Karen L. Sanzo

AN ANALYSIS OF EVALUATION PLANS IN A FEDERALLY FUNDED LEADERSHIP PREPARATION PROGRAM: IMPLICATIONS FOR THE IMPROVEMENT OF PRACTICE

This study is a content analysis of selected federally funded leadership preparation program evaluation proposals for the 2008, 2009, and 2010 School Leadership Program grants. The United States Department of Education (USDE) began awarding funding in 2002 to develop and implement preparation programs for aspiring and current assistant principals and principals. These innovative leadership preparation programs are expected to influence school leaders in a way that enables them to positively impact student achievement and the schools they serve. In order to determine effectiveness, each grant proposal puts forth a comprehensive program evaluation designed to measure the grant's adherence to the program's articulated outcomes. The review and analysis of these program evaluation proposals are offered as a baseline to build upon in understanding how leadership preparation programs are being evaluated. The following themes were found from the analysis of the program evaluation content: development of methods and tools limited, evaluation theory not explicit, data sources and analysis underdeveloped, specific evaluators identified but varying degree of specificity for timelines, and intent of program evaluations exceeded minimum requirements.

School leaders make a difference in the success of schools (Hattie, 2009; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Marzano, Water, & McNulty, 2005; Robinson, Lloyd, & Rowe, 2008). Good school leaders can positively impact school climate, instructional processes, and perceptions of learning in the classroom. Ineffective leaders can be detrimental to the achievement of students and the school. According to Leithwood et al., (2004) "leadership is second only to classroom instruction among all school-related factors that contribute to what students learn at school" (p. 5) and "is widely regarded as a key factor in accounting for differences in the success with which schools foster the learning of their students" (p. 17). In today's challenging and complex world that has "complicated, politically sensitive issues," success begins "with the educational leaders who are selected to lead our schools" (Futrell, 2011, p. 643). The jobs of principals and assistant principals are difficult and the performance expectations of these individuals are high. They are tasked with increasing student achievement, promoting and sustaining change, monitoring and ensuring high-quality instruction, along with the day-to-day tasks of building management from discipline, to buses, to community engagement, and any number of other duties (Perez, Uline, Johnson, James-Ward, & Basom, 2010).

Because of the critical nature of their role in schools and the impact they have on students, it is not surprising there is attention, scrutiny, and criticism of how leaders are prepared (Murphy, 2005; Young, Petersen, & Short, 2002). Over the past two decades there has been a significant push to transform the way school leaders are prepared (Perez et al., 2010). Traditionally, brick-and-mortar universities and colleges of education have held the sole responsibility for school leadership preparation, but new educational leadership preparation providers have emerged during this movement for change. These include not-for-profit entities, online preparation programs, and school districts themselves. While there is an escalating focus on leadership preparation, research on the effectiveness of these programs is relatively limited (Orr, 2010). In fact, there is disagreement on the types of approaches and measures to use to determine if a preparation program has been able to successfully train school leaders (Orr, 2010).

In light of the debate about measures and program effectiveness, I argue here that we must better understand how programs are currently being evaluated. Programs are required to undergo analyses by agencies, whether it is NCATE, TEAC, state-level accreditation, or other program reviews. Rather than assessing for the sake of assessment to satisfy credentialing agencies and requirements of funders, program evaluations should be utilized to improve the preparation programs being evaluated, as well as other leadership programs around the country. To do this, we must first understand the quality of program evaluations, including the tools being used, the data collected, the intent of the evaluation, and other facets surrounding the evaluations.

In this manuscript I analyze the content of selected federally funded leadership preparation program evaluation proposals for the 2008, 2009, and 2010 School Leadership Program (SLP) grants. These innovative preparation programs are expected to influence school leaders in a way that enables them to positively impact student achievement and the schools they serve. In order to determine effectiveness, each grant proposal puts forth a comprehensive program evaluation designed to measure the grant's adherence to the program's articulated outcomes. The review and analysis of these program evaluation proposals are offered as a baseline to build upon in order to understand how leadership preparation programs are being evaluated.

