substance. To the contrary, "decision-making and problem-solving ability are certainly important aspects of the principal's role, yet there are also problems associated with how to observe what are essentially internal processes" (Heck & Marcoulides, 1992, p. 128).

Another substantial problem with principal evaluation systems originates from a lack of clarity in expectations. Some of this problem stems directly from a lack of effective communication between the parties involved. McAdams and Barilla (2003) attest that "communication between superintendents and building principals regarding the purposes and procedures for existing appraisal programs are often weak or non-existent" (p. 19). The resulting confusion about expectations is often not expressed by the principal, who does not wish to appear aloof.

To make matters more challenging, this problem has existed for some time. Peterson (1991) references data collected in 1986 where "only 58 percent of the principals they surveyed said that the expectations for their performance had been made clear prior to each year's evaluation" (p.3). Some of the responsibility for the lack of communicated expectations should be placed squarely on the shoulders of supervisors. However, some responsibility must also be attributed to unclear evaluation instruments that use vague language to avoid addressing expectations directly. "Implicit in the documents is the assumption that principals will be evaluated on the effectiveness they

demonstrate in these (respective) activities. The expected standards of performance, or criteria for making an evaluation decision, were usually not identified” (Thomas, Holdaway, & Ward, 2000, 233).

Ambiguous objectives for the principal evaluation are another problem. This problem manifests itself when it is not clear what the purposes are for conducting evaluation. Heck and Marcoulides (1992) assign the blame to the “lack of theoretically-driven empirical research to establish and validate the appropriate domains of the principal’s role and their collective effects on the school’s achievement at a higher or lower academic level” (p. 125). The result of this research vacuum has been a wide range of perceived objectives for the performance appraisal process, and it has contributed to inconsistent messages. In those cases where research does provide specific guidance as to the selection of evaluative objectives, it is not often truly substantive. Hart (1994) states that the “lists, characteristics, behavioral objects, and competencies in the principal evaluation literature tend to be descriptive and perceptually based” (p. 4). Though the ISLLC Standards are exceptional in their substance, a greater volume of empirical research would go far to support the presence of evaluative objectives that are concrete and based in knowledge of best practice.

It is not surprising that many principals are confused about the ambiguity in evaluation objectives. For that matter, principal evaluation instruments often have very little in common from one district to another, in terms of the performance objectives that are measured. This phenomenon was actually documented in the early 1990’s. “Whim, person-to-person conflict, a bit of political maneuvering, and some “who’s-in-control-here” strategizing are far more likely to guide evaluations of the performance of building

principals than are precise ‘indicators’ of administrative accomplishment” (Smylie & Crowson, 1993, p. 64). Glasman and Martens (1993) state that “essentially, anything goes in evaluating principals – his/her performance, personality, activities, and so on” (p. 49). Further, Peterson observed in 1991 that “principal evaluation often receives short shrift, due, in part, to confusion and misperception about the purpose of evaluation and the formation and application of evaluation criteria” (p. 2).

Yet another problem with principal evaluation systems stems from the difficulty in adequately evaluating many of the critical competencies of the position. The principal’s job is a challenging one to align job responsibilities with measures of performance. As a result, “inconsistencies develop because of the differences between the nature of principals’ work and the nature of the evaluation decoupled from context” (Hart, 1994, p. 4). The challenge of accurately appraising principal performance was also identified in the early 1990’s. Ginsberg and Thompson (1992) mentioned “the nature of principal work, problems with definitional specificity of tasks, the situational nature of the job, (and) the varying expectations for principals . . .” (p. 59) as contributing issues. At the same time, other researchers warned about evaluating principals on personality-based criteria. Heck and Marcoulides (1992) stated that “while attitudes or traits may be correlated with school outcomes, they are more difficult to defend as important indicators of principal effectiveness” (p. 128).

For these reasons, Heck and Marcoulides (1992) assert that “any evaluation model that tries to capture all of the subtleties of the role, and operationalize all of the day-to-day activities of the principal, is doomed to failure” (p. 139). Instead, this problem can only be surmounted with the thoughtful and deliberate application of evaluation

instruments that consider the contextual nature of the setting and circumstances in which principals practices. “Evaluation of principals must somehow account for differing conditions and differing styles of leadership, implying the need for approaches to evaluation unlike traditional systems utilized in education” (Ginsberg & Thompson, 1992, p. 63).

A final problem with many of the current principal evaluation systems is a lack of focus on professional growth or professional development. This is particularly true for those principals who have moved out of the novice role, and matured into experienced or veteran administrators. In addition, the absence of these professional growth components has been de rigeur for some time. In 1993, Glasman and Martens stated that “it appears that there are few supervisors who, on a consistent basis, diagnose evaluation data and suggest to principals ways of self-development” (p. 49). More recently, Brown, Irby, and Neumeyer listed a “lack of connection to professional growth” (1998, p. 1) as one of their specific concerns about the contemporary condition of principal evaluation and performance appraisal.

This problem is disappointing because it strikes at the heart of the question “for what purpose is evaluation conducted?” In order for principals to act in the capacity of instructional leader, it is vital for them to model growth, development, and change for others in the school community. After all, “leaders who tend to be most successful in their organizations also tend to experience personal change and growth as they influence outcomes. Development is a two-way, interactive process” (Hart, 1994, p. 6).

Each of these problems about the state of principal evaluation has at least one respective solution. However, the problems themselves reflect the current practice in a

preponderance of school districts. In the next section, literature will be reviewed that supports the manner in which principal evaluation systems can be improved.

Improvement of Principal Evaluation Systems

As “the job requirements and expectations that accompany the principalship have become steadily more demanding” (Thomas, Grigsby, Miller, & Scully, 2003, p. 40), the importance of addressing the persistent problems with principal evaluations becomes more pressing. There are a number of resolutions that should be considered for the purpose of improving the viability of principal evaluation systems. Some individuals tend to believe that “if the superintendent wants to fire you, he can find a reason” (Stine, 2001, p. 9). Yet, by improving principal evaluation systems to reflect best practice, this type of mentality can be significantly diminished.

It is unfortunate that in the past, “little attention has been given to the internal evaluation instruments used to scrutinize the performance of building-level principals” (Catano & Stronge, 2006, p. 222). This section will discuss the importance of the evaluation instruments themselves, as well as the context, manner, and spirit in which they are used to conduct principal performance appraisals. The literature is limited in that “little discussion has been offered about methods for improving the process in developing a principal evaluation system that leads to fair and valid decisions” (Amsterdam, Johnson, Monrad, & Tonnsen, 2003, p. 222). However, the literature does address a number of key recommendations that can help improve the viability of the principal evaluation process in many school districts.

The first of these recommendations for improvement would be to insure that the various stakeholder groups are involved in the process of developing the evaluation

systems. Amsterdam et al (2003) observe that “greater involvement of stakeholders in the program evaluation process has been advocated in the evaluation literature” (p. 223). This is an important consideration because it helps the individuals who are involved feel as though they have a contributing role in the evaluation process. “Principal evaluation works best when it is not simply imposed from above” (Peterson, 1991, p. 3). Anderson (1991) also supports the value of stakeholder participation, stating that in order “to develop an evaluation process that will be embraced by principals, districts must involve them in the development of the program” (p. 108). Of course, the challenge inherent in involving too many stakeholders in the evaluation process stems from the diverse range of opinions about that which is most important to evaluate. “When any group of stakeholders begins to brainstorm leadership domains, it is likely to think of a very long list of characteristics of leadership performance” (Reeves, 2004, p. 39). As such, it is vital that reasonable limits are placed on the degree of stakeholder involvement that is appropriate in any given setting or situation.

Separately, some researchers advocate the concept of involving external stakeholders as contributors to the evaluation itself. Murphy and Pimental (1996), for example, state that “once they take on the role of evaluator, parents and students move from the sidelines into the heart of the learning enterprise to share responsibility for improving education” (p. 77). Further, they argue that soliciting input from teachers about principal performance “provides principals with crucial feedback, and they are paying attention as never before” (Murphy & Pimental, 1996, p. 77).

A second recommendation for improvement in principal evaluation is rooted in the need for ‘internal consistency.’ The concept of internal consistency addresses the

degree to which the evaluation procedures and instruments are used appropriately, fairly and with proper regard for people and process. Stufflebeam and Nevo (1993) encapsulate this well by stating that “standards require that (principal) evaluations be conducted legally, ethically, and with due regard for the welfare of students, other clients, teachers, principals, and other educators” (p. 38). Within the performance appraisal process, inconsistent practice can yield damaging outcomes. “Having one set of policies on paper another set in practice adds confusion to the system. The drift from “what we are supposed to do” and “what we actually do” contributes to a hidden practice and compliance mentality instead of open practice for continuous improvement” (Kearney, 2005, p. 19).

In the drive to utilize principal evaluation systems successfully, Stine (2001) reminds that “the goal is to develop a system which is fair and provides for accountability” (p. 8). In addition, “the evaluation form is critical. But it is only one dimension. The process is even more significant” (Stine, 2001, p. 8). This sentiment is echoed by McAdams and Barilla (2003): “the effectiveness of an appraisal system is not determined by the type of instrument, but rather by how effectively that instrument is used” (p. 20). Internal consistency also demands that principals receive regular feedback about their performance. Reeves (1998) asserts that “if feedback is to be effective, it must be frequent . . . The most effective feedback systems provide measurable results at least quarterly” (p. 9). Internal consistency is important to assure that the principal evaluation process is being conducted, at least in part, with the objective of helping provide opportunities for genuine growth, development, and improvement in performance.

The third recommendation for improvement is to ensure that the principal evaluation system is grounded in clearly defined policies and procedures. Brown and Irby (1998) call policies and procedures “essential,” and emphasize that “procedures must include a system for communication, a strategy for ongoing monitoring and provisions for a continuous feedback loop between evaluators and principals” (p. 2). Some researchers stress the need for procedures in other aspects of principal evaluation, as well. McAdams and Barilla (2003) advocate that effective evaluation plans must be “understandable, state merit guidelines and personnel policies, and objectively measure job performance to support and assist the school board in determining compensation and personnel-related goals” (p. 20). Additionally, Stine (2001) states that evaluation meeting “dates should be established in advance and the format needs to be clear to both parties” (p. 4).

In striving to use clearly defined policies and procedures, it is most critical that they are established both in law as well as the belief systems and mission of the school district. This must be accomplished far in advance of committing to one evaluation approach or another. “It is important that district leadership identify a philosophy or theory of action from which they are working” (Kearney, 2005, p. 19). Without this, there is the risk that policies and procedures will not align with the overarching goals of the school division. Still, “regardless of the model or the product that is adopted, principals should be left with clarity about the criteria to be used to assess them” (Reeves, 1998, p. 10).

The next recommendation to improve principal evaluation is to ensure alignment with professional standards. This is essential because it links practice to research and the

larger body of professional knowledge. Catano and Stronge (2006) assert that “this alignment would facilitate clearer communication to principals regarding expectations for job responsibilities and their performance of those responsibilities” (p. 232).

Additionally, they state that “this alignment should promote better job satisfaction and effectiveness for school principals by reducing role conflict and consequent role strain” (Catano & Stronge, 2006, p. 232). Stufflebeam and Nevo (1993) make another more elaborate case for the alignment of standards and evaluation:

“These standards provide direction for assuring that principal evaluations will be ethical and legal in their examination of an individual’s aptitudes, competencies, performance, and special achievements; useful for decision making and obtaining direction for improvement; feasible for use within reasonable practical constraints; and accurate in the information they provide” (p. 44).

Reeves (2004) makes perhaps the most distinct argument by asserting that “without standards, evaluations are constantly subject to the shifting sands of relative performance rather than the bedrock of clear, fair, and immutable standards” (p. 21). Equally important is the development of a shared consensus about the organizational interpretation of performance standards. After all, “an essential component of effective leadership is the development of a mutual understanding between the leader and the organization about the specific behaviors and professional practices that are associated with exemplary leadership” (Reeves, 2003, p. 21).

Another recommendation toward improving principal evaluation is to include an appropriate goal-setting component. First, “the principal and evaluator, on an individual

basis, should develop performance objectives and goals (Marcoux, Brown, Irby, & Lara-Alecio, 2003, p. 6). This step is needed to help the principal identify goals that are well suited to his or her needs, as well as the nature of the particular school setting.

“Principals should be intimately involved in the goal-setting process, and they should certainly be fully informed of how the various goals will be weighted and assessed. This knowledge encourages principals to focus on the aspects of their job deemed most important” (Peterson, 1991, p. 3).

Second, “districts are advised to encourage, or even require, principals to set individual and school goals that will enhance their performance and contribute to system wide goals and objectives” (Anderson, 1991, p. 109). In order for this to be effective however, Marcoux et al state that “the evaluation process should enable the principal to set and focus on goals that are aligned with the vision of the school district and campus” (p. 14). Murphy and Pimental (1996) remind that “setting goals is not a precise science. The secret is to aim high but remain realistic. With district goals as a backdrop, each school (should) get a set of improvement goals tailor-made to its circumstance” (p. 78). School districts must then complete the cycle by requiring “each school to develop annual campus improvement plans and . . . requiring each principal to establish annual personal growth plans” (Peterson, 1991, p. 3). It is only through establishing this strong goal-setting component that the principal evaluation process can adequately reflect the objectives and needs of the larger organization.

The principal evaluation process can also be improved by making sure that evaluations are conducted within the proper context of the principal’s role and circumstances. As all school settings are unique, the skills and behaviors that are

necessary to lead in each school are also unique. However, in a desire to implement consistent measures, “school districts might quickly seize on standardized instruments and use them as the measure of principal performance even though they are not sensitive to the particulars of a given principal’s job in a given year” (Stufflebeam & Nevo, 1993, p. 33). Applying the same evaluative lens to various situations without acknowledging the differences in context is both unethical and inappropriate. In 1993, Stufflebeam and Nevo also stated that “many personnel evaluation systems need to be improved in how well they define jobs, how effectively they consider environmental influences, how validly they measure job qualifications and performance, and how effectively they control for various kinds of bias” (p. 40).

A better approach is to accept that a considerable amount of variance exists from one school environment to another, and different leadership will be needed to address these variances. When it comes to evaluation, “principals should be held to a standard of behavior consonant with those in similar circumstances. Standards tie knowledge and action to context. This absolute reality requires principal evaluation frameworks that acknowledge the importance of actions taken in context under unique circumstances” (Weaver, 1994, p. 5).

Unfortunately, evaluators themselves are not always familiar with the dynamics of the specific school setting for which the principal is responsible. “Considering this, organizations should focus on developing knowledgeable supervisors who will direct effective performance appraisal systems” (McAdams & Barilla, 2003, p. 21). Another way to ensure that principals are being evaluated in proper context is “to assess how the principal may contribute, both directly and indirectly, to the processes through which the

school is governed, how it is organized instructionally, and how the climate of the school is developed” (Heck & Marcoulides, 1992, p. 128).

Another recommendation toward improving principal evaluation is to ensure that evaluation systems rely on a variety of data. Though simple in concept, this recommendation is one of the most vital in aligning accurate evaluation and genuine performance. Observation data is the most traditional kind of data that is in use, however there are many other types of data that can also be utilized for the purpose of evaluating principals. “Other examples could include a portfolio of schedules, staff development activities, meeting agendas, school recognition or awards, external funding, budget reports, student discipline data, drop-out rates, innovative programs, parent participation, and articulation activities” (Stine, 2001, p. 6). Additionally, Peterson (1991) advocates a “broad range of sources that can be collected to evaluate principals: attendance and test records, committee reports, newsletters, clippings, and time logs” (p. 4). He also recommends that supervisors should shadow principals “to take extensive notes on their actions and conversations” (p. 4).

Other valuable sources of data are individual stakeholders from the school community. Anderson (1991) states that “districts should also consider soliciting confidential feedback from peers and teachers” (p. 109). He asserts that teachers are in particularly good position to know “whether a principal is performing satisfactorily” (p. 110), although it is important that teachers are guaranteed some degree of anonymity. Clearly, there are a number of potential data sources that can be used in evaluating principals. However, Stine (2001) reminds that although the potential “list could be extensive, (it) should be limited to those items most relevant to the particular school and

it's specific goals" (p. 6).

The principal evaluation process can also be improved by increasing emphasis on the importance of self-reflection. For improved performance, principals should be encouraged to reflect often upon their professional practice and decision-making. Reeves (1998) attests that "educator growth and improved leadership practice have been linked to reflection and self-assessment as integral components of performance evaluation." Further, "reflection and self-assessment offers valuable insights into the principal's effectiveness and needed areas of growth" (p. 10). Anderson (1991) concurs: "districts should also encourage principals to spend time reflecting on their performance. Self-appraisal of performance and progress towards written goals is an activity that principals (should) engage in throughout the year" (p. 109). The purpose of reflection is to help the principal move past the point where professional practice is locked into preferred ways of acting and thinking. "Evaluation processes should help the principal to reflect in order to change behaviors that lead to better practice and decision-making skills" (Marcoux, Brown, Irby, & Lara-Alecio, 2003, p. 14). "In short, the best evaluation systems measure not only competence, but also train the principal to be more competent and effective" (Marcoux et al, 2003, p. 3).

A final recommendation for the improvement of principal evaluation systems rests in the need that they be based squarely on valid job descriptions. Stine (2001) states that "the whole (evaluative) process must begin with precise job descriptions which would provide structure to the summative evaluation instrument; however, the job description alone is insufficient" (p. 4). Unless principals are made clearly aware of the expectations for their position, it is ethically unsound to evaluate them. However, in

some cases job descriptions are present and available but they are not accurate or well suited to the specifics of the position. Stufflebeam and Nevo (1993) assert that “evaluations of the performance of principals can be improved by helping school districts improve their development and use of principal job descriptions” (p. 34). Unless there is strong alignment between job description and the evaluation process, the evaluative results will always contain a certain amount of undesired ambiguity.

Though it is apparent that there are many ways in which principal evaluation can be improved, this is not to imply that it is impossible to evaluate principals in a fair, ethical, growth-oriented, and accurate manner. The key is that school districts utilize the recommendations for improvement in order to review, modify, and implement their own evaluation systems that reflect the unique needs and circumstances of their leaders and their schools. The next section will summarize the literature pertaining to the development of more effective principal evaluation systems.

Development of Effective Principal Evaluation Systems

The literature detailing the development of principal evaluation systems generally centers on the topic of the respective decisions that must be considered when creating a system or putting one in place for the first time. Ultimately, “the goal in evaluation is to develop a program that is both valid and reliable and that helps the principal to reflect on performance (Marcoux, Brown, Irby, & Lara-Alecio, 2003, p. 4). This is not easily accomplished however, especially when a school district is starting the development process from scratch. The most important thing to remember is that “the focus of personnel evaluation should be on measurable performance and not on personal characteristics or traits that are difficult to measure” (Amsterdam, Johnson, Monrad, &

Tonnsen, 2003, P. 222). Keeping this in mind, the research identifies a number of vital considerations that must be addressed when developing a relevant and effective principal evaluation system.

Amsterdam, Johnson, Monrad and Tonnsen (2003) state that “central to the evaluative process are clearly defined criteria for acceptable principal performance” (p.222). Other essential components include a needs assessment that can be used to formulate the principal’s goals, a work plan, and a reflection or self-evaluation portion. Of course, it is the responsibility of the supervisor “to review and react to the principal’s work plan and monitor progress in carrying it out” (Connecticut State Department of Education, 1990, p. 18).

Thomas, Holdaway, and Ward (2000) state that there are “a number of activities that should be incorporated into the supervisory process: a pre-observation conference; careful recording of principals’ words and behaviors; a post-observation conference; feedback; and assistance in developing plans for professional growth” (221). Similarly, Whaley (2002) addresses six essential components of effective principal evaluation, including self-assessment, a goal-setting conference with the supervisor, implementation of strategies to pursue goals, a mid-year progress conference, a summative evaluation conference, and the supervisor’s completion of a summative evaluation document (p. 179). Though there are some differences in the relative value that is placed on each of these components, there is general agreement from researchers that they represent the critical parts that must be included when developing an effective evaluation system.

The next consideration is the topic of which participants should be involved in the evaluation development process. Schools have a variety of stakeholders who hold an

interest in the success or failure of the principal. For this reason, it is vital that there is some collaborative input into the development of a new system. “Furthermore, support for the system may be built among groups represented during its development” (Amsterdam, Johnson, Monrad & Tonnsen, 2003, p. 223), which can go a long way toward legitimizing the work that has been accomplished. Brown and Irby (1998) concur, stating that “those who have successfully collaborated point out that recognizing the need for change in evaluation and the need for broad-based input improve the chances of successful implementation (p. 12). Lastly, Amsterdam et al (2003) assert that “technical accuracy of evaluative decisions is improved through the involvement of these (stakeholder) groups in determining and refining its purpose, evaluation criteria, instrumentation, and procedures for collecting information” (p. 223).