There has been no systemic effort to date focusing on the program evaluation efforts of the School Leadership Program. This initial study's outcomes provide a birds-eye view of the program evaluation efforts of this important national program, and synthesize the evaluation findings.

Preparing School Leaders

A 2010 Wallace Foundation survey found that improving school leadership is one of the top-ranked priorities for school reform (Wallace Foundation, 2011). In fact, principal leadership ranked second only behind concerns about teacher quality. This is not a surprising statistic, especially

in light of the renewed calls for changes in the preparation of leaders in the last decade and the debate around leadership preparation for the last two decades (Hackmann & Wanat, 2007; Levine, 2005; Murphy, 2005). In fact, it has been asserted that a "lack of clear understanding about what educational leadership preparation programs should be and what content, instructional methods, and structures should frame them is at the heart of this tension (La-Magdeleine, Maxcy, Pounder, & Reed, 2009, p. 130).

However, research by scholars such as Orr and Barber (2007), Darling-Hammond, LaPointe, Myerson, Orr, and Cohen (2007), and Orr and Orphanos (2011) have found there are components of high-quality programs that positively impact principals' leadership in schools. These include supportive program structures, intensive internships, competent faculty, student-centered instruction, and an innovative program structure (Orr, 2010). Common characteristics of the currently funded group of SLP grants include a strong mentoring and/or coaching component, intensive internships, an innovative program structure, and an accelerated program model.

Another SLP grant characteristic is that each grant is a partnership effort between multiple stakeholders which include the school district, and may also include a university, a not-for-profit, and other agencies interested in leadership preparation. Partnerships are vital to creating a successful leadership program and "no single organization, group, or individual can create the kind of leadership preparation that our nations' children need and deserve" (Young, Petersen, & Short, 2002). While traditionally the domain of universities and colleges of education, not all USDE SLP grants include an institution of higher education. According to Crow (2006), "evidence regarding the quality of university preparation programs is scant, and most arguments resort to anecdotal evidence or have questionable methodologies" (p. 312). This may not be the sole reason that districts choose to work with alternative certification programs in some of the grants, but it is one possibility. Regardless of the partnership, collaboration is crucial and it must be authentic and connected to practice. Munoz, Winter, and Riccardi (2006) assert that finding examples of these in the literature is difficult, as there are "relatively few examples of successful partnerships" (p. 13). This provides another reason to delve deeper into this study about the U.S. Department of Education (USDE) SLP grants. to explore how these partnerships can prepare effective school leaders.

Current State of Research in School Leadership Preparation

While its use increasing, there is still little empirical research available about the preparation of school leaders. It is a relatively new area for researchers and "limited by scholarly skepticism over the perceived legitimacy and difficulties of launching large-scale comparative research" (Orr & Orphanos, 2011, p. 23). Murphy and Vriesenga (2006) cite the dearth of research on school administration in general, and even more dra-

matically lacking for leadership preparation specifically. "Given the applied nature of the profession and the centrality of preparatory activities to departments of educational leadership, the fact that serious academic work on pre-service training remains a minor element in the school administration scholarship mosaic is as surprising as it is disappointing" (p. 187). In *The Handbook of Research on the Education of School Leaders,* Kottkamp and Rusch (2009) offer the same lament, making an allegorical comparison of the products of researchers of leadership preparation to "a lot of islands sprinkled across a vast sea…a larger number of islands are small… more numerous yet are tiny atolls sprinkled widely apart, with no possibility of building knowledge masses of substance" (p. 80).