Amsterdam, Johnson, Monrad and Tonnsen (2003) also recommend that “principals should participate in the development of their evaluation systems” (p. 223). The assumption is that they are the most knowledgeable of all about the dynamics and nuances of their job, as it exists in their respective schools and school districts. Barth (2001) also advocates principal involvement, but for the reason of facilitating evaluation for professional growth. He argues that involvement in the evaluation development process helps the principal take “ownership of the learning” (p. 148). This can help principals expand in the area of professional development, as a part of the larger evaluation process. “When principals pose and address the important issues about which they want and need to know more, they come alive as learners” (Barth, 2001, p. 148). Involvement in the development of the evaluation process can help principals inject those elements into the system that encourage learning, risk-taking, and growing.

Another important consideration when developing a successful principal evaluation system is the nature of the process that is utilized. What is most important is not to put the cart before the horse, because “an effective appraisal system is (only) developed after district philosophy and goals are established” (McAdams & Barilla, 2003, p. 20). Thomas, Holdaway, and Ward (2000) relate that “the development of a successful evaluation program, which satisfies the need for accountability and performance improvement, requires that school systems carefully plan the process” (p. 221).

A well-planned implementation process must be conducted by a “committee of about a dozen people, one half of them principals” (Peterson, 1991, p. 3). It is important that the committee next “assesses other principal evaluation programs with the aid of a consultant, drafts a plan and submits it to the principals for amendment, and then sends the revised plan to the school board” (Peterson, p.3). Thomas, Holdaway, and Ward (2000) detail a more elaborate process, including nine distinct components. They include the following steps: “1) identify the purposes for evaluation, 2) develop clear performance expectations, 3) involve principals in planning, 4) encourage goal-setting and self-reflection, 5) often observe principals in action, 6) involve peers and teachers in providing feedback, 7) collect artifacts, 8) adopt a cyclical approach to evaluation, and 9) reward outstanding performance” (p. 221). If each of these steps in the development process are followed, the evaluation stands a very good chance of proving to be both effective and relevant. In addition, these steps ensure that the evaluation system has the dual ability to serve both the need for performance appraisal as well as professional growth. These cross-purposes help to make the evaluation process valuable to the

principals, because “striking the right balance of accountability and support” (Tucker & Stronge, 2005, p. 53) is of the utmost importance for them.

Another important consideration is the need to clearly identify those standards that will serve as the basis of the evaluation. What are the modern performance expectations that principals should be encouraged to meet? DuFour and Eaker (1998), for example, state that “although past images of the principalship have focused on principals who were strong, assertive, and forceful leaders, the more promising contemporary view calls for principals who can work collaboratively with others in building consensus” (p. 203). This kind of ‘shift’ in the way that the role of the principal has changed must be considered when developing new appraisal processes, or they will not be reflective of best practice.

The standards themselves should be the bedrock of any evaluation development process. Reeves (2004) reminds that “the essence of a standards-based approach is that the performance is compared to an objective standard, not to the performance of others” (p. 43). The result of this approach is that the standards must be very carefully and deliberately selected. Brown and Irby (1998) state that “frequently, principals comment that their evaluations are not always aligned with their job descriptions and that they are not always certain as to priority expectations. Performance expectations must be determined before you implement a new appraisal system” (p. 3). As such, it would be recommended for an evaluation development committee to first identify the key standards before contemplating any type of performance instrument. Prioritizing the standards to reflect the school district’s mission is yet another important task. Heck and Marcoulides (1992) assert that “in developing such systems, we have argued that choices

must be made about what aspects of the role to emphasize and how to measure those aspects using sound psychometric procedures in relation to identified purposes of evaluation” (p. 130). In the end, the identification of standards provides that “the development of leadership expectations should result in a shared vision of leadership within a school district” (Brown & Irby, 1998, p. 3).

Establishing fair instrumentation and procedures is still another critical consideration when developing an effective principal evaluation system. This is vital because it shows principals that the evaluation process has value. “Often administrators feel that the evaluation is not viewed as important and that, as a result, little time and effort are spent on the process” (Brown & Irby, 1998, p. 11). Clear instruments and procedures ensure that the evaluation process must be followed with at least a certain degree of fidelity, which in turn lends credibility to both the task and the outcome.

Fair instrumentation and procedures also foster equity among those individuals who are being evaluated. Kearney (2005) states that “a well-thought-out set of policies can guide and support fair and equitable practices that both support and assess high quality administrator performance” (p. 19). McAdams and Barilla emphasize that “appraisal systems must also be legal, fair, and reliable. All administrators have the right to know and understand the evaluation procedure and process” (p. 20). Fair instruments are the backbone of an effective evaluation system – one which “allows both the evaluator and the one being evaluated to understand clearly the difference between various levels of performance” (Reeves, 2004, p. 7).

A final consideration in the development of an effective principal evaluation system is the need to pilot the program. Without this step, it is very difficult to ascertain

whether or not the system will meet its desired ends. Glatthorn (1997) stresses that when reviewing evaluative programs, there is a substantial need to “provide for formative and summative evaluation of the program” (p. 100). Consequently, it is also important to pilot the system because it is financially sound to do so before engaging in full-scale implementation. It is also a better alternative than continuing to research the topic, but choosing not to take action to initiate a new program. “In a time of scarce resources, piloting and implementing new systems of administrator support and evaluation using a significant base is a better use of resources than engaging in duplicative research on the work of principals” (Kearney, 2005, p. 20). Reeves (2004) also recommends piloting a new evaluation program, encouraging the use of a “field test”. He states that “a field test should include the voluntary application of the new evaluation tool with leaders in the central office and in schools. Ideally, the field test should include a broad spectrum of leaders, including those who are new to their positions as well as veterans with decades of service” (p. 99). He adds that “field tests should also provide insight about the value of the process” (Reeves, 2004, p. 99). In each case, the critical purpose of the pilot remains the same: to be sure that the new evaluation system is effective and meeting the desired objectives before pressing it into wider service.

When developing new principal evaluation systems, it is also important to consider the type of system that will be implemented. Though most districts elect to use a traditional principal evaluation format (as described throughout this section), there is one alternative that has garnered a good deal of attention. This kind of system, commonly referred to as an “administrative portfolio,” merits particular acknowledgement because of its departure from convention. Advocated most avidly by

Genevieve Brown and Beverly Irby, the administrative portfolio is intended to collect data from a different perspective. Generally described, “the administrator portfolio is a purposeful, self-collected collection of artifacts and reflective entries which represents an administrator’s growth” (Brown & Irby, 1995, p. 3). Constructed by the administrator who is under evaluation, the portfolio allows for the evaluator to view a wide array of self-selected data about the principal’s performance and accomplishments. In addition, “evaluation portfolio provides for principals and supervisors to collaboratively develop and agree on standards, expectations, goals, and/or proficiencies to be evaluated, and it addresses those standards, expectations, goals, and/or proficiencies directly” (Brown & Irby, 2001, p. 37). Though there are challenges inherent in the portfolio system, such as the potential for a principal to aggrandize his or her work, it is both unique and promising enough to merit special acknowledgement in this section.

The development of a new or improved principal evaluation system is an important and valuable undertaking. Still, it must be pursued through a deliberate process that reflects the objectives that the system is intended to serve. At present, “strong principals are crucial . . . but the image of how a strong principal operates needs to be reconsidered” (DuFour & Eaker, 1998, p. 183). Principal evaluation systems are one vehicle that can be used to restructure the nature of the principals’ job. Marcoux, Brown, Irby, and Lara-Alecio (2003) assert that “evaluation processes should help the principal to reflect in order to change behaviors that lead to better practice and decision-making skills” (p. 13). For this reason, the evaluation processes need to be carefully constructed and adhered to closely. The evaluation process also provides a rare opportunity to help principals improve their practice and affect change within schools,

because ultimately “ineffectiveness is rarely the result of a personal defect, but rather the failure of the leader to acquire the necessary knowledge and skills to become effective” (Reeves, 2004, p. 102).

ISAT, NCLB, and School Accountability

In concept, pupil achievement and school performance should go hand in hand. “Across the United States, school accountability is a theme now commonly heard in the regular discourse among state government officials and local community members” (Tucker & Stronge, 2005, p. 12). Through the passage of federal legislation, our contemporary society has expressed a desire that schools are held accountable for measurable pupil achievement. Resultantly, school performance indicators have been instituted as the stick by which to measure this progress. As school leaders, many principals (and other critics) have contested the nature of these school performance indicators, and most explicitly those based upon standardized testing. Reeves (2008) states that “one of the most consistent criticisms of No Child Left Behind (NCLB) has been the unfairness of labeling schools on the basis of annual test-score comparisons” (p. 89). Some other researchers decry the value of standardized tests in determining school performance, claiming that performance is not based on pupil achievement but on other contributing factors. “Most high-performing schools in our highly segregated society have gotten there not by knowing a great deal about instructional practice or improvement but by getting and holding on to students in high socioeconomic groups” (Elmore, 2003, p. 4). Still, state standardized testing is one valuable way to measure pupil achievement, and it is a major tool in use to determine school accountability.

Educational accountability plays a major role in schools and school divisions

today. There are many perspectives on the purpose of school accountability. However, Reeves (2004) reminds us that “the fundamental purpose of effective accountability systems is not the rating, ranking, evaluating, sorting, and humiliating of students, schools, teachers, and leaders. The fundamental purpose of effective accountability systems is the improvement of teaching and learning” (p. 16). The role of the principal in an age of accountability is to facilitate those teaching and learning behaviors that lead to the greatest results, in terms of pupil achievement. Principal evaluation is one tool that can be used to monitor and encourage this behavior. “Through the systematic identification of effective practice and the careful measurement of both student results and leadership actions, systemwide accountability and individual evaluation can transform random acts of good and bad practice into the brains of a learning organization” (Reeves, 2004, p. 17).

Educational accountability provides its share of challenges, both for schools and principals. One of the problems is that it leads to the practice of applying simple and uniform measures to gauge degrees of success. Reeves (2008) reminds that “two schools with identical test scores can have vastly different learning environments and produce vastly different student outcomes in other important areas, such as initiative, teamwork, intellectual curiosity, and physical and emotional health” (p. 89). Elmore (2003) also describes the disproportional impact that principals are expected to play on outcomes. He states that “the U.S. fetish for leadership leads to an overemphasis on the personal attributes of school leaders and a correspondingly weak focus on the technical, cognitive demands of instructional practice and the affective and behavioral responses to those demands” (p. 3). Another problem is our stark inability to discern which of the

innumerable factors inherent in the education process is accountable for student achievement, when it occurs. “The presumption . . . is that the relationship between leader, teacher, and student is an imponderable black box and thus only the output can be examined” (Reeves, 2004, p. 29). This is, of course, off the mark.

The role of the principal in this era of school accountability continues to change and evolve. Elmore (2003) asserts that “successful leaders have an explicit theory of what good instructional practice looks like. They model their own learning and theories of learning in their work, work publicly on the improvement of their own practice, and engage others in powerful discourse about good instruction” (p. 3). While shortsighted principals may encourage disproportionate instruction for those disciplines that are assessed on the standardized exams, most administrators see the folly in this behavior. “Wise leaders know that the lessons learned from good practice in physical education, music, technology, kindergarten, and a host of other nontested subjects and grades can improve instruction in every subject and at every grade level” (Reeves, 2008, p. 89). In the end, leadership matters in promoting pupil and school-wide success. Although “one would not wish to put all of the school improvement eggs in one leadership basket, any discussion of strategies to promote school level reform that establishes accountability and improves educational outcomes must include the role of the principal as a key element” (Heck & Marcoulides, 1992, p. 133).

While educational accountability can be challenging for schools and school leaders to bear, it does yield positive benefit. Elmore (2003) asserts that one valuable lesson is for educators to learn to institute the accountability systems that are necessary through their own hand: “internal accountability precedes external accountability” (p. 3).

Further, “internal coherence around instructional practice is a prerequisite for strong performance, whatever the requirements of the external accountability system” (Elmore, 2003, p. 3). If accountability is here to stay, educational leaders would be sanguine to establish systems of their own that will help to accomplish the kinds of accountability that are needed.

The legislation that was implemented by the federal government to heighten educational accountability is No Child Left Behind (NCLB). While it has played a large part in changing the way schools and school leaders transact business, it has not been without its critics. Though NCLB was passed in 2001, “most of the normal institutional processes that preceded the reauthorization of a major piece of federal policy got short-circuited prior to its enactment, so most of the expert advice on issues of testing, assessment, school improvement, and accountability that would usually have been brought to bear got ignored” (Elmore, 2003, p. 1). Like any legislation, NCLB has also resulted in unintended outcomes. “NCLB’s accountability mandates have a disparate impact on large urban districts with sizable low income and minority populations. These districts are often tempted to adopt quick fixes in an attempt to avoid sanctions” (Hardy, 2006, p. 17). Another major flaw of the legislation is that “it focused primarily on measuring growth in school performance against fixed standards . . . and only incidentally on building the capacity of individual educators and schools to deliver high-quality instruction to students” (Elmore, 2003, p. 1).

On the positive side, “NCLB has certainly focused attention on the performance of poor, minority, and low-performing students” (Hardy, 2006, p. 17). NCLB has also encouraged school districts to consider “the institutional accountability and policy level

of assessment use, where resource allocation, programmatic, policy and other decisions are made by school, district, and community leaders . . . (who) are the people to be held accountable for the quality of schools” (Stiggins, 2008, p. 4). As “NCLB judges a school’s performance by the distance between its current performance level and the performance standard for which the school is being held accountable” (Elmore, 2003, p. 2), parents and community members tend to appreciate the regular (though somewhat inequitable) increments between expected performance levels.

Within the State of Illinois, the Illinois Standards Achievement Test (ISAT) is the standardized assessment used for the dual purposes of determining school performance and measuring pupil achievement. Though it is an imperfect measure of pupil achievement, it does provide one lens through which to consider academic progress. It is also administered to all public school students in grades 3 to 8 throughout the state, so it has value in its ability to provide comparison. Some researchers levy considerable criticism at the ISAT. Wick (2007) states for example, that “citizen groups, school boards, the press, parent groups, and teachers have been lulled into believing that the ISAT represents a perfectly satisfactory barometer of excellence-seeking in this state.” (p. 3). However, school administrators are noticeably absent in his statement. This is because many administrators are quite wary of the ISAT assessment, and particularly of its use as a gauge to measure pupil achievement and school performance. After all, as Reeves (2008) states, “effective accountability must include more than a litany of student test scores” (p. 89). Unfortunately, many administrators are also untrained in working with standardized testing, and are thus unable to help provide others with adequate guidance in this area. Stiggins (2007) states that “lest we believe that (teachers) can turn

to their principals for help, we must also face the fact that assessment training remains nearly nonexistent in university-based leadership preparation programs nationwide” (p. 7). Regardless, the universality with which ISAT is administered does allow its use for the sake of comparison between schools, and as a determining measure of pupil achievement on the objectives that are assessed.

Educational accountability has indisputably changed the landscape of public education. As critics, Leithwood and Jantzi (2006) state that “if policy-makers insist on detailed prescription of local practices, then local leaders should be held accountable for implementing those practices with fidelity and policy-makers should be held accountable for the effects of those practices on students” (p. 224). Strike (2007) reminds us that “test-based accountability is most likely to succeed if educators employ its results sensibly” (p. 133). Yet in many schools the principal merely becomes the interpreter of performance, reminding stakeholders that “changes in test scores may be less a result of teaching and leadership than a reflection of changes in the groups of students who are tested” (Reeves, 2008, p. 89). Educational accountability is clearly here to stay, and savvy school leaders have no choice but to view this as an opportunity. In the end, the goal must be “to promote the most constructive kind of accountability – the kind that will lead to real improvement in education quality” (Reeves, 2008, p. 90).

Sanctions, Rewards, and Evaluation

Although the evaluation of principals is important for reasons of both organizational and professional growth, there are also potential rewards and sanctions at stake as well. Rewards and sanctions are typically characterized as motivational items that are rooted in the concept of ‘behaviorism.’ “Behaviorism is an idea popularized by

B.F. Skinner, the Harvard clinical psychologist who, in the 1930's, theorized that human behavior is motivated by external stimuli (rewards and punishments)” (Strickler, 2006, p. 26). For years, the evaluation process has been one of the most commonly used methods of allocating the distribution of rewards and sanctions to employees. The reverse is also true, as even today “managers continue the behaviorist strategy of offering rewards and punishments to motivate workers to behave in a prescribed manner” (Strickler, 2006, p. 27).”

“The traditional motivational rule – “What gets rewarded gets done” – has its place, but by itself it is neither powerful nor expansive enough to provide the kind of motivational climate needed in schools” (Sergiovanni, 1992, p. 26). Though motivation to earn reward (and to presumably avoid sanction) can be either extrinsic or intrinsic, the evaluation process is considered to be primarily extrinsic in nature. Interestingly, Strickler (2006) asserts that this limits the effectiveness of evaluation being used as a tool to motivate behavior. “Human beings respond best to intrinsic motivators such as earning the respect of co-workers, . . . having the responsibility of doing one’s own work without supervision or rigid rules and policies, and being held accountable for delivering real results for customers and for the organization within which they work” (Strickler, 2006, p. 28). Similar findings can be found in the work of Herzberg (1968). “According to Herzberg’s theory, once a certain level of (worker productivity) is obtained, hygiene or maintenance factors do not motivate workers to higher performance” (Shen, Cooley & Wegenke, 2004, p. 59). However, “motivators, which include achievement, challenging work, increased responsibility, and recognition, can lead to higher performance” (Shen et al, 2004, p. 59). Yet, “most practical American managers continue to believe that by

pushing the right motivational buttons, they can motivate (manipulate) workers into doing more of what management wants” (Strickler, 2006, p. 27).

In 2002, Andreoni, Harbaugh, and Vesterlund conducted a study on rewards, punishments, and cooperation. Their data indicated that the use of either sanctions or rewards had differing influences on the motivation of their test subjects. “The process suggested by our data is that the stick can help by getting people to move away from perfect selfishness and to test the waters of cooperation. The carrot can then take over by encouraging further cooperation, rendering the stick a rarely used but important and necessary tool” (Andreoni, Harbaugh, & Vesterlund, 2002, p. 26).

Their study helps to highlight the implications of both rewards and sanctions on the principal evaluation process. Andreoni, Harbaugh, and Vesterlund (2002) state that the “results show that when devising incentive systems it is important to recognize that the absence of a reward is not equivalent to a punishment – it is important that both tools be present” (p. 26). They further state that “rewards and punishments act to complement one another and, even though only one can be used at a time, the availability of both tools leads to the greatest degree of cooperation” (Andreoni et al, 2002, p. 25). As such, effective principal evaluation systems must incorporate both rewards and sanctions as necessary, and these rewards and sanctions must be levied on a systematic and deliberate basis in order to affect principal motivation, behavior, and cooperation in the desired manner.

“The theory that measuring performance and coupling it to rewards and sanctions will cause schools and the individuals who work in them to perform at higher levels underpins performance-based accountability systems, now operating in most states and

thousands of districts” (Elmore & Fuhrman, 2001, p. 9). Unfortunately, the use of sanctions is much more common than the use of reward in these accountability systems. Elmore and Fuhrman studied the use of organizational and individual sanctions in 2001, and they found that “negative outcomes included increased pressure and stress to improve results, fear of being labeled a “school in decline,” and the accompanying professional embarrassment, loss of freedom through state-directed assistance or “takeovers,” and expanded work hours” (p. 9). These are not insignificant outcomes for principals.

Separately, Shen, Cooley, and Wegenke (2004) noted several types of rewards desired by administrators: “Workers sought appreciation for completed tasks, input into decision-making, opportunities for promotion and growth, job security, and good wages” (p. 59). While it is important to note that “the distribution of rewards and sanctions within a given accountability system often raises unanticipated problems” (Elmore & Fuhrman, 2001, p. 10), it is still critical that they be incorporated as part of the evaluative system. For effective and equitable principal evaluation, the most important thing is that “in the event of deficient performance, the procedure needs to provide for a mutually agreed-upon plan for improvement; and for superior performance there should be suitable rewards” (Andrews, 1990, p. 3).

Principal Perceptions of Evaluation

It is the role of the principal to “oversee the learning process effectively, manage the flood of paperwork, guide staff development, meet student needs, oversee the financial and physical resources of the school, plan and innovate, manage the crises and disruptions of each day, and be everybody’s friend” (Ginsberg & Thompson, 1992, p. 60). Of all people, the principal herself is the first to know the complexity of the job.

However, not all of those who evaluate principals have an accurate understanding of the role, but they must still appraise the principal's performance. While some principals find this a bitter pill, others embrace evaluation as an opportunity to grow and develop. For example, Ediger (1998) states that "a school leader who is truly professional will desire to have comments pertaining to the improvement of the environment in education and the curriculum" (p. 3). It is not surprising then, to find diverse principal perceptions about the subject of evaluation. These perceptions clearly merit investigation and review.

One area in which there seems to be considerable agreement is the importance of reflection as a component of effective principal evaluation systems. "Reflection – thinking about what we do before, during, and after our actions – is our cognitive guide for growth and development, a way of thinking that we should engage in continuously" (Lambert, 2003, p. 22). Evaluation systems that include an element of introspection make "administrators feel more professional through the inclusion of self-reflection – the evaluation is done by themselves and done unto them" (Brown & Irby, 1996, p. 11). In the circumstances of highly successful principals, Ediger (1998) states that "much reflection on the part of the principal occurs when reaching toward higher levels of achievement" (p. 9). This may be attributed to the practice of successful leaders learning from past experiences. Brown and Irby (2001) state that "principals report that the analysis of past events assists them in becoming more proactive and developing valuable alternatives that enhance program effectiveness and improve schooling" (p. 4). Additionally, principals "believe that to improve both their own performance and, ultimately, students' performance, they must make self-assessment an integral part of their routine" (Brown & Irby, 2001, p. 27). These examples are some of the ways that

principals perceive reflection to be a valuable part of the evaluative process.

As a group, principals also have important perceptions about the feedback that they receive from evaluation. In order for the feedback to have value, Ediger (1998) reminds us that “first, principals should desire to obtain feedback on the quality of performance given” (p. 3). In many cases however, this is difficult. In their study on the politics of principal evaluation (1999), Davis and Hensley found that “principals did not completely trust the formal evaluation process nor the motives or intentions of their evaluators” (p. 399), making the feedback considerably less desirable. Of the evaluation feedback that they received, “principals indicated that most feedback was qualitative and subjective in nature” (Davis & Hensley, 1999, p. 391). The perceived value of feedback is also compromised by the cumulative effect of value-less evaluations over the years. “Over time, principals have perceived that, in general, evaluation systems do not improve performance, do not promote professional growth or school improvement, (and) do not relate to what contributes to principal effectiveness . . .” (Brown & Irby, 2001, p. 5). Some principals also question the value of consistently strong evaluation feedback that does not indicate any substantial areas for improvement or growth. Reeves (2004) asserts that “when an evaluation system makes a leader appear to be uniformly outstanding in every domain of leadership, it is quite likely that the evaluations are deeply flawed or that some domains are missing” (p. 38). Valuable feedback is even important in situations where systems of self-evaluation are the norm. From their study in 2000, Thomas, Holdaway, and Ward state that “(principals) indicated that, although self-evaluations are important, they would appreciate “informed feedback” from their evaluators that extended beyond the self-evaluation” (p. 227).

Many principals also share common perceptions about the input and criteria that are used as part of the principal evaluation process. A study conducted by Rallis and Goldring in 1993 found that “principals believe that the most important influence on their evaluation as a principal by their superiors is running an efficiently administered school” (p. 11). Further, they state the suggestion that these principals “largely believe that they are evaluated by keeping an orderly, tight ship, both in terms of the students (disciplinary environment) and in regard to their teachers and other staff (having efficient administration)” (Rallis and Goldring, 1993, p. 11). Interestingly, the principals in the study also indicated that “parental and community reactions, and student outcomes, such as academic achievement and college admission, have relatively little impact on their performance appraisal” (p. 11).

In 1999 however, Hensley and Davis found principals to generally believe that “positive school outcomes such as high test scores, low dropout rate, and low campus crime had a positive political effect on principals’ evaluations” (p. 396). Principals also “believe that evaluators need to be more aware of the school culture and principal performance based on direct contact with a principal regarding performance” (Thomas, Holdaway, & Ward, 2000, p. 232). Yet a separate study from 1998 dealing with principals’ loss of employment found that ultimately “behaviors relating to a principal’s personal characteristics and relationships with others far outweigh any other (evaluative) factor related to job failure” (Davis, p. 59).

Principals sometimes find themselves conflicted about the criteria that are used to evaluate their performance. Ginsberg and Thompson (1992) quote a principal expressing frustration about the use of too much objective criteria in his evaluation: “I do much more

than can be gleaned from the data sources; How do you document, for example, that I resolve conflict?; It falls short of giving a complete picture of your performance and what you do as a principal . . .” (p. 62). Principals also have differing opinions of the role that standardized testing should play in evaluation. In 2001, the Public Agenda study found that “41% (of principals) agree that standardized tests of student achievement are important and well used in their districts, almost one-half (48%) of principals surveyed think it is a “bad idea” to hold principals accountable for test scores in their own building, and 34% think it is a good idea” (Kaplan, Owings, & Nunnery, 2005, p. 30).

As a group, principals do have strong perceptions about the inputs that they believe are used to evaluate them. Davis and Hensley (1999) interviewed a group of twenty principals about the subject of principal evaluation, and “all principals expressed concern that judgments were being formed by the board or superintendent without accurate information and honest forthright communication” (p. 394). They also state that “every principal strongly felt that a few vocal, opinionated, or influential parents or teachers had the ability to negatively impact their evaluations” (Davis & Hensley, 1999, p. 394). Hart (1992) concurs, stating that “principals continue to believe that superintendents rely most heavily on external measures of performance while reporting that they rely on internal measures” (p. 39). For these reasons, it is not unexpected for Davis and Hensley to conclude that “principals believed that their evaluations were seriously compromised by various political influences and pressures (1999, p. 399). Lastly, Davis (1998) acknowledges that “political pressures placed on a superintendent to remove a problem principal may, in fact, have less to do with the principal’s insufficiencies than with the perceptions, preferences, or particular issues held by a few

influential parents, teachers, administrators, or board members” (p. 80).

Principal perceptions about the quality of their evaluations vary, though there are commonalities. Davis and Hensley (1999) found that “most principals did not find the formal evaluation process helpful in shaping or directing their professional development or in promoting school effectiveness” (p. 399). Their study also asserts that “principal evaluation methods rarely, if ever, included systematic feedback from teachers, parents, or students” (Davis and Hensley, 1999, p. 399). This is a disservice to principals, for it eliminates a viable source of potential evaluation data that could be used in addition to traditional data. After all, “principal work is situational and they face a myriad of expectations, and such work does not lend itself to standardized evaluation practices” (Ginsberg & Thompson, 1992, p. 67). Some principals already feel as though the superintendents with whom they work base evaluation decisions on limited data. “Such biases may lead to incongruent perceptions between superintendents and principals regarding principal competence, thereby raising principal concerns about fairness and about the quality of performance evaluations” (Davis & Hensley, 1999, p. 387). Clearly, effective school principals desire to participate in evaluation processes that provide quality feedback that is accurate and helpful. As Reeves (2004) states, “the best leaders of today and the most promising leaders of tomorrow will not accept an interview for a position that fails to provide an evaluation system that is constructive, fair, and clear – that is, in brief, robust” (p. 25).

Principals also have perceptions about the use of portfolio evaluation systems, in those situations where they are in place. Brown and Irby (2001) state that “principals recognize the merits of portfolios and use them for professional growth, evaluation,

career advancement, and academic progress” (p. 1). There tends to be less concern about equity in evaluation in these situations, presumably because the principal has a large hand in the data that is chosen and utilized for the evaluation. Portfolio systems also offer an opportunity for professional growth that may not be found in traditional evaluation systems. “Principals who have been involved with the development of portfolios report that the processes of selecting viable samples of work and writing accompanying reflections have been beneficial in denoting areas of needed improvement, assisting in maintaining focus, and providing new perspectives and creative insights” (Brown & Irby, 2001, p. 27).

In the realm of principal evaluation, the adage that “perception is reality” holds a good deal of truth. As such, school districts must implement principal evaluation systems that are not only fair and equitable, but also transparent to the individuals to which they are applied. “In the current climate, which emphasizes accountability and school effectiveness, school systems must pay careful attention to the evaluation of principals” (Thomas, Holdaway, & Ward, 2000, p. 235). On their part, principals must also realize that continued professional growth and development are important and defining parts of the evaluative process. As Ediger states, “continuous growth toward attaining more complex objectives should be an ideal of the school leader. The principal has never arrived at being the ideal school administrator but is always moving in the direction of the ultimate” (1998, p. 7).

Constituency Perceptions of Principal Performance

In considering the realm of principal evaluation, it is critical to address the matter of constituency perceptions of leadership performance. Every school principal is

accountable to a variety of potential constituencies, including teachers, parents, students, supervisors, and community members. Likewise, all of these stakeholder groups have their own idea about the type and quantity of leadership that is needed in the school setting. “Principals are role players who operate in an interpersonal arena or role set. Within this role set there are a number of subgroups, each having their own role expectations for the principal” (Duke & Iwanicki, 1992, p. 30). The manner in which these constituency perceptions affect principal evaluation is an issue that must be explored.

Constituency groups have a sizable impact on the manner in which school organizations operate, and principals must be responsive to them. Ginsberg and Thompson (1992) state that “in practical terms, the consumers that principals should respond to would include teachers, students, parents, higher level administrators, board members, and perhaps other staff and community members” (p. 70). Principals must establish, develop, and maintain effective relationships with this broad range of stakeholders in order to accomplish the necessary goals and objectives of the school. In reality, “school leadership is as much (if not more) about relationships with others in the school community, as it is with the more traditional managerial/administrative aspects of the role” (Buttignol & Diamond, 2003, p. 448). In some cases, the need for principals to respond to these constituency groups has burgeoned. Langer and Boris-Schacter (2003) relate how “principals across the country complain that they are constantly accessible by their cell phones, pagers, and e-mail to an ever-expanding constituency” (p. 15).

Yet these constituency groups are also important because of the influence they have on leadership and decision-making. One critical task they accomplish for successful

principals is to validate their leadership. “An effective principal can only achieve influence beyond the ordinary, minimum levels enforced by formal authority when his leadership has been endorsed or legitimized by teachers and other members of the school organization. Validation is a social, not an individual, process” (Hart, 1994, p. 6). Another role that constituency groups play is providing a variety of essential expectations about the key functions and obligations of the principal. Of course, these varied expectations will inevitably invoke stress on the principal as well. “Concerning the variety of expectations, the dilemma for principal evaluation relates to the need to have them considered, while the variety of sources of expectations creates often competing demands” (Ginsberg & Thompson, 1992, p. 65). While some might be concerned that this situation could lead to more negative than positive implications, savvy leaders understand the legitimacy that can be gained from acknowledging and meeting these expectations, if only in part. Duke and Iwanicki (1992) assert that “effective principals are good at scanning the school environment and identifying what their constituencies really expect them to do” (p. 31).

Constituency groups also have perceptions about principal effectiveness, and these opinions can be meaningful to the principal evaluation process. Ginsberg and Thompson (1992) state that “principals must respond to a variety of expectations held by the assorted “publics” they serve . . . (Yet) each constituency has a well-developed and forcefully asserted view of how the school should be run” (p. 63). In order to accomplish their objectives, principals must make frequent choices that have effect on constituent groups. “These choices among competing goals and principal actions will invariably cause some groups to raise or lower their opinions concerning the principal’s

effectiveness” (Snyder & Ebmeier, 1992, p. 77). Still, “no consensus on what principals can and should do is easily drawn” (Ginsberg & Thompson, 1992, p. 61).

The manner in which principals maneuver through the politics of constituency expectations is often a predictor of their success or failure in the evaluation process. Ginsberg and Thompson (1992) state that “because of the type and nature of contact between principals and these individuals and groups, it is essential for effective evaluation that role expectations be explicit” (p. 64). Explicit role expectations are also helpful in buffering principals when being pulled between incompatible interests.

Regardless, “for principal evaluation to be useful, principals must be knowledgeable about the expectations held by various reference groups” (Ginsberg & Thompson, 1992, p. 64). In those cases when constituency groups are given the opportunity to provide feedback, “upon examining the information the principal may notice what is desired from respondents. It behooves the principal then to implement the findings . . .” (Ediger, 2001, p. 6). Unless principals are willing to take action when confronted with their feedback, constituency groups will perceive the principals in a negative light. For although constituent groups often have conflicting opinions and agendas, they usually share the similar expectation that principals “must exhibit characteristics that motivate teachers, students, and parents to higher levels of involvement and ultimately improved student achievement” (LoVette & Watts, 2002, p. 4).

In acknowledgement that constituency groups often have worthy input to share, some researchers advocate soliciting their feedback for the purpose of principal evaluation. Whaley (2002) states that “information gathered from teachers, parents, and

students provides valuable insight to how well principals are performing on the job” (p. 193). He further states that effective principal evaluation “can incorporate 360-degree processes, seeking feedback from multiple sources” (Whaley, 2002, p. 187). In order to promote principal growth, Gil (1998) suggests the importance that “regular surveys of community, staff and students are conducted and feedback is considered seriously and incorporated for improvement actions” (p. 29). Ginsberg and Thompson (1992) take the recommendation one step further, encouraging that constituency groups “through surveys, narratives, and interviews or other means should provide evaluative information concerning both processes utilized and outcomes” (p. 70). Specifically, they state that “student, parent, and teacher surveys or interviews conducted by supervisors could form the basis for . . . data collection” (Ginsberg & Thompson, 1992, p. 70).

Teachers appeal to some researchers as having “preferred” status among constituent groups. Although they did not advocate for the evaluative contributions of other constituent groups, LoVette and Watts (2002) assert “that a plausible method for determining the effectiveness of principals would be to conduct assessments based on teacher perception” (p. 5). Teachers tend to agree that they have valuable input to share. In a 1988 study of the perceptions of Texas elementary principals and teachers regarding principal evaluation, “teachers felt far more strongly than principals that teachers should have a significant input into the evaluation of their principals” (Mullins et al, 1988, p. 7).

Some caveats are offered about the use of constituent input toward the principal evaluation process, however. For example, “people who are either very upset or very happy with your district are most likely to complete the surveys” (Whaley, 2002, p. 193). Peterson (1991) also cautions about the manner in which input is collected, because

“those who are supervised by the principal should, of course, enjoy anonymity” (p. 4). Still, the benefits of considering constituent perceptions within the principal evaluation process do have merit. In the real world, “expectations (of the principal) go beyond technical competence to include such elusive qualities as interpersonal style, values, beliefs, and judgment. The extent to which an administrator meets expectations is a function of peoples’ perceptions . . .” (Duke & Iwanicki, 1992, p. 34).

Principal Evaluation and Pupil Achievement

If the most critical task for the school is to provide high-quality instruction, then the most critical outcome for the school must be student achievement. Though the research varies on the amount of impact that the school principal actually has on pupil achievement, it is still the main emphasis of the larger objectives of teaching and learning. Performance appraisal for the principal plays a role as well. For it is true with both teachers and principals that “when the focus of supervision is on teaching and learning, evaluation is an avoidable aspect of the process” (Sergiovanni, 1995, p. 282).

Researchers are divided on the degree of impact that the school principal has on pupil achievement, although more recent research emphasizes a larger influence than that of early research. Glasman (1992) for example, indicates that the role of the principal in pupil achievement varies with the circumstances of each school. “Principals in improving schools tend to “own” the problem (of test scores) more than principals in declining schools. The latter group tends to delegate responsibilities in dealing with the problem, or to claim that it is not under their control” (p. 113). Further, “principals in improving schools tend to . . . collect a large amount of data and analyze these data alone and with other staff members. Less involvement in such activities is reported by most

principals in declining schools” (Glasman, 1992, p. 116).

Snyder and Ebmeier (1992) report that the effect of the principal on student outcomes is only indirect, at best. In their study, they state that “principal behaviors had significant direct effects on all teacher outcomes and on all teacher perceptions of school functions. On the other hand, principal behaviors did not have significant direct effects on any student outcomes . . .”(p. 96). While the impact of the principal on teachers is considerable, “a principal’s work often is decoupled from the instructional process, and the principal apparently exerts little direct control over learning or attitude formation – at least at the individual student level” (Snyder & Ebmeier, 1992, p. 101). However, “just because the effects of principals are mediated by other school factors does not diminish the importance of principal contributions to school effectiveness” (Hoy & Miskel, 2001, p. 302).

Assertions about the limited value of principal leadership to student achievement are countered by the research of Kaplans, Owing, and Nunnery (2005). They state that “although the principal’s effect on student achievement may be indirect, it is crucial. The principal controls the most important factors affecting a school’s teaching and instructional quality . . .” (p. 29). They also note that “similarly Waters, Marzano, and McNulty’s (2003) meta-analysis on 30 years of research on the effects of principals’ practices on student achievement found a significant, positive correlation of .25 between effective school leadership and student achievement” (Kaplan et al, 2005, p. 29. “Thus, principals of successful learning communities work with their staffs to articulate clear and measurable goals, to identify indicators that offer evidence of progress, and to develop systems for monitoring those indicators on a continuous basis” (DuFour &

Eaker, 1998, p. 194).

Waters, Marzano, and McNulty (2003) add, “given the perceived importance of leadership, it is no wonder that an effective principal is thought to be a necessary precondition for an effective school” (p. 5). This is important because “whether a school operates effectively or not increases or decreases a student’s chances of academic success” (Waters et al, 2003, p. 3). By this line of logic, the presence of principal leadership skills does have an impact on student academic success, even if tangential. Waters et al (2003) are very clear that this effect on students can actually be “profound” (p. 32) in nature.

How is pupil achievement typically measured, however? Four common measures of pupil achievement are in conventional use, including standardized testing, formative assessment, student grades, and the use of class data. Class data is used when educators gather a range of information about student progress, and then use the information to adjust instruction. “Rather than being tools for rating, ranking, sorting, or humiliation, data displays . . . are celebrations of teacher effectiveness” (Reeves, 2006, p. 89) when the information is used in this way. Another kind of group data is “value-added,” in which individual student and cohort achievement are measured over time in order to determine the growth impact of each year of instruction. Kaplan, Owings, and Nunnery (2006) state that “value-added has become an important, empirical way of noting educators’ effect on student achievement” (p. 29). In addition, “empirically connecting the consistency of principal leadership with their schools’ student achievement would offer opportunities for principals to receive important feedback about their professional effectiveness on crucial school leadership dimensions” (Kaplan et al, 2006, p. 29).

Student grades are another gauge by which pupil achievement is often determined. Depending on the circumstances, this may or may not be appropriate. “Letter grades do not reflect student achievement in an astonishing number of cases” (Reeves, 2006, p. 113) because of the use of some questionable classroom grading practices. Reeves states that “amazingly, teachers regularly use and leaders frequently tolerate grading systems that may appear to be accurate but are devoid of even the most basic elements of mathematical reasoning and are neither fair nor effective” (2006, p. 119). In cases such as Reeves describes, pupil achievement is not what is reflected in a class grade, but often some other objective, such as effort, work completion, or participation. As such, principals and other educational leaders must be wary of utilizing student grades as a true indicator of pupil achievement.

Another manner by which pupil achievement is typically determined is through the use of formative assessments. Formative assessments meet an important need in furthering understanding about student progress during the course of the instructional process. “Assessments, in contrast to tests, are formative, provided during the year, designed to improve teaching and learning, and accompanied by immediate feedback” (Reeves, 2006, p. 86). This measure is a highly valuable indicator of student achievement, but it must be facilitated through principal support. Reeves (2006) underscores that “assessment informs teaching; leadership provides the time and resources for teachers to respond to assessment results; and students use assessment feedback as a series of cues for improved performance” (p. 87). It is critical to note that each of the respective components is vital in reinforcing the effectiveness of this measure of pupil achievement.

The most commonly used measure of pupil achievement is standardized testing. Standardized tests are intended to provide data that is valid, reliable, and based on statistical norms at a variety of levels, including national, state, or local. Daresh (2002) explains that he believes the “increasing emphasis on statewide tests of student achievement” is primarily based in the desire for teacher accountability (p. 84). Though they are widely accepted by some advocates as firm evidence of both school and student performance, standardized tests do have certain limitations that must be acknowledged. As only part of the story regarding pupil achievement, “leaders should neither ignore test scores nor embrace them as the sole indicator of student learning, teacher ability, or school quality” (Reeves, 2004, p. 30). Sergiovanni (1995) states that this is because “not all desired student outcomes can be accounted for by such tests, and not all such desired student outcomes can be specified with precision” (p. 208). Interestingly, over-reliance on standardized tests can also lead to dysfunctional behaviors on the part of principals. Reeves (1998) asserts that “it doesn’t take most principals long to determine that if an accountability system rewards only test scores, then the easiest way to look good is to find a school with a record of high achievement . . . and denying leadership in schools where it is most needed” (p. 6). In the end, “an organization that values only test scores will unwittingly validate mediocrity and ineffective leadership practices because it fails to differentiate among the many leadership variables linked to student performance” (Reeves, 2004, p. 31).