Perhaps research on leadership preparation is behind other social sciences because of an assumption that we know how to prepare leaders, or a reliance on historical comfort to replicate the status quo, or a potential fear of innovation and change. This is conjecture, but whether this hypothesis is true or not, the way school leaders are trained must change. School leaders impact student achievement in significant ways (Hattie, 2009; Leithwood et al., 2004; Marzano, Water, & McNulty, 2005; Robinson, Lloyd, & Rowe, 2008) yet problems continue to persist in the United States educational system. Ball and Forzani (2007) suggested in the Wallace Foundation Distinguished Lecture in *Educational Researcher,* "The fact that educational problems endure despite repeated efforts to solve them suggests the fallacy of this reliance on common sense. Disciplined research on problems and solutions could help in education, just as it does in other domains" (p. 529).

Another critique by Murphy and Vriesenga (2006) is the lack of enhanced quality in empirical studies with the increase in methodological scaffolding supporting the research. Perhaps this is attributed to the lack of funding that is available for research on leadership preparation and is often conducted "out of [professors'] back pockets," relying on the kindness of current program students and graduates, as well as anecdotal records (p. 189). Kottkamp and Rusch (2009) state the "resource availability for expanding research on leadership preparation appears to be next to nil" (p. 78). This lack of funding to research on leadership preparation creates a compelling urgency to maximize the research potential for current projects, such as the USDE SLP.

This examination of the program evaluation efforts of selected grants provides a baseline for further exploration into the effectiveness of these programs. One of the challenges in researching projects such as the SLP grants is that they are designed to be service, and do not have an explicit research component articulated beyond reporting on the required program outcome measures (which in general limit the reporting data to such information as details on the number of individuals trained and placed in leadership positions). The onus appears, in fact, to be on the project personnel themselves to transform the service project data points collected in the program evaluation into usable research that is disseminated to the field at large.

Program Evaluation

Program evaluations are not new, and are used in many facets of society (Fitzpatrick, Sanders, & Worthen, 2011). There is a difference between research and evaluation, which can make the use of program evaluations based on the data collected politically challenging. They have a different purpose: evaluations provide information that is useful to stakeholders in order to make a decision or a judgment, while research contributes to the knowledge within a field, and can either generate the development of a theory or add to an existing one (Fitzpatrick et al., 2011). Patton (2008) highlights the challenge of getting people to use the knowledge produced from evaluations. Regardless of the purpose of the evaluation, whether for formative or summative purposes, the information should be used (Morabito, 2002). It is critical to analyze the initial program evaluation design, including the potential tools to use and data to collect in order to explore the potential application of the evaluation for research purposes.

The development of a program evaluation can take on many different designs. There are evaluation models which focus on accountability, compare contexts, make a cost-benefit analysis, focus on critical issues, examine internal programs and processes, focus on mission or goals, assess outcomes, or program theory, and address utilization issues, among others (Patton, 2008). For a more extensive and detailed list, see Patton's (2008) *Utilization-Focused Evaluation.*

Each program in the USDE SLP grant initiative requires a program evaluation component. The federal government outlines specific guidelines for reports due to be submitted to the program officer twice a year. Grantees are required to report on principal certification percentages, hiring percentages of program completers as assistant principals and principals, and retention rates for hires. Within these reporting categories, grantees are expected to provide data collection procedures, rationales for missing data, information on progress or lack thereof in achieving their program objectives, and an explanation about the program completers. Additional information about the project is also requested in both qualitative and quantitative formats where appropriate in both the annual performance reports and final performance report.

While simply stated here, the information required to complete the reports is complex and therefore ripe for research on leadership preparation programs, as well as secondary and tertiary impacts of the principals and assistant principals on schools, faculty, and student achievement. In order to discern that information, a preparation program with well-articulated goals and objectives needs to be developed, implemented, and assessed with a carefully designed program evaluation that utilizes both formative and summative assessment methods. It is neither the purpose nor intent of this manuscript to evaluate the quality of the leadership preparation program proposal, nor to make a judgment on the quality of the evaluation. Rather, this study

serves as an initial baseline effort to begin to tease out the nuances of program evaluation in federally-funded school leadership program grants, for which funding is currently available. With better understanding about the nature, processes, and content of the program evaluations, further research on leadership preparation programs can build on this foundation.