The primary question is whether or not principal evaluation can be used to improve pupil achievement. Again, some of the answer depends upon the degree to which researchers believe principals have the ability to affect student academic

performance. Snyder and Ebmeier (1992) assert that “principals can be evaluated directly in terms of their effects on teachers but only indirectly for their effects on students and parents” (p. 76). They attribute this to the fact that principals are not given direct responsibility for many of the “factors of production” that directly impact student learning. “If principals were afforded more control over the input variables such as staff selection and budget authority, and if school outcomes were clearly defined, then principals might have more control over achievement . . . and could more reasonably be accountable for student outputs” (Snyder & Ebmeier, 1992, p. 102). In order to contribute to pupil achievement, it is also critical that effective evaluation instruments and procedures are utilized. Unfortunately, “common evaluation schemes, on the other hand, often emphasize processes such as student behavior management and control and communication skills” (Hart, 1994, p. 7). In order to have impact on pupil learning, the principal evaluation process should ensure that “the principal is responsible for monitoring teaching and learning in her or his school and does so by visiting classrooms, touring the school, talking with people, and visiting with students” (Sergiovanni, 1995, p. 284). Unless principal evaluation is rooted in student achievement, it will not yield administrative practices that support student achievement.

Most researchers assert that principals do have a considerable impact on pupil achievement, and they enumerate ways in which evaluation can be used to facilitate student academic performance. Glasman (1992), for example, addresses nine different evaluative items that can be included to “assess principals’ performance for accountability” of student achievement:

- (a) becoming aware of student achievement problems

- (b) recognizing the importance of the problems
- (c) recognizing the pressure to solve the problems
- (d) gathering pertinent data about the problems
- (e) considering obstacles and opportunities to solve the problem
- (f) choosing alternative solutions
- (g) choosing ways to implement the chosen solutions
- (h) allocating resources (money, time, personnel, space) to the implementation of the problems
- (i) monitoring the implementation of the solutions” (p. 120)

It is also vital that school principals are evaluated on their ability to implement academic improvement strategies that have a high cost/benefit ratio. As Marzano, Waters, and McNulty (2005) have noted, “the school leader’s ability to select the right work is a critical aspect of effective leadership. It might be the case that teachers and administrators in a low-performing school are working “hard” but not working “smart” in that they select interventions that have little chance of enhancing student academic achievement” (p. 97).

Another consideration that is needed to ensure that principal evaluation remains focused on pupil achievement is the element of standardized testing. The Oklahoma State Department of Education includes the following indicator in their Oklahoma Criteria for Effective Administrative Performance instrument: “The administrator provides a written analysis of student test scores and other data to assure that the various student populations are benefiting from the instructional program” (1999, p. 26).

Effective principal evaluation tools must also put forth the expectation that principals

should review and analyze the instructional program to maximize student learning. Hart (1994) states that “superintendents and other supervisors can examine principals’ use of organizational analysis techniques that can enhance their success as school leaders and provide opportunities to promote the instructional practices and goals valued by the school district” (p. 8). Lastly, evaluation that is based in pupil achievement must afford principals with the chance to be reflective about their practice. Without this component, evaluation will become mired in management-based competencies. “Perhaps most important, reflective principals do not confuse evaluation processes with the substance of evaluation” (Sergiovanni, 1995, p. 211). With the benefit of reflection, principals are able to implement the strategies and initiatives that are required to further pupil achievement, without excessive anxiety about the evaluative implications of those interventions.

Although researchers are divided on the extent to which principals affect pupil achievement, it is clear that evaluation is one vehicle that can be used to help maximize principal impact on student learning. Focusing on standardized test scores is one way to help guide principals’ efforts. “As a weather forecaster considers data on wind, temperature, history, and personal experience, so also must those evaluating educational leaders consider test scores as an important piece of data, but not the only piece on which they will predict future leadership performance” (Reeves, 2004, p. 30). Sergiovanni agrees that standardized test scores must be part of a more balanced approach: “reflective practice in school evaluation requires a far more complex view than is associated with simple outcome-based conceptions and with the measurement stance” (1995, p. 210). In the end, it seems that the most viable way for principals to effect pupil achievement is by

showing leadership with teachers, which should also be an important part of every principal's evaluation. "If a teacher's work focuses on enabling students to know and be able to do what is laid out in student standards, then a principal's work is to enable teachers to be successful in accomplishing that" (Kearney, 2005, p. 20). Unless principals are held accountable in playing this role, pupil achievement will undoubtedly suffer.

Summary

A scholarly review of the literature identifies a number of vital themes. One of these themes is the primacy of effective leadership on the part of the principal. Principals must exert leadership that is appropriate to the educational setting, ethical in nature, and based upon the values of fairness and trust. Effective principals also place emphasis on the importance of relationships and technical knowledge. The essential principal leadership competencies have been successfully distilled in the Interstate School Leaders Licensure Consortium Standards (ISLLC) for School Leaders. The principal evaluation system is a useful vehicle by which school districts can identify and stress the importance of principal leadership behaviors.

Another key responsibility for effective principals is that of staff supervision. Critical supervision tasks include collaborative performance evaluation of faculty and staff, the coordination of appropriate staff development, and human resource planning. Effective principal evaluations address the manner in which these tasks are accomplished efficiently, ethically, and with deliberate purpose.

A number of problems typically exist with principal evaluation systems. Three of the most common issues are principal-based problems, instrument-based problems, and

lack of clarity in expectations. Other problems with principal evaluation include ambiguous evaluation objectives, difficulty evaluating competencies of the position, and lack of focus on professional growth and development for the principal. While all of these problems bring forth unique challenges in the task of evaluating the principalship, they are not sufficient causes to overshadow the importance of conducting the evaluative process.

As such, the improvement of principal evaluation systems is needed. One of the recommendations to accomplish this objective includes insuring that the various stakeholder groups are involved in the process of developing the evaluation systems. Another recommendation is to address the degree to which the evaluation procedures and instruments are used appropriately, fairly and with proper regard for people and process (called internal consistency). A third recommendation for improvement is to ensure that the principal evaluation system is grounded in clearly defined policies and procedures. Other recommendations include the need to ensure alignment with professional standards, including an appropriate goal-setting component, and making sure that evaluations are conducted within the proper context of the principal's role and circumstances. More recommendations are ensuring that evaluation systems rely on a variety of data, increasing emphasis on the importance of self-reflection, and requiring that evaluations be squarely based on valid job descriptions.

For many districts, the development of principal evaluation systems is a central objective. The first consideration in this process is to identify those components that are necessary to accomplish the goals of the evaluation system. Secondly, it is important to identify which participants should be involved in the evaluation development process.

The nature of the process that is utilized should also be considered, as should the need to clearly identify those standards that will serve as the basis of the evaluation. Lastly, establishing fair instrumentation and procedures is critical, as well as piloting the new program properly. As principal evaluation systems are a reflection of the goals and objectives of the larger organization, the development of a new system is a process that must be undertaken with careful planning and a firm philosophical foundation.

Standardized testing and school accountability have become factors in principal evaluation, as with many aspects of education. As such, achievement test indicators have been instituted as the stick by which to measure both school and principal performance. In Illinois, the standardized test that is used to measure student achievement is the Illinois Standards Achievement Test (ISAT). The use of standardized testing has presented both opportunities and challenges for school districts, and standardized test scores are now often used as a criteria in the evaluation of principals.

Although the evaluation of principals is important for reasons of both organizational and professional growth, there are also potential rewards and sanctions at stake as well. To be effective, principal evaluation systems must incorporate both rewards and sanctions as necessary, and these rewards and sanctions must be levied on a systematic and deliberate basis in order to affect principal motivation and cooperation in the desired manner.

Principals have diverse perceptions about the subject of evaluation. Important considerations about evaluation include the need for professional reflection, the value of quality evaluative feedback, and the significance of the input and criteria that are used as part of the principal evaluation process. In order to improve principals' perceptions,

school districts must implement principal evaluation systems that are not only fair and equitable, but also transparent to the individuals to which they are applied.

Constituency groups have a sizable impact on the manner in which school organizations operate, and principals must be responsive to them. Yet these constituency groups are also important because of the influence they have on leadership and decision-making. In addition, constituency groups have perceptions about principal effectiveness, and these opinions are often utilized as part of the principal evaluation process.

Researchers are divided on the degree of impact that the school principal has on pupil achievement, although more recent research emphasizes a larger influence than that of early research. Four common measures of pupil achievement are in conventional use, including standardized testing, formative assessment, student grades, and the use of class data. Conflicting research indicates that it is difficult to discern whether or not principal evaluation can be used to improve pupil achievement. Unless principal evaluation is rooted in student achievement, it will not yield administrative practices that support student achievement.

Principal evaluation is a multi-faceted issue upon which many influences play. It is impacted by organizational, individual, professional, and cultural influences, and has evolved as evaluation processes and systems have changed over time.

CHAPTER THREE METHODOLOGY

Introduction

In broad strokes, the purpose of this study was to investigate the topic of the evaluation of school principals and the role it plays in fostering student achievement. The overarching research question for this study was the following: “Is there a relationship between the type and intensity of principal evaluation and pupil performance in DuPage, Will, and Lake County, Illinois public schools at the K-8 level?”

This question contains two components, both of which are central to understanding the larger issue. The initial objective was to determine the type and intensity of the methods that are in use for the purpose of principal evaluation in DuPage, Will, and Lake County, Illinois public schools at the K-8 level. This objective was best accomplished through the use of a survey designed for this purpose, and was administered to the principals (Appendix A). The second objective was to correlate the identified evaluative methods with pupil performance, which were based upon Illinois Standards Achievement Test (ISAT) school scores for the 2007-2008 school year. This standardized assessment is administered to all public school students in the state of Illinois for grades 3-8, and the scores are publicly available information. It is important to note that only principals who have served for at least two years at each data collection site were considered viable participants in this study, because the standardized

assessment scores could not be correlated without establishing an overlap between principal employment and student testing.

To supplement the primary research question, the evaluative methods that were identified were also correlated with the principals' perceptions as they pertained to two specific topics. The first topic addressed the principals' perceptions about the influence of principal evaluation on pupil performance. The second topic addressed the principals' perceptions about the influence of principal evaluation on their own professional performance. Since this research study utilized simultaneous mixed methods, open-ended questions were nested within the larger survey that allowed respondents to indicate their perceptions about the effect of the type and intensity of principal evaluation on student achievement, as well as their professional performance (Appendix A). The qualitative data that was acquired from this component added richness and validity to the study.

The main constructs for the survey are listed below, and include all of the following:

- *The type/intensity of principal evaluation methods that are in predominant use in DuPage, Will, and Lake Counties*
- *The principals' perceptions of the relationship between school performance level on the ISAT assessment and type/intensity of principal evaluation*
- *The principals' perceptions about the relationship between principal evaluation of pupil performance and type/intensity of principal evaluation*
- *The principals' perceptions about the influence of principal evaluation on their own professional performance and type/intensity of principal evaluation*

Research Design

In determining the research design for this study, several issues were considered. The first of these issues was the desire to gather the data through a survey format. Creswell states that “a survey design provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population” (2003, p. 153). Although all manner of research possess shortcomings, “those shortcomings are made clearer in survey research than in other social research methods, thereby permitting more considered evaluations of their implications” (Babbie, 1990, p. 40). In this study, a random sample was not drawn from which to collect data. Instead, all individuals in the target population were asked to participate. Still, those who elected to participate likely had varying reasons to do so thus introducing possible differences in the sample compared to the underlying population.

Survey research offers a logical approach to “test complex propositions involving several variables in simultaneous interaction” (Babbie, 1990, p. 41.) Survey research also offers the ability for the researcher to tailor the questions included in the data collection instrument to operationalize the specific research questions under study. If the survey constructs are carefully and thoughtfully defined, and the subsequent survey questions are closely aligned to them, it is possible to create a survey instrument that is well suited to the unique needs of the study being conducted. Creswell (2003) speaks to the “economy of design” (p. 154) that survey research offers in attempting to address the specific constructs that are being investigated. Though a separate issue, the “rapid turnaround in data collection” (Creswell, 2003, p. 154) inherent in survey research is another appealing

aspect of this methodology.

Lastly, the main purpose of survey research is to “generalize from a sample to a population so that inferences can be made about some characteristic, attitude, or behavior of this population” (Babbie, 1990, p. 43). The information that was obtained through the use of the survey instrument was valuable for the implications it holds within the realm of education for the areas of principal evaluation and student achievement. As Creswell asserts, survey research provides the “advantage of identifying attributes of a large population from a small group of individuals” (2003, 154).

In this study, it was also desirable to gather qualitative perceptions from the respondents. This need required the study to utilize a mixed method approach. Mixed methods research procedures offer the researcher the ability to use “both open- and closed-ended questions, both predetermined and emerging methods, and (the potential to) analyze multiple forms of data drawing on all possibilities” (Creswell, 2003, p. 17). The main benefit of a mixed methods approach to this research is that the qualitative aspect added needed depth, richness, and context to the data.

As a result of these considerations, the research design that was employed was the “Concurrent Nested Strategy.” This mixed methods research approach “can be identified by its use of one data collection phase, during which both quantitative and qualitative data are collected simultaneously” (Creswell, 2003, p. 218). In this approach, a predominant method exists (quantitative) that guides the research, while a secondary method (qualitative) is “nested” within the larger study. This study investigated the relationship between student achievement and the type and intensity of principal evaluation as the predominant component, while the principals’ perceptions about the

influence of principal evaluation on student achievement and professional performance served as the “nested” component.

This study was cross-sectional, and was self-administered in an online format. The online format was selected primarily because of its cost-effectiveness in administration for a large group of potential respondents. Also, there are additional strengths and weaknesses of this format. Strengths include the ability of the researcher to receive data in a time-sensitive manner, the ability to target recipient groups directly, and the convenience of administration. Weaknesses include the difficulty of securing respondent e-mail addresses, the risk that the survey might be interpreted by respondents as “junk mail,” and the possibility that respondents may not be technologically savvy enough to use the survey or distrust the link provided to connect to the online survey driver.

Research Site

The research sites for this study were the respective K-8 public schools in DuPage, Will, and Lake Counties, three populous “collar counties” outside of Chicago, Illinois. With the exception of those schools excluded as described in the subsequent “Participants” category, all of the public schools in these counties that contain the grades K-8 were included as research sites, under two conditions. First, all of the schools included in the study must administer the ISAT to students in at least one grade level in the testing range (presently grades 3-8). Second, the schools must be considered “regular attendance centers” for students in their respective communities. For example, alternative schools for discipline or special education centers for students with profound disabilities were not considered research sites.

At the time, there were 199 public schools in DuPage County from 42 separate school districts that met the criteria. Likewise, there were 122 public schools in Will County from 29 school districts meeting the criteria. Finally, there were 159 public schools in Lake County from 48 school districts that met these criteria. It is important to note that the schools that were eligible as research sites are configured in many different ways. The configurations included: grades PK-8, grades K-3, grades K-4, grades K-5, grades K-6, grades K-8, grades 1-3, grades 1-5, grades 2-5, grades 3-5, grades 4-5, grades 5-6, grades 5-8, grades 6-8, and grades 7-8. While non-identifiable demographic data was collected to allow for stratification of results, the school configurations were neither a determinant of inclusion nor exclusion for the study. However, all schools included in the study enrolled students ranging from K-8, were considered “public” in nature, were considered “regular attendance centers,” and administered the ISAT to at least one grade level of students.

Participants

The participants for this study were public school K-8 principals practicing in DuPage, Will, and Lake Counties, in the State of Illinois. These generally affluent counties outside of Chicago have 480 public schools serving students in grades ranging from K-8 in a number of different program configurations, with all schools serving at least one grade level between 3-8. However, as public school principals, all of the target participants were required by the Illinois School Code to receive evaluation based upon their job performance. Because of their mutual employment in DuPage, Will, and Lake Counties, all of the principals work under the purview of their respective Regional

Offices of Education. Yet, the Illinois School Code subjects them all to the same applicable evaluation requirements (105 ILCS 5/24 A-1).

Four DuPage County principals were excluded from the study in advance. Two of the exclusions consisted of the researcher himself and his school district colleague. The other two exclusions were principals in the community in which the researcher resides. However, the researcher has no known relationship with any of the other potential participants in this study. As a result of the aforementioned exclusions, the total number of school principals recruited to participate was 476.

Procedures for Data Collection

The procedures used to conduct the study adhered to the rigorous standards of survey research. As stated previously, the survey was conducted in an online format (Appendix A). This format required that the researcher attained all participants' e-mail addresses, which were accessible through publicly available information at the Illinois State Board of Education. A federal Freedom of Information Act (FOIA) request was made to acquire this information. An e-mail message was then sent to potential participants that provided an introduction to the survey, with a link to the survey instrument included within the message. The link then led the participant directly to the consent form, which was embedded in the first page of the online survey. The message included information about the purpose of the survey as well as language necessary to conform to the principles of ethical research (see content in "Ethical Issues in Data Collection" below).

After the survey was available for several days, an additional e-mail was sent to

request the research subjects to participate, if they have not already. Another e-mail reminder was sent several days later. Lastly, a final e-mail was sent to inform non-respondents that the survey was closing on a specific date, either in 24 or 48 hours. The survey was then closed to participants when the window had expired.

In order to protect the confidentiality of participants, the instrument was only sent and available to those respondents in a group with similar “school ISAT performance levels”. In order to determine “school ISAT performance level”, the percentage of all students ‘meeting expectations’ or ‘exceeding expectations’ on each of the tested ISAT areas (by subject and grade level) was calculated. These percentages were averaged at the school level, resulting in one aggregate “school ISAT performance level” score. Once the “school ISAT performance level” scores were identified for each school, three performance tiers were devised with an equal number of data collection sites. Three different links to the secure computer server were used to differentiate the groups and collect the data. While the surveys that each group received were identical, the method of providing different survey links worked well to effectively sort the responses by performance tier. As each group of survey responses was collected, the online survey closed. This method allowed the researcher to collect data within each of the ISAT performance groupings without compromising the confidentiality of the respondents. The members of each group did not know which tier they were in, nor did the researcher know the identity of the respondents. School ISAT scores were collected through the use of the Illinois Interactive Report Card website (www.iirc.niu.edu), and the Illinois State Board of Education website (www.isbe.state.il.us).

Additional steps were also taken to ensure security of the data, such as storing

the data on a secure server, using 'https' encryption technology, and ensuring that the researcher was the only individual with access to the information. The online survey driver that was used to collect the data was selected carefully. The most important criterion for selection was the requirement that the survey driver have rigorous security standards.

The final data was ultimately collected at the conclusion of each online survey window. The greatest concern at this stage of the study was the response rate. If an adequate number of responses were not obtained, the online surveys would have had to remain open until that threshold was attained. If necessary, the researcher would have sent additional follow-up emails reminding respondents to participate.

Once each survey window closed, the online survey tool provided a composite overview of the quantitative data. Online survey tools do have the capacity to perform basic data analysis functions, but they do not provide for the complexity of data analysis that is required in this study. As a result, the final quantitative data was converted and downloaded into SPSS for the purpose of analysis. The qualitative data also required transformation (coding) in order to be analyzed.

Data Analysis

Data were downloaded from the survey driver into SPSS and analyzed. Techniques for the imputation of non-systematic missing data were used so that any bias resulting from the absence of data was minimized.

Regarding the principal evaluation component of the research question, frequency counts were measured and descriptive statistics calculated (means, modes, and standard deviations). Within the survey instrument, a matrix was used to collect data on both the

type and intensity of principal evaluation. A seven-point matrix was used to measure the degree of the participants' responses. A second matrix was used to collect data pertaining to the degree to which the participants perceived that principal evaluation accomplishes the objectives for which it is intended. As a result, Cronbach's alpha was used to determine the degree to which these items were interrelated on the inter-item correlation matrices.

The second part of the research question attempted to determine the degree to which the evaluation methods correlated with student performance on the ISAT. "The measure of the degree or strength of the relationship (between two variables) is represented by a correlation coefficient" (Howell, 2004, p. 164). Correlation is determined through the use of bivariate statistics; it was expected that a Pearson product moment statistic would be useful in determining the strength of the relationship as well as the statistical significance.

Supplementing the overarching research question, respondents were asked to provide qualitative data about their perceptions through the use of a limited number of open-ended questions. For the analysis, these data were coded, described, and interpreted in an appropriate theoretical context. As these qualitative data were embedded ("nested") within a predominantly quantitative research study, analysis had to combine the two forms of data to "seek convergence among the results" (Creswell, 2003, p. 222).

Ethical Considerations

There were numerous ethical issues that had to be considered in the collection of the data, and all of them had to be communicated to potential respondents. One of the most important issues was that of voluntary and informed consent, which underscores the

fact that participants have no obligation to participate and that they may elect to cease participation at any time. Another consideration mandates that participation in the study cause the participant no harm. As part of this, research subjects had to be informed of any potential negative ramifications that might occur as a result of their participation.

Participants also had to be informed of the sponsorship and genuine purpose of the survey. The fact that a Loyola graduate student was conducting the research for the purpose prescribed had to be communicated very clearly. Reporting of results had to be addressed as well, and respondents made aware of the manner in which the results would be available to them.

Though these were not the only ethical considerations that had to be taken into account when collecting data, they represented some of the most vital concerns. When reviewing the research proposal, the Loyola Institutional Review Board for the Protection of Human Subjects also required other specific actions that had to be undertaken to protect the research subjects.

Prevention of Bias

As an acting school principal, it was vital to acknowledge the possibility that the researcher might have identified personal bias while undertaking this study. Though he has no known predilections in the topic that was researched, the researcher was responsible to provide for measures to prevent bias from affecting the results of the study.

In survey research, the greatest danger from bias stems from the manner in which the survey questions are worded and presented. Babbie (1990) asserts that “survey data are created, rather than simply collected” (p.131). What is meant by this claim is that the manner in which the questions are asked often influences the participants’ responses.

In order to avoid this danger, it was essential to “carefully examine the purpose of (the) inquiry and construct items that will be most useful to it” (Babbie, 1990, p. 132). It was critical to ensure that the constructs for the survey were linked directly to the formulation of the survey questions, and that the questions did not lead to an implication that certain views represent preferred ways of thinking. The risk that the survey instrument might lead respondents to believe that viewpoints may either be “right” or “wrong” based upon the way in which the questions were asked was minimized.

So that the risk of unexpected bias was minimized, the researcher solicited the input of the dissertation director and committee members regarding the content of the questions contained within the survey instrument.

Validity and Limitations

In survey research, each survey instrument is uniquely created in order to collect data relative to the specific constructs. As such, no two survey instruments are identical (unless they are addressing the exact same constructs). This fact leads some to question the reliability and validity of survey instruments in use. However, reliability and validity concerns were minimized in this research survey through the use of three distinct practices.

First, the dissertation committee was solicited for input regarding the wording of the survey questions in relation to the survey constructs. The question was posed: “Are these survey questions crafted so that they garner the required information about the constructs from the participants?” Questions could not be confusing, misleading, or biased, and peer-review of the survey questions helped to minimize these risks.

Second, the survey was pre-tested with a small group of respondents. The

respondents were then provided with the opportunity to share feedback about the integrity of the instrument and its fidelity to those constructs being measured. This was the ideal point at which to identify potential threats to internal validity, had they existed.

Lastly, the survey was piloted with an independent group of respondents who were not in the survey participation group, but have like background. The results of this administration were highly valuable in determining the degree to which the survey functioned as designed. Although these steps were obviously unable to either guarantee the validity or reliability of the instrument, they increase confidence in the outcome.

Summary

The method that was employed in this survey was intended to provide for both quantitative and qualitative feedback about the research question. By utilizing the survey in this manner, the researcher was able to “gain broader perspectives as a result of using the different methods as opposed to using the predominant method alone” (Creswell, 2003, p. 218).

Creswell emphasizes that another advantage to the use of the described method is that the qualitative data would be able to “be used to describe an aspect of a quantitative study that cannot be quantified” (2003, p. 218). The questions about the respondents’ perceptions of principal evaluation added depth and richness to the data that was collected about the type and intensity of the evaluations they have experienced personally. This was very helpful in assisting the researcher to determine and understand some of the shades of perspective that were reported by the respondents.

In considering the research method for this study, primary considerations included ethical research practices, proper survey/question development, participant response rate,

effective data collection, and careful data analysis. Each of these concerns was essential in providing the structure for a properly conducted research study.

CHAPTER FOUR ANALYSIS OF DATA

This study was created to investigate the topic of school principal evaluation and the role it plays in fostering student achievement and principal professional development. The study has illustrated that there are a number of important purposes for principal evaluation, that the methods of principal evaluation differ substantially from one school district to another, and that the issue of principal evaluation is a matter that merits more thorough investigation.

After a comprehensive review of the literature, it was also established that principal evaluation plays an important role in the manner in which public schools are effectively managed. Though often applied in varied forms and with inconsistency, principal evaluation instruments have the capacity to foster positive personal and organizational outcomes. While principal evaluation systems can suffer from a wide range of problems in both development and implementation, they do offer promise in helping school administrators improve their instructional impact on students, as well as their own personal leadership skills and management abilities.

The main research question for this study addressed whether a relationship exists between the type and intensity of principal evaluation and pupil performance, specific to the setting of K-8 public schools in DuPage, Will, and Lake Counties (IL). To supplement the primary research question, the identified evaluation types were correlated

with the principals' perceptions as they pertained to two distinct topics. The first of these topics was the matter of principals' perceptions about the effectiveness of principal evaluation on principal performance and professional development. The second topic was the issue of principals' perceptions about the impact of principal evaluation on pupil performance and professional development.

Before reporting the statistical test results on these topics, key aspects of this study must be described in further detail. This chapter first addresses the accuracy of the findings, and then provides clear and thorough description of the survey and the sample. Following this, the demographic data is reported, as well as an overview of the data reduction. Finally, quantitative and qualitative data analyses describe the results.

Accuracy of Findings

Because of the nature of this study, it is important that the issue of the accuracy of findings is addressed. While there are numerous strategies that can be used within a research study for this purpose, there are two main strategies employed in this study that can accomplish this goal (Creswell, 2003, p.196).

The first of these strategies is clarifying the bias of the researcher. As with all research, the investigator in this study has bias about the topic of principal evaluation. The researcher is a practicing K-8 public school principal, and has had his own experiences with principal evaluation over the course of his career. While the researcher's personal evaluation experiences have generally been very positive, he has anecdotal knowledge from professional colleagues that runs counter to this. As a result, the researcher brings bias to the study in his belief that the manner in which principal evaluation is conducted varies substantially, and that the evaluation process has differing

impact upon the individuals who are subjected to its' use.

The second strategy that is used to increase the credibility of the findings is to openly present discrepant information. While it is anticipated that research findings will yield varied data, it is important that the consumers of this research can have confidence that any data running counter to the main findings will be genuinely reported in a forthright manner.

Description of Survey

The survey instrument was administered online, and was designed to take approximately 10-12 minutes to complete. It consisted of fourteen separate questions, including two matrices. The first matrix asked the respondents about those evaluative methods that had been used to formally evaluate them during their career. The second matrix questioned the respondents about their perceptions regarding the ability of the principal evaluation process to accomplish a variety of instructional purposes. In addition, respondents were asked about the length of time they had already served as principal in their career, as well as the length of time they had served as principal at their present school of employment. They were asked about the frequency with which they were evaluated, and by whom. Respondents were also asked about the type of school and district in which they worked, and the county in which they currently were employed.

Finally, four open-ended questions were asked addressing the respondents' beliefs. Two of these questions targeted the respondents' beliefs regarding the effect of principal evaluation on principal performance, and also about principal evaluation and professional development. The other two questions were about the respondents' beliefs about the impact of principal evaluation on pupil performance, and about principal

evaluation and their own professional development.

In order to gather data that reflected the ISAT performance level of each respondent's school of employment, respondents were provided with one of three different links to the survey. While the surveys themselves were identical, this allowed the researcher to stratify the responses into three different tiers, with each tier comprising responses from principals at schools with similar ISAT performance levels.

Description of Sample

The population for this study was all of the K-8 public school principals presently employed in either DuPage, Lake, or Will Counties (IL). While the link to the online survey was sent to each of the potential respondents in this population, the survey was only completed by a percentage of them. Sample size is also limited by the requirement that respondents in their first year as principal be excluded from the study. This requirement was established to ensure that each principal was employed at his or her respective school during the year that the student achievement data (ISAT scores) were actually collected. As a result, the final sample size (N) is a subset of the total number of survey respondents. Table 1 represents the cumulative sample size for the study, as well as the sample sizes for each of the sub-groups (Tiers 1-3).

Table 1 – Matrix Display of Sample Size

	Actual Respondents	First Year Principals	Final Sample Size (N)	% of Actual Respondents
Tier 1	54	11	43	.796
Tier 2	51	7	44	.862
Tier 3	50	7	43	.860
Total	155	25	130	.839

For this research study, the respondents were administered an online survey. The

survey was constructed in order to gather both frequency and perception data, and it consisted of both closed and open-ended questions. The data were collected over an eight-day period in April 2009 through the use of a secure, proprietary online survey driver. Potential respondents were emailed and provided with a link connecting the user directly to the survey instrument. The response rate was nearly equal for the three tiers. Table 2 represents the cumulative response rate for the study, as well as the response rates for each of the sub-groups (Tiers 1-3). The table takes into account the diminished number of potential and final respondents once first year principals were removed from the data set.

Table 2 – Matrix Display of Response Rates

	Potential Respondents	Final Respondents	Response Rate
Tier 1	147	43	.292
Tier 2	153	44	.287
Tier 3	151	43	.284
Total	451	130	.288

Descriptive Data - Demographics

The demographic data in this research study reflect several different respondent attributes. These include the number of years that the respondents have served as principal at the school of their present employment, the total number of years that each respondent has served as a school principal, the school structure at the place of each principal's employment, and the frequency of evaluation. Other attributes include the respective job positions of each principal's evaluator, their county of employment, and the number of schools in the district where each principal is employed. When reviewing this data, it is not possible to determine which subjects indicated that they were in their

first year as principal in the school where they were presently employed. As such, the demographics must be reported using the data from the total group of respondents.

Table 3 addresses the descriptive statistics for the research variable relating to the number of years that respondents have served as principal at their current school, while Table 4 displays the same statistics for the research variable about the total number of years that respondents have served as a school principal. Table 3 shows that the mean number of years that the respondents have served in their current school is 5.32, with a standard deviation of 4.5 and a positive skew of 1.65. This skew supports the clustering of scores found on the low end of the scale. The positive kurtosis of 2.86 indicates that the distribution of scores is peaked in nature.

Table 3 – Descriptive Statistics, Principal Years at Current School

	N	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis		
								Std. Error	Std. Error
								Stat.	Stat.
Years at Current Sch.	138	1	22	5.32	4.503	1.650	.206	2.863	.410
Valid N (listwise)	138								

Table 4 relates that the mean number of total years that respondents have served as a school principal is 8.36, with a standard deviation of 6.64. Once again, a positive skew of 1.27 indicates that the scores are clustered on the low end of the scale. The positive kurtosis of 1.19 indicates that the score distribution is somewhat peaked in shape.

Table 4 – Descriptive Statistics, Total Years of Principal Experience

	N	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis
						Std. Error	Std. Stat. Error
Total Years, Principal	138	0	30	8.36	6.638	1.272	.206
Valid N (listwise)	138						

In a different format, Table 5 also addresses the issue of the total number of years that each respondent has served as a school principal. As is evidenced in the table, there is considerable variability in this data. Though there are a large number of principals having served five years or less, there are also a considerable number of principals having served in the range of six to fifteen total years. This data is helpful to consumers of the research in that it provides an overview of the professional experience level of the population surveyed.

The school structures in which the principal respondents' are presently employed are demonstrated in Table 6. The data in this table indicates that the respondents were most frequently employed in a traditional K-5 school structure. The next most frequent structure comprised grades 6-8. Other school structures are represented with varying and limited frequency, but none of them exceeded 6.0 of the total number of research subjects. It is notable that there was at least one individual who elected to participate in the survey from each of the respective school structures, although some of the categories are very sparsely populated.

Table 5 – Matrix Display of Years of Principal Experience

Years of Experience	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Total Respondents	12	12	8	17	9	12	5	11	6	10	2	6	2	3	6	2	1	0	0	2
Percentage of Total Years	.08	.08	.05	.11	.06	.08	.03	.07	.04	.06	.01	.04	.01	.02	.04	.01	.01	.00	.00	.01

Years of Experience	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	Missing	Total
Total Respondents	4	0	1	2	0	1	3	0	0	1	0	0	0	0	0	17	155
Percentage of Total Years	.03	.00	.01	.01	.00	.01	.02	.00	.00	.01	.00	.00	.00	.00	.00	.11	1.000

Table 6 - Matrix Display of School Structure of Principals' Employment

	Pre-K	K-3	K-4	K-5	K-6	K-8	1-3	1-5	2-5
Subjects	1	1	6	61	9	6	1	6	1
Percentage of N	1.0	1.0	4.0	39.0	6.0	4.0	1.0	4.0	1.0

	3-5	4-5	5-6	5-8	6-8	7-8	Missing	Total
Subjects	2	3	1	5	29	6	17	155
Percentage of N	1.0	2.0	1.0	3.0	19.0	4.0	11.0	100.0

The frequency with which the respondents reported being evaluated varies somewhat, though it was very clear that most individuals were evaluated on an annual basis. Of the total participants, 122 responded to the question about evaluation frequency. Table 7 illustrates the respective categories for this question and the frequencies reported. Approximately 78 percent of the participants responded to the question. Of those cases, over 65 percent indicate that they are evaluated annually. The next most frequent evaluation category is “semi-annually.” While only one respondent indicates that he or she was “never” evaluated, almost 4 percent of the respondents state that they were only evaluated “sporadically.” Only about 3 percent of the respondents indicate that they were evaluated either every other year or on a three-year cycle. Clearly, a very large proportion of the respondents report being evaluated at least annually. In addition, almost 25 percent of the respondents indicate that they are evaluated more than once each year; with a very small percentage (1.3 percent) reported that they are evaluated monthly.

Table 7 –Matrix Display of Principal Evaluation Frequency

	Frequency	Percent	Valid Percent	Cumulative Percent
Monthly	2	1.3	1.6	1.6
Quarterly	9	5.8	7.4	9.0
Semi-annually	19	12.3	15.6	24.6
Annually	80	51.6	65.6	90.2
Every other year	4	2.6	3.3	93.4
Every three years	1	.6	.8	94.3
Sporadically	6	3.9	4.9	99.2
Never	1	.6	.8	100.0
Total	122	78.7	100.0	
Missing	33	21.3		
Total	155	100.0		

The titles of those individuals who were responsible for evaluating the principals fell into one of four categories. Table 8 displays the frequency with which the respondents selected from these position titles. On this question, 89 percent of the participants responded. Though this question on the survey instrument allowed for an open-ended response, none of the participants identified an alternate position title beside those offered. The data in Table 8 indicates that almost 77 percent of the survey participants who responded to the question express that they were evaluated by the superintendent of the school district in which they were employed. Approximately 23 percent of the respondents state that they were evaluated either by the deputy superintendent, an assistant superintendent, or a director (though only 3.6 percent report being evaluated by the deputy superintendent).

Table 8 – Matrix Display of Evaluator Title Frequency

	Frequency	Percent	Valid Percent	Cumulative Percent
Superintendent	106	68.4	76.8	76.8
Deputy superintendent	5	3.2	3.6	80.4
Assist. superintendent	15	9.7	10.9	91.3
Director	12	7.7	8.7	100.0
Total	138	89.0	100.0	
Missing	17	11.0		
Total	155	100.0		

Table 9 illustrates the number of schools that are found in each respondent's school district. The data is well-dispersed in this area; 27.5 percent of respondents indicate that they work in school districts consisting of more than 15 schools. About 20 percent of respondents also state that they worked in districts comprised 1-3, 4-6, and 7-9 schools, respectively, with districts of 10-15 schools at 12 percent.

Table 9 – Matrix Display of Schools in District

	Frequency	Percent	Valid Percent	Cumulative Percent
1-3	28	18.1	20.3	20.3
4-6	29	18.7	21.0	41.3
7-9	27	17.4	19.6	60.9
10-15	16	10.3	11.6	72.5
15 or more	38	24.5	27.5	100.0
Total	138	89.0	100.0	
Missing	17	11.0		
Total	155	100.0		

The final demographic attribute that is surveyed are the respective counties where each of the principals are employed. As the survey was only administered to K-8 principals working in Dupage, Will, and Lake Counties (IL), these choices were the only alternatives provided. Out of 155 survey participants, 137 responded to the question. Of these, 67 respondents (48.9 percent) identify themselves as working in DuPage County. Another 30 individuals (21.9 percent) identify themselves as working in Will County. Finally, 40 participants (29.2 percent) indicate that they are employed in Lake County.

Data Reduction

The primary emphases of the survey instrument are two matrices and four open-ended qualitative questions. One matrix asks respondents about the extent to which various evaluative methods have been used to formally evaluate their performance as a principal, and the other asks about the extent to which the respondents perceive principal evaluation accomplishing various objectives. Two of the open-ended questions ask about the respondents' beliefs pertaining to evaluation and principal performance, as well as evaluation and professional development. The other two open-ended questions ask about the respondents' beliefs regarding the impact of principal evaluation on pupil performance, and the impact of principal evaluation on professional development.

Before any statistical analyses could be performed on these variables, data reduction was needed. In the case of the matrices, the first step was to calculate total scale scores. For both matrices, the scales were calculated by adding together all of the items that comprised the scale, on a case-by-case basis. This resulted in a new variable reflecting the total scores for each case. As none of the items in the matrices were negatively worded, it was not necessary to reverse-score any of the items. On both

matrices, cases were excluded pairwise in creating the scales.

The first matrix was comprised of ten items (various evaluative methods), and the respondents were asked about the extent to which the methods were used to formally evaluate their performance as principals, if at all. The degree of normality for this scale was assessed through the use of descriptive statistics. The valid N on this matrix is 118, out of 130 total respondents. With 12 missing cases, the valid N is 90.7% of the total. The scale minimum is 10.00 and the maximum is 70.00. The mean for the scale is 35.22, with a standard deviation of 10.95. The skewness is .217, indicating scores clustering somewhat to the left at the low values. The kurtosis is -.204, indicating a relatively flat distribution with an increased number of cases at the extremes.

The second matrix was comprised of eleven items (various objectives of evaluation), and the respondents were asked about the extent to which they perceive principal evaluation accomplishing the various evaluation objectives, if at all. Descriptive statistics are again used to assess the degree of normality for this scale. The valid N on this matrix is 114, out of 130 total respondents. With 16 missing cases, the valid N is 87.7% of the total. The scale minimum is 11.00 and the maximum is 77.00. The scale mean is 47.79, with a standard deviation of 13.85. The skewness is -.362, indicating scores clustering to the right at the higher values. The kurtosis is -.289, indicating a relatively flat distribution with an increased number of cases at the extremes.

The next step was to check the reliability of the scales for both matrices. Reliability addresses the internal consistency of the scale, which can also be described as the degree to which the items on the scale successfully “hang together.” On each matrix, creating an index and calculating the Cronbach alpha coefficient for the scale helped to

determine reliability.

The variables on the first matrix reflect those methods in use for the purpose of principal evaluation. For this scale, all 10 items yield Cronbach alpha = .704. As the item-to-total criterion is set at .7, all of the individual items are retained and can reliably be used for analysis (Babbie, 1990).

The variables on the second matrix reflect the objectives for which evaluation is utilized. For this scale, all 11 items yield Cronbach alpha = .881. The item-to-total criterion is again set at .7, so all of the individual items are retained and can reliably be used for analysis.

In the case of the open-ended questions, the data reduction process required that the participants' responses be coded into appropriate response categories. To facilitate the reliability of the coding process, an additional rater was secured in order to code the responses independently from the researcher. On those circumstances when data were missing, cases were excluded pairwise.

Upon review of the open-ended responses, it became clear both to the researcher and the independent rater that two distinct coding categories existed for each question. The first two open-ended questions asked the participants to share their opinions about the effect (if any) of their experiences with evaluation, both on their beliefs about principal performance and also on their own professional development. For both of these questions, the responses distinctly indicate that the participants perceive evaluation as either ineffective or effective in impacting either principal performance or their own professional development. As such, the open-ended responses that were provided for these two questions are coded as either "Ineffective" or "Effective," utilizing the

following specific definitions:

Ineffective - *the response indicates no support for the position that evaluation is effective in impacting either principal performance or professional development; while response may indicate a small degree of support for the concept, the support is highly qualified or negated by subsequent comments.*

Effective – *the response indicates support for the position that evaluation is effective in impacting either principal performance or professional development; while response may be qualified, support for the position is identifiably positive (specific comments, evidence, or examples are provided).*

The second set of open-ended questions asked the participants to share their beliefs about the impact (if any) that the principal evaluation process has on pupil performance, and also on their own professional development. Once again, the responses on these two questions are distinctly polarized. The participants indicate their belief that the principal evaluation process either has limited/no positive impact or that it has considerable positive impact. Yet, none of the respondents volunteer that the principal evaluation process has a distinctly negative impact. As a result, the open-ended responses that are provided for these two questions are coded as either “No/Limited Positive Impact” or “Considerable Positive Impact,” utilizing the following specific definitions:

No/Limited Positive Impact – *the response indicates no support for the position that the principal evaluation process has impact on either pupil performance or professional development; while response may indicate a small degree of support for the concept, the support is highly qualified or negated by subsequent comments.*

Considerable Positive Impact – *the response indicates support for the position that the principal evaluation process has impact on either pupil performance or professional development; while response may be qualified, support for the position is identifiably positive (specific comments, evidence, or examples are provided).*

In order to maximize inter-rater reliability, both the researcher and the independent rater coded the responses for all of the open-ended questions separately. Afterward, the coded responses were compared and inconsistent codes were identified. The researcher and the independent rater then discussed each of these responses on an individual basis. Inconsistent coding was reconciled through discussion and consistent application of the category definitions, yielding the final coded responses.