Methods

The purpose of this manuscript is to present findings from a content analysis of selected program evaluations from the U.S. Department of Education (USDE) School Leadership Program (SLP) proposals. The research questions addressed are:

1) What evaluation methods and tools are proposed in the SLP grant applications?

2) What evaluation theories are used in the program evaluation proposals?

3) What are the data being collected and how are they analyzed?

4) Who are the evaluators and what are the timelines they have established?

After several initial reviews of the evaluation proposals, as the content analysis began, a fifth question emerged:

5) What is the purpose of the program evaluation?

An initial analysis of the proposals indicated a wide variety of purposes for their evaluations.

The grant applications used for this study were purposefully selected from the larger pool of USDE SLP proposals. The competitive USDE SLP program, which began in 2002, is designed "to support the development, enhancement, or expansion of innovative programs to recruit, train, and mentor principals (including assistant principals) for high-need LEAs" (USDE SLP, n.d.). Over $158,000,000 has been appropriated for 90 funded programs. Programs can focus on aspiring leadership preparation for aspiring assistant principals and principals, training for current school leaders (assistant principals and principals only), or a combination of both. While any type of district—rural, suburban, or urban—may be a part of the grant application, eligibility is limited to only high-need local education agencies. Currently 43 active USDE SLP grants have been awarded. For the purpose of this grant, proposals focused on single urban school districts were analyzed for the currently funded grants (2008, 2009, and 2010 award cycles). Seventeen proposals met the initial selection criteria. Two evaluations were excluded from the content analysis due to the inability to identify program evaluation components unique to the SLP grant (multiple grants outside of the SLP were interwoven into the evaluation plan). Fifteen evaluations were reviewed for this study.

A content analysis was conducted on components of the proposals for the USDE SLP grant (Patton, 2002). Manual coding was utilized

for organizing, classifying, and coding the data. Each program evaluation section was initially reviewed in the selected grant proposals, followed by an overall review of each complete grant application. This process allowed the researcher to identify the specific information required from each grant application that would be needed for a more detailed analysis to address the research questions. This analysis included the following items: proposal abstract, program evaluation, management plan, and appendices containing relevant information related to the evaluation proposal (i.e. logic model and timeline). The researcher also took notes during this step in the analysis process to inform the preliminary coding schema. These materials were reviewed once using a provisional "start list" to guide the process (Miles & Huberman, 1994). An outside expert reviewed the start list and the preliminary coding guide developed from reviews to in order to confirm the coding guide. Feedback on this schema was solicited from outside experts, and revisions were made (see Appendix A).

Limitations

The study analyzes only the proposed program evaluations for single-district applications in the urban district context, excluding programs in suburban and rural settings, as well as multi-district programs. However, it is argued that these types of programs are potentially qualitatively different from one another and could potentially require a different program evaluation approach and require separate reviews and analyses outside the scope of this study.

The program evaluations could also change; the grantees have indicated in their proposals that they will change based upon the development of the grant over the course of the five-year award period. While these changes should be explored and documented, it is important to detail the program evaluation proposals, methodologies, tools, and other components of the evaluation in the proposal stage. This provides an understanding of initial program evaluation design in the leadership program and insight into what the federal government deems acceptable for a program evaluation proposal at the time of the awarding of the grant.

Findings

The results of the content analysis are presented below. The following themes were identified from the analysis of the program evaluation content: development of methods and tools were limited, evaluation theory not explicit, data sources and analysis underdeveloped, specific evaluators identified but varying degree of specificity for timelines, and intent of program evaluations exceeded minimum requirements. These results reflect only what was found specifically in the grant application, and may not reflect what is currently in practice with the funded USDE SLP grant program evaluations.