Kappa Measure of Agreement is used to assess the consistency of inter-rater agreement. This statistic was derived for each of the four open-ended questions on which the responses were coded by the researcher and the independent rater, and the results are indicated as follows.

The first open-ended question asked the respondents to share their beliefs about evaluation and principal performance. On this question, the Kappa Measure of Agreement value is .90 with a significance of $p < .0005$. According to Peat (2001, p.228), a Kappa value of .5 indicates moderate agreement, above .7 indicates good agreement, and above .8 indicates very good agreement. As such, the Kappa value on this question reflects very good agreement between the raters.

The second open-ended question asked respondents about their beliefs regarding evaluation and their own professional development. In this case, the Kappa value is .919 with a significance of $p < .0005$. This Kappa value also represents very good agreement

between the two raters.

The third open-ended question asked the respondents about their beliefs regarding the impact of principal evaluation on pupil performance. The Kappa value for this question is .844 with a significance of $p < .0005$. Again, this Kappa value represents very good agreement between the raters.

The final open-ended question asked respondents to share their beliefs about the impact that principal evaluation has on their own professional development. In this situation, the Kappa value is .883 with a significance of $p < .0005$. This Kappa value also indicates very good inter-rater agreement.

Data Analysis

The initial objective of this research study is to determine the type and intensity of the methods that are in use for the purpose of principal evaluation. This objective is accomplished through the use of the “evaluation methods” matrix, with which respondents identify the degree (if at all) that various methods were used in the process of their performance evaluation. Table 10 indicates the intensity that the participants reported about the use of the various methods.

The most intensely used evaluative method is “Narrative evaluation by supervisor,” with a mean score of 5.53 and a standard deviation of 1.834. This method is followed by “Narrative self-evaluation,” with a mean score of 4.65 and a standard deviation of 2.255, and “Anecdotal evidence,” with a mean score of 4.41 and a standard deviation of 2.117.

By a sizable margin, the respondents indicate that the least intensely used evaluative method is “Peer supervision/review,” with a mean score of 1.64 and a standard

deviation of 1.418. This method is followed by “Survey data from teachers, parents, and students,” with a mean score of 2.63 and a standard deviation of 2.029, and “Portfolio/dossier,” with a mean score of 3.00 and a standard deviation of 2.298.

Table 10 – Matrix Display of Evaluation Method Intensity

	Mean	Std. Deviation	N
Narrative self-evaluation	4.65	2.255	118
Portfolio / dossier	3.00	2.298	118
Checklist / rating system	3.14	2.296	118
Supervisor observation	3.67	2.215	118
Narrative evaluation by supervisor	5.53	1.834	118
Data-based evaluation	3.53	2.183	118
Survey data from teachers, parents, students	2.63	2.029	118
Peer supervision / review	1.64	1.418	118
Anecdotal evidence	4.41	2.117	118
Perception feedback from stakeholders	3.03	2.172	118

The combined scale mean for the variables on the “evaluative methods” matrix is 35.23, with a standard deviation of 10.954. The grand mean for the scale is 3.52 with a standard deviation of 1.095.

Further analysis of the data from this matrix is possible through the use of an inter-item correlation matrix, investigated using Pearson product-moment correlation coefficient. The inter-item correlation matrix provides correlation statistics for each item on the scale in relationship to one another. This gives an indication as to the presence of possible association between each of the respective evaluative methods identified on the matrix. Note that all correlation analyses for this study were conducted by excluding

cases pairwise when data were missing.

Upon review of the inter-item correlation matrix, three noteworthy relationships surface. For interpretation of the significance of these correlation coefficients, Cohen (1988, p. 79) recommends the following guidelines, assuming that no violations of the assumptions of normality are found:

Small $r = .10 \text{ to } .29$

Medium $r = .30 \text{ to } .49$

Large $r = .50 \text{ to } 1.0$

Table 11 displays the correlation coefficients among all of the variables on the “evaluation methods” inter-item correlation matrix. On this matrix, several notable relationships are indicated at or above the .40 level. An example is found between the items “portfolio/dossier” and “data-based evaluation.” A medium, positive correlation exists between the two variables, $r = .409$, $n = 118$, $p < .05$. To determine how much variance the two variables share, one can calculate the coefficient of determination. By squaring the r value and multiplying by 100, this ‘percentage of variance’ can be determined. For these two variables, the use of this procedure results in a coefficient of determination of 16.7%. Separately, a medium, positive correlation is also identified between the variables “peer supervision/review” and “perception feedback from stakeholders,” $r = .426$, $n = 118$, $p < .05$. This yields a coefficient of determination of 18.1%.

The vast majority of the correlations displayed in Table 11 are positive. The correlations that are found to be negative are extremely small, and they range from $-.015$ to $-.042$.

Table 11 – “Evaluation Methods” Inter-Item Correlation Matrix

		Narr self-evaluation	Portfolio	Checklist	Supervisor observation	Narrative-supervisor	Data-based	Survey data	Peer supervision	Anecdotal	Perception feedback
Narr self-evaluation	Pearson Corr.	1	.224*	.183*	.068	.240**	.246**	-.042	.198*	.084	-.019
	Sig. (2-tail.)		.015	.047	.468	.009	.007	.654	.032	.368	.842
Portfolio	Pearson Corr.	.224*	1	.253**	.143	.195*	.409**	.209*	.210*	.263**	.135
	Sig. (2-tail.)	.015		.006	.123	.035	.000	.023	.023	.004	.144
Checklist	Pearson Corr.	.183*	.253**	1	.223*	.191*	.362**	.171	.192*	.299**	.074
	Sig. (2-tail.)	.047	.006		.015	.039	.000	.064	.037	.001	.423
Supervisor observation	Pearson Corr.	.068	.143	.223*	1	.139	.254**	.043	.212*	.238**	.081
	Sig. (2-tail.)	.468	.123	.015		.135	.006	.646	.021	.009	.386
Narrative-supervisor	Pearson Corr.	.240**	.195*	.191*	.139	1	.075	.020	-.010	.159	-.015
	Sig. (2-tail.)	.009	.035	.039	.135		.423	.834	.914	.085	.869
Data-based	Pearson Corr.	.246**	.409**	.362**	.254**	.075	1	.290**	.305**	.344**	.332**
	Sig. (2-tail.)	.007	.000	.000	.006	.423		.001	.001	.000	.000
Survey data	Pearson Corr.	-.042	.209*	.171	.043	.020	.290**	1	.407**	.117	.610**
	Sig. (2-tail.)	.654	.023	.064	.646	.834	.001		.000	.206	.000
Peer supervision	Pearson Corr.	.198*	.210*	.192*	.212*	-.010	.305**	.407**	1	.129	.426**
	Sig. (2-tail.)	.032	.023	.037	.021	.914	.001	.000		.162	.000
Anecdotal	Pearson Corr.	.084	.263**	.299**	.238**	.159	.344**	.117	.129	1	.120
	Sig. (2-tail.)	.368	.004	.001	.009	.085	.000	.206	.162		.197
Perception feedback	Pearson Corr.	-.019	.135	.074	.081	-.015	.332**	.610**	.426**	.120	1
	Sig. (2-tail.)	.842	.144	.423	.386	.869	.000	.000	.000	.197	

*. Correlation is significant at the 0.05 level (2-tail.).

**. Correlation is significant at the 0.01 level (2-tail.).

In addition, a strong, positive correlation is identified between the variables “survey data from teachers, parents, or students” and “perception feedback from stakeholders,” $r=.610$, $n=118$, $p<.05$. The resulting coefficient of determination of 37.2% indicates a strong relationship between these two variables, for those individuals who responded to the matrix questions.

Table 11 shows that correlations between .20 and .36 exist between several of the variables included on the inter-item correlation matrix. However, the data indicates that the strongest correlations on the matrix are among variables that are in little use. At the same time, some of the oft-used evaluation methods (such as Narrative-supervisor) show much smaller correlations with other variables. This data seems to indicate that some of the commonly used evaluation methods are likely used in isolation from other methods.

The second primary objective of this research study is to correlate the various evaluative methods with pupil performance. In order to accomplish this objective, School ISAT performance levels were calculated using the method described in Chapter 3 (Methodology). The schools were then assigned to one of three ISAT performance tiers, and principals were asked to respond to the survey that was available to their respective tier. This approach allows for the investigation of a potential association between the evaluation methods and school ISAT performance level (indicated by tier).

The correlation coefficients were calculated between each of the variables and the school ISAT performance level using Pearson product-moment correlation coefficient. While the direction of the relationship varies depending on the specific evaluation method, it is notable that there is no correlation greater than $\pm.11$ and none was significant. In other words, there is no evidence of an association between the use of any

of the evaluation methods and the school ISAT performance level of the school where the principal is employed. To the contrary, the correlation coefficients indicate that there is virtually no relationship at all between these variables. Ultimately, the data from the collected tiers did not matter in the objective of correlating the various evaluative methods with pupil performance, as demonstrated through school ISAT performance. This was true for all of the evaluation methods that were investigated.

An important clarification should be emphasized for consumers of this research. The respondents for this survey who indicated that they were serving in their first year as principal at the school where they were presently employed were excluded from this analysis, because ISAT scores are only publicly available the year after their administration. Though the decision to exclude participants limits N, it would not be statistically valid to correlate the methods by which the respondents were evaluated with student performance if there is no possibility that the performance measure could have been affected by the principal's evaluation.

The supplemental objective of this research study is to correlate the respondents' perceptions with issues pertaining to the impact and effectiveness of principal evaluation. First, the participants were asked about their beliefs regarding the effectiveness of evaluation on principal performance and professional development. Second, the respondents were questioned about their beliefs pertaining to the impact of principal evaluation on pupil performance and professional development. These questions were asked in an open-ended manner, allowing the respondents to share their beliefs in a manner as unique and rich as they wished. As previously described, the responses were then coded for quantitative analysis. Pearson product-moment correlation coefficient was

used to investigate the relationships, if any, between the respective evaluation methods and the respondents' beliefs. Missing cases were excluded pairwise for analysis.

Table 12 displays the results of the analysis. For each of the identified evaluative methods, the table indicates the respective correlation coefficients for the coded qualitative responses. In the table, the variable titled "Effective: prin. perf." refers to principals' beliefs about the effectiveness of principal evaluation on principal performance. The variable "Effective: prof. dev." refers to principals' beliefs about the effectiveness of principal evaluation on their own professional development. The variable titled "Impact on pup. perf." refers to principals' beliefs about the impact of the principal evaluation process on pupil performance. Finally, the variable "Impact on prof. dev." refers to principals' beliefs about the impact of the principal evaluation process on their own professional development.

Statistical significance is found between several variables. This includes a small, positive correlation between the variables "Narrative self-evaluation" and "Impact on pup. perf.," $r=.245, n=103, p<.05$, and between "Narrative self-evaluation" and "Impact on prof. dev.," $r=.236, n=104, p<.05$. A small, positive correlation is also indicated between "Portfolio/dossier" and "Effective: prin. perf.," $r=.203, n=102, p<.05$.

A small, positive correlation is also present between the variables "Data-based evaluation" and "Effective: prof. dev.," $r=.211, n=100, p<.05$, and between "Data-based evaluation" and "Impact on prof. dev.," $r=.283, n=103, p<.01$. Another small, positive correlation is indicated between the variables "Peer supervision/review" and "Effective: prin. perf.," $r=.214, n=101, p<.05$, and between "Peer supervision/review" and "Impact on prof. dev.," $r=.219, n=103, p<.05$.

Table 12 – Correlation of Eval. Methods and Qual. Principal Perception Data

		Effective: prin. perf.	Effective: prof. dev.	Impact on pup. perf.	Impact on prof. dev.
Narrative self-evaluation	Pearson Corr.	.132	.335**	.245*	.236*
	Sig. (2-tailed)	.186	.001	.013	.016
	N	102	101	103	104
Portfolio / dossier	Pearson Corr.	.203*	.164	.120	.100
	Sig. (2-tailed)	.041	.101	.229	.312
	N	102	101	103	104
Checklist / rating system	Pearson Corr.	.162	.046	.100	.147
	Sig. (2-tailed)	.107	.650	.318	.141
	N	100	99	101	102
Supervisor observation	Pearson Corr.	.023	.060	.168	.083
	Sig. (2-tailed)	.817	.556	.091	.406
	N	101	100	102	103
Narrative evaluation by supervisor	Pearson Corr.	.140	.128	.138	.036
	Sig. (2-tailed)	.160	.204	.165	.715
	N	102	101	103	104
Data-based evaluation	Pearson Corr.	.114	.211*	.345**	.283**
	Sig. (2-tailed)	.256	.035	.000	.004
	N	101	100	102	103
Survey data	Pearson Corr.	.091	.123	.115	.171
	Sig. (2-tailed)	.367	.221	.248	.084
	N	101	100	102	103
Peer supervision / review	Pearson Corr.	.214*	.153	.156	.219*
	Sig. (2-tailed)	.032	.129	.118	.026
	N	101	100	102	103
Anecdotal evidence	Pearson Corr.	.106	.148	.088	.009
	Sig. (2-tailed)	.291	.143	.380	.925
	N	101	100	102	103
Perception feedback from stakeholders	Pearson Corr.	.024	.127	.053	.121
	Sig. (2-tailed)	.813	.209	.600	.224
	N	101	100	102	103

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Two correlations of medium statistical significance are also displayed in Table 12. A medium, positive correlation is identified between the variables “Narrative self-evaluation” and “Effective: prof. dev.,” $r=.335, n=101, p<.01$. This resulting coefficient of determination is 11.2%. Another medium, positive correlation is indicated between the variables “Data-based evaluation” and “Impact on pup. perf.,” $r=.345, n=102, p<.01$. This yields a coefficient of determination of 11.9%.

For each of the given dependent variables, the significant univariate correlations were further analyzed to determine if a useful complement of evaluation methods exists from the point of view of those being evaluated. The inter-correlations of the variables that had a positive correlation to the dependent variables were checked using Cronbach alpha. Combining those variables that had significant correlations resulted in the creation of new variables. Finally, the new variables were subsequently correlated to the dependent variables using Pearson product-moment correlation coefficient. The dependent variables include “Effective for principal performance” (EPP), “Effective for professional development” (EPD), “Impact on pupil performance” (IPP), and “Impact on professional development” (IPD). By combining the original variables, the resulting correlations become stronger.

“Complement 1” is a new variable consisting of the original variables “Portfolio/dossier” and “Peer supervision/review,” while EPP is a dichotomous coded dependent variable. A small, positive correlation is identified between “Complement 1” and “EPP,” $r=.252, n=102, p<.05$. The new variable “Complement 2” consists of the original variables “Narrative self-evaluation” and “Data-based evaluation,” while “EPD” is a dichotomous coded dependent variable. A medium, positive correlation is identified

between “Complement 2” and “EPD,” $r=.340, n=100, p<.01$.

“Complement 3” is another new variable consisting of the original variables “Narrative self-evaluation” and “Data-based evaluation,” while “IPP” is a dichotomous coded dependent variable. Another medium, positive correlation is identified between “Complement 3” and “IPP,” $r=.372, n=102, p<.01$. Finally, “Complement 4” consists of the original variables “Narrative self-evaluation,” “Data-based evaluation,” and “Peer supervision/review,” while “IPD” is again a dichotomous coded dependent variable. A medium, positive correlation is also identified between “Complement 4” and “IPD,” $r=.335, n=102, p<.01$. For each dependent variable, correlation is positively increased when combining the significant variables into a complement of evaluation methods.

It should be noted that no negative correlation coefficients exist between any of the variables in Table 12, indicating with consistency that when one variable increases, so too does the other. It is also important to underscore that the majority of the correlation coefficients show no or extremely low levels of significance. Most of the evaluative methods, including “checklist/rating system,” “supervisor observation,” “narrative evaluation by supervisor,” “survey data,” “anecdotal evidence,” and “perception feedback from stakeholders,” display virtually no relationship to any of the qualitative variables.

The unanticipated result of the qualitative data being dichotomous also allows for the qualitative data to be analyzed by school ISAT performance level, using the Chi Square test for independence. The Chi Square test for independence is used to explore the relationship between two categorical variables; in this case, the variables are the school performance tiers and the respective open-ended questions.

When conducted however, the Chi Square tests for independence did not yield

any values of significance. As the assumptions concerning minimum expected cell frequencies were not violated (each cell contained counts over 5), the values for each of the respective tests indicate that there is no association between the responses on the open-ended questions and school ISAT performance levels. In fact, the response variations were generally minimal from one item to the next.

In order to seek convergence with the qualitative perception data, analysis of the “perceived evaluation objectives” matrix data was also necessary. This matrix asked that respondents identify the degree (if at all) that principal evaluation was perceived to accomplish various objectives of evaluation. Table 13 indicates the intensity that the participants reported about their perception of the various objectives.

Table 13 – Matrix Display of Perceived Evaluation Objectives

	Mean	Std. Deviation	N
Satisfy district accountability requirements	5.75	1.583	114
Increase standardized assessment scores	3.65	1.932	114
Provide principals with professional growth	4.30	1.904	114
Document sub-standard principal performance	4.67	1.773	114
Identify the needs for principal professional dev.	4.38	1.879	114
Provide incentive for performance improvement	3.67	1.922	114
Reward exemplary principal performance	3.55	2.009	114
Ensure adherence to policies and procedures	4.99	1.706	114
Support the maintenance of instructional program	4.49	1.845	114
Improve pupil achievement	4.18	1.929	114
Foster positive school climate	4.18	1.965	114

The most intensely perceived objective of evaluation is “Satisfy district

accountability requirements,” with a mean score of 5.75 and a standard deviation of 1.583. This evaluation objective is followed by “Ensure adherence to policies and procedures,” with a mean score of 4.99 and a standard deviation of 1.706, and “Document sub-standard principal performance,” with a mean score of 4.67 and a standard deviation of 1.773.

By a narrow margin, the least intensely perceived objective of evaluation is “Reward exemplary principal performance,” with a mean score of 3.55 and a standard deviation of 2.009. This method is followed by “Increase standardized assessment scores,” with a mean score of 3.65 and a standard deviation of 1.932, and “Provide incentive for performance improvement,” with a mean score of 3.67 and a standard deviation of 1.922.

The combined scale mean for the variables on the “perceived evaluation objectives” matrix is 47.80 with a standard deviation of 13.860. As the scale contains 11 items, this converts to a combined mean score of 4.35 with a standard deviation of 1.260.

As before, further analysis of the data from this matrix is possible through the use of an inter-item correlation matrix, investigated using Pearson product-moment correlation coefficient. The inter-item correlation matrix provides correlation statistics for each item on the scale in relationship to one another. This gives an indication as to the presence of possible association between each of the respective evaluative methods identified on the matrix. Note that all correlation analyses for this study were conducted excluding cases pairwise when data was missing.

While several of the variables on the “perceived evaluation objectives” matrix correlate with medium-to-strong significance (and none of them were negatively

correlated), three of the correlation coefficients are particularly noteworthy. On the inter-item correlation matrix, a considerable association is found between the items “Foster positive school climate” and “Support the maintenance of the instructional program.” A strong, positive correlation exists between the two variables, $r=.674$, $n=114$, $p<.05$. This results in a coefficient of determination of 45.4%. Separately, a strong, positive correlation is also identified between the variables “Improve pupil achievement” and “Support the maintenance of the instructional program,” $r=.687$, $n=114$, $p<.05$. This yields a coefficient of determination of 47.2%.

Finally, a markedly strong, positive correlation is identified between the variables “Increase standardized assessment scores” and “Improve pupil achievement,” $r=.814$, $n=114$, $p<.05$. The resulting coefficient of determination of 66.3% indicates a particularly strong relationship between these two variables, for those individuals who responded to the matrix questions.

CHAPTER FIVE DISCUSSION

This chapter will now discuss the research questions through the lens of the study findings. Prior to that discussion, it will briefly review the purpose, population, research design, validity of the research, and limitations of the study. A discussion of the conclusions will follow as the research objectives are addressed. Finally, recommendations for future research will be provided.

Review of the Study

Purpose

The purpose of this study was to investigate the topic of the evaluation of school principals and the role it plays in fostering student achievement. The overarching research question for the study was the following: “Is there a relationship between the type and intensity of principal evaluation and pupil performance in DuPage, Will, and Lake County, Illinois public schools at the K-8 level?”

This question contains two primary components, both of which are central to understanding the larger issue. The initial objective was to determine the type and intensity of the methods that are in use for the purpose of principal evaluation in DuPage, Will, and Lake County, Illinois public schools at the K-8 level. The second objective was to correlate the identified evaluative methods with pupil performance, which are based upon Illinois Standards Achievement Test (ISAT) school scores for the 2007-2008 year.

To supplement these primary research goals, principals' perceptions were also investigated pertaining to the relationships between principal evaluation and pupil performance, principal performance, and professional development. Lastly, the study explored the correlation between principals' perceived objectives of evaluation and the evaluative methods in predominant use.

Population

The participants for this study were public school K-8 principals currently practicing in DuPage, Will, and Lake Counties, in the State of Illinois. These mostly affluent counties outside of Chicago have 480 public schools serving students in grades ranging from K-8 in a number of different program configurations, with all schools serving at least one grade level between 3-8. However, as public school principals, all of the target participants are required by the Illinois School Code to receive evaluation based upon their job performance. Because of their mutual employment in DuPage, Will, and Lake Counties, all of the principals work under the purview of their respective Regional Offices of Education. Yet despite the local peculiarities and nuances of each county, the Illinois School Code subjects all public school principals to the same applicable evaluation requirements (105 ILCS 5/24 A-1).

Four DuPage County principals were excluded from the study in advance. Two of the exclusions consisted of the researcher himself and his school district colleague. The other two exclusions were principals in the community in which the researcher resides. The initial number of respondents for the research study totaled 155. However, a number of respondents were also excluded from analysis as they were serving as in their capacity as first-year principals at the school where they were presently

employed. As a result, the final N=130.

Research Design

The research design that was employed was the “Concurrent Nested Strategy.” This mixed methods research approach “can be identified by its use of one data collection phase, during which both quantitative and qualitative data are collected simultaneously” (Creswell, 2003, p. 218). In this approach, a predominant method exists (quantitative) that guides the research, while a secondary method (qualitative) is “nested” within the larger study. The study investigated the relationship between student achievement and the type and intensity of principal evaluation as the predominant component, while the principals’ perceptions about the relationship of principal evaluation on student performance, principal performance, and professional development served as the “nested” component. Qualitative data was coded for the purpose of data analysis, and findings from both methods were interpreted concurrently in order to seek convergence among the results.

This survey research study was cross-sectional, and was self-administered in an online format. A random sample was not drawn from which to collect data. Instead, all individuals in the target population were asked to participate. The survey was both confidential and anonymous in nature. The online format was selected primarily because of the cost-effectiveness in administration for a large group of potential respondents. There are additional strengths, as well as weaknesses, of this format. Strengths include the ability of the researcher to receive data in a time-sensitive manner, the ability to target recipient groups directly, and the convenience of administration. Weaknesses include the difficulty of securing respondent e-mail addresses, the risk that respondents might

interpret the survey as “junk mail,” and the possibility that respondents may not be technologically savvy enough to use the survey or distrust the link provided to connect to the online survey driver.

Validity of the Research

In every research study, it is important to take deliberate steps to provide assurance that the instrument being used to collect the data is properly suited to the needs of the study. This idea, known as ‘content validity,’ refers to “the degree to which a measure covers the range of meanings included within the concept” (Babbie, 1990, 134). Content validity must be “determined by expert judgment. There is no formula by which it can be computed and there is no way to express it quantitatively” (Gay, 1992, 157). Though it is never possible to ensure complete content validity between a survey instrument and the survey constructs, proactive steps should still be taken to increase confidence in the level to which a study possesses content validity.

In this study, two critical tasks were completed in order to maximize content validity. First, a pre-test was conducted. Babbie (1990) explains that pre-testing commonly refers to the practice of administering “a draft questionnaire to a group of subjects” (p. 220). For this study, pre-testing was accomplished by sharing the survey with the members of the researcher’s dissertation committee. Their feedback was then incorporated into revisions of the survey instrument, as appropriate.

The second step that was employed to maximize content validity was a survey pilot. Babbie (1990) states that “unlike a pre-test, a pilot study should be directed to a representative sample of the target population” (p. 226). In this study, a representative sample of the population was identified who were not targeted as participants for the full

study. Consisting of 10 practicing K-8 principals, the members of this group were primarily employed in Cook County, IL. Babbie (1990) recommends that “the pilot study questionnaire should contain all the intended questions in the wording, format, and sequence that pre-testing has indicated are best for the final survey” (p. 226). In the pilot for this study, the members of the pilot group were all administered the final version of the survey using the same method that was utilized with the full study. The results of the pilot were then reviewed to insure that participants were responding to the questions in a manner that reflected the survey constructs. As the pilot results showed consistency between the participants’ responses and the constructs of the survey, the researcher expresses confidence that overall content validity was established.

Additionally, it is important to acknowledge any potential threats to external validity. Creswell (2003) explains that “external validity threats arise when experimenters draw incorrect inferences from the sample data to other persons, other settings, and past or future situations” (p. 171). There is often great temptation on the part of investigators to extrapolate meaning from research that is not generalizable to a different population. While this is often done with the best of intentions, the decision to generalize results without legitimate basis is inappropriate at best and potentially quite damaging at worst. In this study, there are three potential threats to external validity that must be addressed.

The first of these is a kind of reactive arrangement. While reactive arrangements can occur in several different ways, all of them deal with the manner in which the research subjects are affected by the specific nature of the research. In the case of this study, the greatest concern is the type of reactive arrangement that “results from the

subjects' knowledge that they are involved in an experiment or their feeling that they are in some way receiving 'special' attention" (Gay, 1992, p. 313). The respondents for this study are all K-8 public school principals who are currently practicing in Will, Lake, or DuPage Counties in Illinois. If, for example, this unique group of respondents were to feel that their inclusion as part of the research group was affording them special attention, it would subsequently represent a threat to the external validity of the study. Though there is no viable way to determine the presence of a reactive arrangement, it is important to state that no evidence has surfaced indicating this occurrence.

Another potential threat stems from the problem of 'selection-treatment interaction'. Selection-treatment interaction threats occur when samples are not randomly chosen from the larger population, and they can have dire effect if not considered when drawing research conclusions. Gay (1992) explains that "the very fact that subjects are not randomly selected from a population severely limits the researcher's ability to generalize since representativeness of the sample is in question" (p. 309). This being the case, it is important that consumers of this research are mindful that the respondents were deliberately culled from a unique pool of potential participants by design. While outside inferences can only be made in a tenuous fashion, the researcher believes that some practical generalizability does exist to other similar populations in similar settings.

The last possible threats to external validity in this study are 'experimenter effects.' Experimenter effects occur as the result of the researcher's behavior, although often subconsciously and unobtrusively. There are "a number of ways in which the experimenter may unintentionally affect execution of study procedures, the behavior of

subjects, or the assessment of their behavior, and hence results” (Gay, 1992, p. 311).

The aggregate influence that some of these actions may have upon the outcome of the study can be substantial. While efforts were made to minimize experimenter effects in this study by limiting the researcher’s interaction with subjects, one must accept the unknown possibility of their occurrence. However, it is important to underscore that the researcher is not aware of the presence or identification of any experimenter effects throughout the course of this research study.

While not a primary concern for survey research, it is also important to acknowledge the role that threats to internal validity can play in compromising research. Creswell (2003) defines internal validity threats as “procedures, treatments, or experiences of the participants that threaten the researcher’s ability to draw correct inferences from the data” (p. 171). In 1971, Campbell and Stanley identified eight major threats to internal validity. These threats include History, Maturation, Testing, Instrumentation, Statistical Regression, Differential Selection of Subjects, Mortality, and Selection-Maturation Interaction.

As this study did not attempt to employ traditional experimental design, the typical threats to internal validity were not a factor in assessing the overall validity of the study. However, concern is merited in those occurrences where survey participants only provided partial data. This phenomenon required that the researcher conduct analysis with less information than if full data was available for each of the respondents. While this was not the norm within the study, it did happen with a limited degree of frequency on the open-ended survey questions. With the goal of disclosure and accuracy, this phenomenon should be openly shared.

Limitations of the Study

As with all research models, there are limitations inherent in the use of the simultaneous mixed-methods research model. These limitations are unavoidable, but must still be acknowledged as they do impact the validity of the research.

Gay (1992) states that “a limitation is some aspect of the study that the researcher knows may negatively affect the results or generalizability of the results but over which he or she probably has no control” (p. 108). In the case of simultaneous mixed-methods research, three predominant types of limitation exist.

The first of these limitations originates directly from the research design. In simultaneous mixed-methods research, the methods are not employed on an equal basis. Creswell (2003) emphasizes that “because the two methods are unequal in their priority, the (simultaneous mixed-methods) approach also results in unequal evidence within a study, which may be a disadvantage when interpreting the final results” (p. 219). While this limitation may be inherent in the nature of the research model, it must be communicated clearly by the researcher or there is a risk that readers may interpret the results inappropriately.

Another limitation of this model is found in the qualitative coding process. Qualitative coding is the process that is used to encode and convert qualitative data into quantitative data so that it can be further analyzed. Though qualitative data can be encoded effectively, “no ways exist of perfectly replicating the researcher’s analytical thought process” (Patton, 2002, p. 433). This challenge places a limitation on the degree to which the researcher can assure the accuracy of the encoded data. In this study, it has already been demonstrated through Kappa Measure of Agreement that the coding process

was conducted with a high degree of agreement, however.

Finally, limitations can stem from the fact that simultaneous mixed-methods studies yield two different kinds of data. This can make analysis difficult for the researcher if there is an occasion when the quantitative and qualitative data findings do not support each other. When this happens, Creswell (2003) admits that “there is little advice to be found for how a researcher should resolve discrepancies that occur between the two types of data” (p. 218). In this study, it is fortunate that the findings displayed a large degree of convergence, thereby minimizing this limitation greatly.

The limitations of the simultaneous mixed-methods model are not pervasive, nor do they call the research validity of the results from this study into question in any way. Still, the limitations are important to acknowledge and explain, as they do place some specific parameters on the way in which the results can be interpreted.

Conclusions

Based upon the data analysis, a number of conclusions can be drawn from the research. Several of the research conclusions from this study are based upon correlation data, which is primarily used to determine the presence of relationship between given variables. While this study relates a number of cases in which relationships can be found between variables, it is important to emphasize that this does not indicate causality, nor should causality be inferred.

The following conclusions are the result of the research that was conducted on the topic of principal evaluation:

1. Principals are currently being evaluated in DuPage, Will, and Lake Counties using a variety of evaluation methods, which are being conducted with widely

varying intensity and frequency. The method “Narrative evaluation by supervisor” was employed most often (Mean=5.53). The method “Peer supervision/review” was used with the least intensity (Mean=1.64). Low scores on many of the other methods indicate a prevalence of one-dimensional leadership evaluation systems.

2. A paired occurrence exists between the evaluation methods “survey data from teachers, parents, or students” and “Perception feedback from stakeholders” (strong, positive correlation, $r=.610$, $n=118$, $p<.05$). This indicates an evaluation “package” of methods that might effectively be used together. However, the actual use of these methods in practice was reported with low frequency and intensity.
3. Correlation coefficients indicate that there is virtually no relationship at all between any of the identified evaluation methods and the school ISAT performance level. As a result, the selective use of one or another method cannot be expected to impact pupil performance on the ISAT assessment.
4. There is a limited association between the “narrative self-evaluation” method and respondents’ perception that evaluation is effective for the purpose of professional development (medium, positive correlation, $r=.335$, $n=101$, $p<.01$).
5. There is a limited association between the “data-based evaluation” method and respondents’ perception that the principal evaluation process has an impact on pupil performance (medium, positive correlation, $r=.345$, $n=102$, $p<.01$).
6. When using principal evaluation for specific purposes, responses indicate that evaluation “complements” exist that seem to be more effective than any of the

methods alone. They include:

- Improving principal performance - “Portfolio/dossier” and “Peer supervision/ review”
 - Principal professional development - “Narrative self-evaluation” and “Data-based evaluation”
7. By a considerable margin (.76), principals most intensely perceive that the objective of evaluation is to “Satisfy district accountability requirements,” followed by “Ensure adherence to policies and procedures,” and “Document sub-standard principal performance.” These data indicate that principals generally have a custodial perception about the principal evaluation process that limits its potential to impact professional growth and development.
8. Inter-item correlation matrix indicates three strong, positive correlations of note:
- “Foster positive school climate” and “Support the maintenance of the school program” ($r=.674, n=114, p<.05$)
 - “Improve pupil achievement” and “Support the maintenance of the school program” ($r=.687, n=114, p<.05$)
 - “Increase standardized assessment scores” and “Improve pupil achievement” ($r=.814, n=114, p<.05$)

Clearly, it is important to review each of these conclusions in more detail. The first conclusion from this research is that principals are presently being evaluated using a number of different evaluation methods, and these evaluations are being conducted with widely varying intensity and frequency. While almost two-thirds of the principals indicated that they were evaluated on an annual basis, most of the remainder stated that

they were evaluated more often than that. This is promising information because it reflects that supervisors generally place importance on the evaluation process. Further, less than 6 percent of the respondents stated that they were not evaluated on any consistent basis.

The methods that are used to conduct the evaluation process in DuPage, Will, and Lake Counties vary substantially. On the survey instrument, the respondents were presented with ten different principal evaluation methods and asked to identify which of them had been used to evaluate their performance, and with what intensity. Though the rate of use ranged considerably from one method to another, the responses from the participants indicated that all of the methods were used in some degree to evaluate their performance as principals. The method with the lowest mean score was “Peer supervision/review (1.64),” indicating that this method was used very infrequently and with low intensity. “Narrative evaluation by supervisor (5.53)” was the method with the highest mean score, followed by “Narrative self-evaluation (4.65)” and “anecdotal evidence (4.41)”. The remaining evaluation methods surveyed resulted in mean scores ranging from 2.63 to 3.67. Considering that the highest possible mean for each of these variables is 7.0, the “Narrative by supervisor” method score was particularly strong. This was the evaluative method that was clearly used with the greatest intensity, and was the most predominantly used for the principals who were surveyed. Unfortunately, the lower mean scores for most of the other evaluation methods included in the survey indicated a prevalence of one-dimensional leadership evaluation systems being used throughout those counties that were studied.

The evaluation frequency/intensity data were also analyzed using Cronbach alpha

to determine if a common evaluation ‘package’ seems to exist among any of the methods. The goal was to determine if a relationship is present between any of the variables that indicate a paired occurrence. A strong, positive correlation was identified between “Survey data from teachers, parents, or students” and “Perception feedback from stakeholders,” $r=.610$, $n=118$, $p<.05$. This is logical as both of these evaluative methods are predicated on the use of perception data from constituents. However, survey respondents reported both of these methods with low frequency and intensity, so the significance of this correlation should be considered carefully. While the correlation coefficient may be statistically significant, it is of somewhat limited practical significance since neither of these evaluation methods is in frequent use.

The next conclusion from this research relates to the issue of whether a relationship exists between the type/intensity of principal evaluation and pupil performance, as evidenced by “School ISAT performance level.” Pearson product moment correlation was calculated between the school ISAT performance level variable and each of the respective evaluative method variables. The correlation coefficients indicated that there was virtually no relationship at all between any of the identified evaluation methods and the school ISAT performance level. All of the correlation coefficients fell in the range from $-.051$ to $.113$. As a result, the conclusion can be made that none of the evaluation methods commonly used in DuPage, Will, and Lake Counties were associated with school ISAT performance level, and the selective use of one or another method cannot be expected to result in any relationship with pupil performance on the ISAT assessment.

Several conclusions from this research relate to the respondents’ perceptions

about the effectiveness and the impact of the principal evaluation process. The survey asked four open-ended questions about these topics, allowing the respondents to share their perceptions in their own words. It is important to note that the number of respondents on these open-ended questions was smaller than on other questions in the survey. However, this can be interpreted that the respondents perceived that the other portions of the survey provided them with a “voice” to fully express their opinions and beliefs about the topics.

During the data reduction phase, the responses from each of these qualitative questions were coded to allow for quantitative analysis. It was clear to both the researcher and the independent rater that two distinct coding categories existed for each qualitative question, as described in Chapter 4. One of the strengths of the qualitative research approach is that research subjects are given the opportunity to respond to the topic in a highly personal manner. However, they apparently elected not to do so, as their responses clearly fell into one or the other of two distinct categories. This occurrence reinforces the decision that was made to code these data, since this approach has maintained the essential integrity of the data’s meaning.

Employing Pearson product-moment correlation, correlation coefficients were calculated to determine whether a relationship exists between the respondents’ perceptions about the effectiveness and impact of the evaluation process and the identified evaluation methods. While all of the correlation coefficients were positive, the majority of the relationships were not significant. However, several small correlations were evident, and two correlations of medium statistical significance occurred.

The first of these medium, positive correlations was identified between the

variables “Narrative self-evaluation” and “Effective for professional development,” $r=.335, n=101, p<.01$. This correlation coefficient indicated the presence of a limited association between the use of the narrative self-evaluation method and the respondents’ perception that evaluation is effective for the purpose of professional development. This data is compelling, mainly because “Narrative self-evaluation” is used frequently, and is one of the only identified evaluation methods that is self-reflective in nature.

Another medium, positive correlation was indicated between the variables “Data-based evaluation” and “Impact on pupil performance,” $r=.345, n=102, p<.01$. This correlation coefficient indicated the presence of a limited association between the use of data-based evaluation and the respondents’ perception that the principal evaluation process has impact on pupil performance. In an age of student achievement testing, school accountability, and data collection, it is not surprising that the participants’ responses would indicate an association between these variables. It is safe to state that in both of the aforementioned cases, the correlations coefficients display a relationship between the variables that is noteworthy. This is particularly notable since the variables reflect qualitative perception data.

The value of this qualitative data was expanded further after the evaluation methods that displayed significant correlation coefficients were identified and isolated. Once this was achieved, it became possible to derive a complement of evaluation methods that are most relevant for a given purpose from the point of view of those being evaluated. For example, by isolating the evaluation methods “Portfolio/dossier” and “Peer supervision/review” from the other non-significant variables in the qualitative perception data, it was evident that the respondents perceived the combined use of these

two evaluation methods as more effective in improving principal performance than any of the methods used independently. When employing the evaluation process for this specific purpose, the tandem use of these two evaluation methods thereby resulted in a stronger evaluation complement than the evaluative items used alone. It is noteworthy that principals perceived the use of the “Portfolio/dossier” and “Peer supervision/review” evaluation methods to be most effective in improving principal performance, yet these two methods are not in common use. This point should be carefully considered by those individuals involved in conducting and developing principal evaluations. Likewise, the same approach also indicates that the respondents perceived the combined use of the evaluation methods “Narrative self-evaluation” and “Data-based evaluation” to be the most effective when using evaluation for the purpose of professional development compared to any of the methods used alone. Yet, these two evaluation methods are more commonly used than many of the other methods.

When using principal evaluation toward the goal of impacting pupil performance, the respondents indicated their perception that the evaluation methods “Narrative self-evaluation” and “Data-based evaluation” again comprised the strongest complement of evaluation methods. However, the respondents perceived the use of the evaluation methods “Narrative self-evaluation,” “Data-based evaluation,” and “Peer supervision/review” to result in the most effective evaluation complement when the objective is to impact principals’ professional development through evaluation. Administrators who are responsible for conducting principal evaluations may wish to consider the application of these findings to other related areas, such as principal perception, principal motivation, and the evaluation effectiveness.

In order to seek convergence between the qualitative and quantitative results, the respondents were also asked to identify the degree (if any) that principal evaluation was perceived to accomplish various objectives of evaluation. Eleven different evaluation objectives were presented in matrix format, allowing the respondents to indicate the intensity of their perceptions about each variable. On a scale from 1 to 7, the combined score for all of the items was 4.35. This suggests that the respondents generally supported the idea that the identified objectives represented legitimate purposes for evaluation. Yet, by a substantial margin (.76), the most intensely perceived objective of evaluation was “Satisfy district accountability requirements” with a mean score of 5.75. A distant second was the variable “Ensure adherence to policies and procedures” with a mean score of 4.99, followed by “Document sub-standard principal performance” with a mean score of 4.67.

The data suggest that these custodial perceptions about evaluation have the effect of limiting the potential of using principal evaluation for positive benefit. The variables with the highest mean scores tended to reflect a negative perception of the purposes of evaluation by those who are being evaluated. Though the respondents did perceive value in objectives such as “Provide principals with professional growth” and “Foster positive school climate,” these variables were scored much lower than those that focused on the accountability component of the evaluation process. While it is difficult to ascertain the reasons for these perceptions, they certainly impede the overall possibility that evaluation might be increasingly used for the purposes of professional growth, improved principal performance, and pupil achievement.