Development of Methods and Tools Were Limited

The grant applications indicated the use of either "qualitative and quantitative measures" or "mixed-methods approaches" in the evaluation plan. The plans consistently lacked specificity about the types of tools and measures that were being used. Most of the applications indicated that "interviews, focus groups, and surveys" would be conducted. However, generally there were no descriptions of interviews and surveys, how these were developed or were going to be developed, whether these were internally developed for the purpose of the grant and validated, externally developed and validated, or the protocols for the interviews and focus groups. Several of the applications also stated the tools would be made or selected at the time of the grant award or during the grant process. For example, language about this from applications included "appropriate instruments" would be used, a "variety of techniques" would be employed, "tools and protocols" would be adopted from existing instruments in the literature, that some tools were already validated and others were "under development," and "all surveys and observation protocols will be developed for use in formative assessments."

Only a few tools were specifically mentioned in the grants. The following list indicates the number of grants that indicated the use of these tools in their proposal:

- School Leadership Licensure Assessment (SLLA) (2)
- Vanderbilt Assessment of Leadership in Education (VAL-ED) (2)
- NASSP 360 Assessment (2)
- School Leadership Preparation and Practice Survey (1)
- District Climate Survey (1)
- Decision Making Inventory (1)
- School District Self-Assessment for Leaders (1)
- Problem Solving Inventory (1)
- Leadership Skills inventory (1)
- NASSP Developing the Twenty-First Century Principal Assessment (1)
- Gallup Principal Perceiver (1)

Additional information on the types of tools and methods were lacking.

Evaluation Theory Not Explicit

It was impossible to discern any underlying evaluation theory or theory of action from the presentation of the program evaluation proposals. Five grant proposals contained logic models, some more detailed and correlated to the grant goals and objectives than others, but there were no

articulated evaluation theories. This is not to say there were no theories used; none were provided, however.

Data Sources and Analysis Underdeveloped

Many of the data sources proposed for collection in the program evaluations were vague. For example, most of the program evaluation proposals indicated that surveys would be conducted at varying times throughout the grant process, but the types of data ranged from "participant survey" to more specific school climate data, mentor surveys, faculty and staff surveys, and "pre/post surveys." In general these types of data were to be "analyzed" with little to no description about the analysis process. In addition, most of the grants were designed to conduct observations of such activities as program meetings, courses, and professional development activities. There was no description of either how these observations were being conducted or the method of analysis.

A number of documents were listed in the grants that were intended to be used for both formative and summative purposes. Some of the grants simply indicated that documents would be collected and analyzed while others were somewhat more specific (although still vague, i.e. "products"). These included: cohort rosters, training schedules, curricula, program products, case study activities used in the training programs, certification documents, attendance sheets, and leadership placement documents. In general these types of data did not have a specific mode of analysis.

Twelve grants identified the use and analysis of student achievement data using standardized assessments as a gauge for grant success. Details about the analysis of this information varied. Some proposals included a variety of statistical methods to demonstrate impact on student achievement while others only indicated that the data "would be analyzed". For example, one grant noted the use of student achievements on state level assessments would be used as a proxy for student learning. The achievement of students would be compared over a five year period in schools matched by student, community, and pre-learning variables. Another grant indicated that aggregate school-level achievement data would be used to determine the rates of academic progress of students in schools led by graduates of the preparation program and compared with similarly situated leadership peers who did not participate in the grant-funded initiative.

Data collected for the program evaluations included all measures required for reporting purposes to the USDE. These mandated metrics are more managerial in nature, disconnected from previously cited research. Included in the reporting requirements are the number of applicants in the leadership preparation program and the number of program members who received initial principal (or assistant principal) certification. Reports required grants to also collect data that demonstrated the number of program completers who secured positions as assistant principals and principals, as

well as retention of these school leaders. The reporting requirement also includes information demonstrating some form of academic achievement although the specific metric for this was left up to each individual program.

Overall results of data collection were presented in generalized formats which included surveys, interviews, observations, feedback, and products, as examples. Other types of data identified for collection during the evaluation process were presentations, simulation responses, assessment processes, exit interviews, screening data, needs assessments, and human resource data. The methods of analysis were also vague. While some grant proposals provided what seemed to be specific data analysis methods, such as using a matched program control with a "quasi-experimental design using district-wide hierarchical linear modeling and a comparative time series analysis," very few were this specific. For example, one grant proposed the use of multivariate analysis, but without details regarding the data sources for that particular analysis approach.