As a result of these findings, it is important to underscore the need for

superintendents and other individuals who are responsible for conducting principal evaluations to take heed of the principals' perceptions about the effectiveness of the evaluation process. Evaluators must accept that the evaluation process is not currently being perceived as successful in meeting the objectives of increasing principal performance, improving pupil achievement, or aiding in principal professional development. It is imperative that evaluators begin to look more closely at the evaluation systems that are presently in place, as well as question whether these systems are effective at accomplishing the desired goals. Likewise, it is vital that superintendents and evaluators consider the review and replacement of those evaluation systems that are not being found effective. Under the circumstances that a need for a new evaluation system is identified, it is equally important that the principals who will be affected by a new system be brought to the table as part of the evaluation development process.

Another conclusion about the perceived objectives of evaluation was identified in the data from the inter-item correlation matrix. A strong, positive correlation was found between the variables "Foster positive school climate" and "Support the maintenance of the instructional program," $r=.674, n=114, p<.05$. Another strong, positive correlation was identified between the variables "Improve pupil achievement" and "Support the maintenance of the school program," $r=.687, n=114, p<.05$. Lastly, a very strong, positive correlation was found between the variables "Increase standardized assessment scores" and "Improve pupil achievement," $r=.814, n=114, p<.05$.

These data suggest that it is clear to principals which objectives are mutually necessary to accomplish their desired aims as school leaders. It is apparent that the school principals surveyed are very capable of identifying the goals that are most critical

to the genuine success of the students and their school. In addition, they demonstrate logical perspectives about the complementary nature of some of these critical objectives.

Clearly, some disconnect is apparent between the way in which principal evaluation is perceived and the way in which it is utilized. What may be needed moving forward is both a shift in the manner in which principal evaluation is conducted by supervisors, as well as increased openness on the part of principals to accept the potential for a well-managed evaluation process to impart positive benefit. By changing the current way of thinking, evaluation systems have the beneficial potential to make considerable impact on the quality of education provided for our students, the manner in which our schools are monitored and supervised, and the professional growth of our school leaders.

Recommendations for Future Research

The primary purpose of this study was to investigate the topic of the evaluation of school principals and the role it plays in fostering student achievement. While much of the information garnered from the results seems to affirm “common sense,” it is intriguing that some of the results also yielded unanticipated outcomes. It is clear that much more research is needed.

Although efforts were made to avoid them, this study had shortcomings of which future researchers should be aware. First, the population for the study was limited in scope. For both practical and financial reasons, the study only included research subjects from the DuPage, Will, and Lake County (IL) areas. If this study were replicated, a random sample from a much larger population would allow for greater generalizability of the results. This would generate valuable information about principals’ perceptions and

their experiences with evaluation on a much larger scale.

Another shortcoming of this study was the timing with which it was administered. The survey was disseminated to the research subjects during the month of April, which is a very busy time of year for educators. An increased number of participants might have resulted if the study was administered during one of the summer or autumn months. However, as the research was conducted as part of a doctoral program, the researcher did not have the practical ability to collect the data at a different time of year. Future researchers replicating this study might find that administering the study at another time of year could result in a higher response rate, thereby adding validity to the research findings.

Questions also remain as to several aspects of the relationship between principal evaluation and pupil achievement. One of these unresolved questions is whether increased frequency of principal evaluation would lead to an increased impact on pupil achievement. In other words, would it be beneficial in improving student performance if principal evaluation cycles were shorter, or were conducted at more frequent intervals? Another valid question is whether principal perceptions about the evaluation process, including their effects on pupil achievement, change over time. It would be intriguing to determine if there is any kind of relationship between principals' perceptions about evaluation and the number of years that they have served as a school principal. Still another unresolved question is how principals believe the evaluation process can be changed to make it more effective, both for their professional growth and to increase pupil achievement.

Though this research study helped to provide answers to several important

questions, the topic of principal evaluation has many facets. More is now known about the subject matter than before the study was conducted, but a number of questions still remain that merit future investigation. It is hoped that future researchers will continue to pursue answers to some of these questions, and will raise more of their own, as we strive to better understand the complexities of principal evaluation.

APPENDIX A: SURVEY INSTRUMENT

SURVEY INSTRUMENT

Please answer the following questions to the best of your ability:

Question 1

1. Is this the first year in which you have served as principal at the school where you are currently employed?
 - Yes (If yes, then *skip/logic* to respondent information)
 - No

Questions 2 –11 (*Randomly presented in survey driver*)

To what extent, if at all, have the following methods been used to formally evaluate your performance as principal?

	Not At All		Neutral		Very Much		
	1	2	3	4	5	6	7
2. Narrative self-evaluation							
3. Portfolio/dossier							
4. Checklist/rating system							
5. Supervisor observation							
6. Narrative evaluation by supervisor							
7. Data-based evaluation							
8. Survey data from teachers, parents, or students							
9. Peer supervision/review							
10. Anecdotal evidence							

11. Perception feedback from stakeholders

Question 12

How frequently are you evaluated in your current role as principal?

- Monthly
- Quarterly
- Semi-annually
- Annually
- Every other year
- Every three years
- Sporadically
- Never

Question 13-23 (*Randomly presented in survey driver*)

To what extent, if at all, do you perceive principal evaluation accomplishing the following objectives?

Not At All			Neutral			Very Much
1	2	3	4	5	6	7

13. Satisfy district accountability requirements

14. Increase standardized assessment scores

15. Provide principals with professional growth

16. Document sub-standard principal performance

17. Identify the needs for principal professional development

18. Provide incentive for performance improvement

19. Reward exemplary principal performance
20. Ensure adherence to policies and procedures
21. Support the maintenance of the instructional program
22. Improve pupil achievement
23. Foster positive school climate

Some ideas do not lend themselves to check box format. I would greatly appreciate if you would respond to the following four questions with your thoughts.

Question 24

What effect, if any, has your experience with evaluation had on your beliefs about principal performance? Please explain.

Question 25

What effect, if any, has your experience with evaluation had on your beliefs about your own professional development? Please explain.

Question 26

Do you believe that the principal evaluation process has an impact on pupil performance? Please explain.

Question 27

Do you believe that the principal evaluation process has an impact on your own professional development? Please explain.

Respondent Information
Question 28

Please indicate the school structure that best describes the environment in which you work.

Grades:

- PreK-8
- K-3
- K-4
- K-5
- K-6
- K-8
- 1-3
- 1-5
- 2-5
- 3-5
- 4-5
- 5-6
- 5-8
- 6-8
- 7-8

Question 29

Please indicate the number of years for which you have served as principal at the school where you are presently employed.

Number of years:

(open field)

Question 30

Please indicate the total number of years for which you have served as a school principal.

Number of years:

(open field)

Question 31

How many schools are in the district in which you are employed?

Number of schools:

- 1-3
- 4-6
- 7-9
- 10-15
- 15 or more

Question 32

Please indicate the title that best describes the supervisor who is responsible for evaluating you.

Title:

- Superintendent
- Deputy Superintendent
- Assistant Superintendent
- Director
- Other (*open field*)

Question 33

Please indicate the county in which you are presently employed.

- DuPage County
- Will County
- Lake County

REFERENCES

- Amsterdam, C. A., Johnson, R. L., Monrad, D. M., & Tonnsen, S. L. (2003). A collaborative approach to the development and validation of a principal evaluation system: A case study. *Journal of Personnel Evaluation in Education, 17*, 221-242.
- Anderson, M. E. (1991). *Principals: How to train, recruit, select, induct, and evaluate leaders for America's schools*. ERIC Clearinghouse on Educational Management, University of Oregon-Eugene, OR. (ERIC Document Reproduction Service No. ED 337 843)
- Andrews, C. (1990). Evaluating principals. *National Association of Elementary School Principals Research Roundup, 6*(2), 3-6.
- Babbie, E. (1990). *Survey Research Methods*. 2nd ed. Belmont, CA: Wadsworth Pub. Co.
- Barth, R. S. (2001). *Learning by Heart*. San Francisco, CA: Jossey-Bass Inc.
- Bolman, L. G., & Deal, T. E. (2003). *Reframing Organizations: Artistry, Choice, and Leadership*. San Francisco: Jossey-Bass.
- Brown, G., & Irby, B. J. (1995). *Administrative portfolios: A contemporary construct for evaluation*. (EDRS Document Reproduction Service No. ED 400 247).
- Brown, G., & Irby, B. J. (1996). *Administrative portfolios: The development of a portfolio system*. (EDRS Document Reproduction Service No. 400 248).
- Brown, G., & Irby, B. J. (1998a). Seven policy considerations for principal appraisal. *School Administrator, 55*(9), 10-11.
- Brown, G., & Irby, B. J. (1998b). Taking the lead: One district's approach to principal evaluation. *National Association of Secondary Schools Principals Bulletin, 82*(602), 18-25.
- Brown, G. & Irby, B. J. (2001). *The Principal Portfolio*. 2nd ed. Thousand Oaks, CA: Corwin Press.
- Budd, J. W., & Scoville, J. G. (Eds.). (2005). *The Ethics of Human Resources and Industrial Relations*. Champaign, IL: Labor and Employment Relations Association.

- Buttingnol, M., & Diamond, C. T. P. (2003). Gladiator meets Athena: The lure of the warriorress in educational leadership. *Curriculum Inquiry*, 33, 441-453.
- Campbell, D. T., & Stanley, J. C. (1971). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Canty, J., Clausen, K., Kozleski, E. B., & McGill, M. V. (2006). Administration. *Teachers College Record*, 108, 1-11.
- Catano, N., & Stronge, J. (2006). What are principals expected to do?: Congruence between principal evaluation and performance standards. *National Association of Secondary Schools Principals Bulletin*, 90, 221-237.
- Cohen, J. W. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Connecticut State Department of Education. (1990). *A guide to the process of evaluating school principals*. East Lyme, CN: Connecticut Principals' Academy. (ERIC Document Rep. Service No. ED 319 141)
- Creswell, J.W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.
- Danielson, C., & McGreal, T. L. (2000). *Teacher Evaluation to Enhance Professional Practice*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Daresh, J. C. (2002). *What It Means to be a Principal; Your Guide to Leadership*. Thousand Oaks, CA: Corwin Press, Inc.
- Davis, S. H. (1998). Superintendents' perspectives on the involuntary departure of public school principals: The most frequent reasons why principals lose their jobs. *Educational Administration Quarterly*, 34, 58-90.
- Davis, S. H., & Hensley, P. A. (1999). The politics of principal evaluation. *Journal of Personnel Evaluation in Education*, 13, 383-403.
- DuFour, R., & Eaker, R. (1998). *Professional Learning Communities at Work: Best Practices for Enhancing Student Achievement*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Duke, D. L., & Iwanicki, E. (1992). Principal assessment and the notion of "fit". *Peabody Journal of Education*, 68, 25-36.

- Ediger, M. (1998a). *Appraising the school principal*. (ERIC Document Reproduction Service No. ED 427 023)
- Ediger, M. (1998b). Change and the school administrator. *Education*, 118, 541-548.
- Ediger, M. (2001). *Assessment of school principal achievement*. (EDRS Document Reproduction Service No. ED 452 235)
- Elmore, R. (2001). Research finds the false assumption of accountability. *Education Digest*, 67(4), 9.
- Elmore, R. (2002). Hard questions about practice. *Educational Leadership*, 59(8), 22.
- Elmore, R. (2003). A plea for strong practice. *Educational Leadership*, 61(3), 6-10.
- Fredericks, J., & Brown, S. (1993). School effectiveness and principal productivity. *National Association of Secondary School Principals Bulletin*, 77(9), 9-16.
- Gay, L. R. (1992). *Educational Research: Competencies for Analysis and Application*. 4th ed. New York, NY: Macmillan Publishing Company.
- Gil, L. S. (1998). Principals evaluating peers. *School Administrator*, 55(9), 28-31.
- Ginsberg, R., & Thompson, T. (1992). Dilemmas and solutions regarding principal evaluation. *Peabody Journal of Education*, 68, 58-74.
- Glasman, N. S. (1992). Toward assessing test score-related actions of principals. *Peabody Journal of Education*, 68, 108-123.
- Glasman, N. S., & Martens, P. A. (1993). Personnel evaluation standards: The use in principal assessment systems. *Peabody Journal of Education*, 68, 47-63.
- Glatthorn, A. A. (1997). *Differentiated Supervision*. 2nd ed. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Goldring, E., & Rallis, S. (2000). *Principals of Dynamic Schools: Taking Charge of Change*. 2nd ed. Newbury Park, CA: Corwin/Sage.
- Green, R. L. (2001). New paradigms in school relationships: Collaborating to enhance student achievement. *Education*, 121, 737-743.
- Hardy, L. (2006). Questions about NCLB's effectiveness continue to divide education leaders. *American School Board Journal*, 193, 16-17.

- Hart, A. W. (1992). The social and organizational influence of principals: Evaluating principals in context. *Peabody Journal of Education*, 68, 37-57.
- Hart, A. W. (1994). *Evaluating principals in light of context and socialization*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Rep. Service No. ED 380 879)
- Heck, R. H., & Marcoulides, G. A. (1992). Principal assessment: Conceptual problem, methodological problem, or both?. *Peabody Journal of Education*, 68, 124-144.
- Heppner, P. P., Kivlighan, D. M., & Wampold, B. E. (1991). *Research Design in Counseling*. 3rd ed. New York: Brooks/Cole.
- Howell, D. C. (2004). *Fundamental Statistics for the Behavioral Sciences*. 5th ed. Belmont, CA: Thomson Wadsworth.
- Hoy, W. K., & Miskel, C. G. (2001). *Educational Administration: Theory, Research, and Practice*. 6th ed. New York: McGraw-Hill.
- Illinois Association of School Boards. (2006). *Illinois School Code, 2006 Edition*. Charlottesville, VA: Matthew Bender and Company, Inc.
- Jianping S., Cooley, V. E., & G. L. Wegenke. (2004). Perspectives on factors influencing application for the principalship: A comparative study of teachers, principals and superintendents. *International Journal of Leadership in Education*, 7, 57-70.
- Kaplan, L. S., Owings, W. A., & Nunnery, J. (2005). Principal quality: A Virginia study connecting interstate school leaders licensure consortium standards with student achievement. *National Association of Secondary Schools Principals Bulletin*, 89(643), 28-44.
- Kearney, K. (2005). Guiding Improvements in Principal Performance. *Leadership*, 35, 18-21.
- Lambert, L. (2003). *Leadership Capacity for Lasting School Improvement*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Langer, S., & Boris-Schacter, S. (2003). Challenging the image of the American principalship. *Principal*, 83, 15-18.
- Lashway, L. (2003). Improving principal evaluation. *ERIC Digest*. (ERIC Document Reproduction Service No. ED 482 347)

- Leithwood, K., & Hallinger, P. (Eds.). (2002). *Second International Handbook of Educational Leadership and Administration, Part Two*. Norwell, MA: Kluwer Academic Publishers.
- LoVette, O., & Watts, S. (2002). *Teachers' perceptions of principal: Refining the instrument*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN. (EDRS Document Reproduction Service No. ED 471 351)
- Marcoux, J., Brown, G., Irby, B. J., & Lara-Alecio, R. (2003). *A case study on the use of portfolios in principal evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 475 710)
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School Leadership That Works: From Research to Results*. Aurora, CO: Mid-continent Research for Education and Learning (McREL).
- McAdams, R. P., & Barilla, M. A. (2003). Theory and practice in performance appraisal of school administrators. *Research for Educational Reform*, 8, 19-35.
- McCleary, L. (1979). Evaluation of principals. *Theory Into Practice*, 18, 45-49.
- Mullins, T. H., et al. (1988). *Perceptions of Texas elementary principals and teachers Regarding evaluations of principals*. Paper presented at the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 305 703)
- Murphy, J. A., & Pimental, S. (1996). Grading principals: Administrator evaluations come of age. *Phi Delta Kappan*, 78, 74-83.
- Oklahoma State Department of Education. (1999). *Oklahoma criteria for effective teaching and administrative performance activities workbook*. Oklahoma City, OK: State Department of Education Printing Services. (EDRS Document Reproduction Service No. ED 439 103)
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*. 3rd ed. Thousand Oaks, CA: Sage Publications.
- Peat, J. (2001). *Health Science Research: A Handbook of Quantitative Methods*. Sydney: Allen and Unwin.
- Peterson, D. (1991). Evaluating principals. *ERIC digest series number 60*. (ERIC Document Reproduction Service No. ED 330 064)

- Rallis, S. F., & Goldring, E. B. (1993). Beyond the individual assessment of Principals: School based accountability in dynamic schools. *Peabody Journal of Education*, 68(2), 3-23.
- Rebore, R. W. (2001). *The Ethics of Educational Leadership*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Rebore, R. W. (2004). *Human Resources Administration in Education: A Management Approach*. 7th ed. Boston, MA: Pearson Education, Inc.
- Reeves, D. B. (1998). Holding principals accountable. *School Administrator*, 55(9), 6-11.
- Reeves, D. B. (2004). *Assessing Educational Leaders: Evaluating Performance for Improved Individual and Organizational Results*. Thousand Oaks, CA: Corwin Press.
- Reeves, D. B. (2006). *The Learning Leader: How to Focus School Improvement for Better Results*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Reeves, D. B. (2008). Waiting for NCLB. *Educational Leadership*, 65(6), 89-90.
- Sergiovanni, T. J. (1992). *Moral Leadership: Getting to the Heart of School Improvement*. San Francisco, CA: Jossey-Bass.
- Sergiovanni, T. J. (1995). *The Principalship: A Reflective Practice Perspective*. 3rd ed. Boston: Allyn and Bacon.
- Sergiovanni, T. J., & Starratt, R. J. (1993). *Supervision: A Redefinition*. 5th ed. New York: McGraw-Hill Inc.
- Smith, R. E. (2001). *Human Resources Administration: A School Based Perspective*. 2nd ed. Larchmont, NY: Eye on Education, Inc.
- Smith, R. M., Montello, P. A., & White, P.E. (1992). Investigation of interpersonal management training for educational administrators. *Journal of Educational Research*, 85, 242-245.
- Smylie, M. A., & Crowson, R. L. (1993). Principal assessment under restructured governance. *Peabody Journal of Education*, 68,
- Snyder, J., and Ebmeier, H. (1992). Empirical linkages among principal behaviors and intermediate outcomes: Implications for principal evaluation. *Peabody Journal of Education*, 68, 75-107.

- Starratt, R. J. (2004). *Ethical Leadership*. San Francisco, CA: John Wiley & Sons, Inc.
- Stiggins, R. (2008). *Assessment Manifesto: A Call for the Development of Balanced Assessment Systems*. Portland, OR: ETS Assessment Training Institute.
- 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. (ERIC Document Reproduction Service No. ED 452 278)
- Strickler, J. (2006). What really motivates people? *Journal for Quality & Participation*, 29(1), 26-28.
- Strike, K. A. (2007). *Ethical Leadership in Schools: Creating Community in an Environment of Accountability*. Thousand Oaks, CA: Corwin Press, Inc.
- Stufflebeam, D., & Nevo, D. (1993). Principal evaluation: New directions for Improvement. *Peabody Journal of Education*, 68(2), 24-46.
- Stufflebeam, D. L., & Pullin, D. (1998). Achieving legal viability in personnel evaluations. *Journal of Personnel Evaluation in Education*, 11, 215-230.
- Tennant, M. & Pogson, P. (1995). *Learning and Change in the Adult Years*. New York: John Wiley & Sons.
- Thomas, D. W., Holdaway, E. A., & Ward, K. L. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education*, 14, 215-240.
- Thomas, D. D., Grigsby, C. J., Miller, J. C., & Scully, W. M. (2003). Networking: A life net for embattled principals. *Principal*, 83, 40-44.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking Teacher Evaluation and Student Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Whaley, J. (Ed.). (2002). *Developing the effective principal: Hiring, evaluation, and retention practices for the superintendent*. Gaithersburg, MD: Aspen Publishers, Inc.
- Wick and Associates. (2007). *A Psychometric Audit of the Illinois Standards Achievement Test (2006-7), ISAT 2007 Audit*. [Brochure] Glencoe, IL: Wick, J.

VITA

Edward J. Condon was raised in Hinsdale, Illinois. He earned a B.A. in History from the University of Richmond (1991), a M.A. in Educational Administration from the College of William and Mary (1996), and an Education Specialist degree in Administration and Supervision from the University of Virginia (2003). Over the course of his career, he has served as a high school Social Studies teacher, a high school assistant principal, and a junior-high school principal.

Currently, Ed serves as the Principal of Butler Junior High School in Oak Brook, Illinois. He lives with his family in nearby Hinsdale, Illinois.