Specific Evaluators Identified but Varying Degrees of Specificity for Timelines

All but one of the fifteen grants specially identified the evaluators. Seven grants used an external program evaluation company and two used program evaluators that were independent contractors. One program evaluator was a university-based group internal to the lead grant, one used a faculty member internal to the lead grant, and one used a private company internal to the lead grant. Two of the projects used program evaluators that were listed as grant partners (one was listed as an evaluation group and one identified the person) and one grant did not have an evaluator identified at the time of the proposal. Several of the grants used external evaluators as well as internal evaluators and clearly delineated the roles and responsibilities for data collection and for analysis and project reports.

The grant proposals included timelines for implementation. Several of the program evaluations used their own evaluation-specific timelines. Of the remaining grants, evaluator activities were embedded into the overall project timeline. Eight of the proposals contained very detailed timelines, and the other seven provided vague data collection points and limited information about when the evaluation components would be implemented. While many of the grant proposals did not include specific evaluation tools, specific data to collect, or methods of analysis, the more detailed timelines contained very specific information on when the yet-to-be-created/adopted tools were to be deployed (i.e. course products, participant surveys, focus group, and interviews).

Intent of Program Evaluations Exceeded Minimum Requirements

The program evaluation proposals had to address the following minimum requirements as outlined in the USDE SLP Request for Proposals:

1) The extent to which the methods of evaluation include the use of objective performance measures that are clearly related to the intended outcomes of the project and will produce quantitative and qualitative data to the extent possible.

2) The extent to which the methods of evaluation will provide performance feedback and permit periodic assessments of progress toward achieving intended outcomes.

All of the proposals met the minimum requirements. The proposals described how the evaluations were designed to evaluate the different goals and objectives articulated in the project narrative. As indicated in the findings within this paper, a variety of qualitative and quantitative measures were proposed (whether explicit or not), along with varying degrees of descriptive specificity related to analysis and evaluation.

Beyond these minimum requirements, all of the grants focused on both formative and summative components of the evaluations. While the evaluations did not have an expressed theory of action or program evaluation theory underpinning the process, many of the evaluations posed research questions that helped guide the data collection process. Feedback loops were written into the evaluations, including timelines for the evaluators and project personnel to meet, discussing formative evaluation findings, and discussing recommended changes in program implementation that would positively affect the grant. These loops were integral components of those evaluations which were designed to integrate ongoing recommendations from the evaluations into the project in order to improve the fidelity of implementation and project participant success.

Presenting findings that positively impact leadership preparation, and disseminating them to the field at large, were identified as the articulated purpose in several grants. Some of the proposals also designed a feedback loop to the larger organization(s), connected in some way to the grant partners, to impact related programs in a positive way based on the results of the formative and summative evaluation components. Several grant applications made reference to identifying "what works" or "best strategies" in preparing aspiring school leaders. It appeared that while the primary purpose of these evaluations was to inform the grantees about formative changes that needed to be made in the project, many of the grantees were also interested in how their projects could be used to impact leadership preparation at the national level and in other programs.

Discussion and Conclusion

Recognizing the importance and impact of school leaders on the success of students and schools (Hattie, 2009; Leithwood et al., 2004; Marzano, Water, & McNulty, 2005; Robinson, Lloyd, & Rowe, 2008), and the need to prepare highly qualified principals and assistant principals, this study examined how federally funded school leadership preparation programs proposed to evaluate the impact of grant activities. Acknowledging that this research is only the first step in exploring a large-scale study of the USDE SLP programs, I sought to establish a baseline understanding of how the program evaluations were conceptualized at the funded application phase through a content analysis of federally funded program evaluations in selected 2008, 2009, and 2010 USDE SLP grants. Single urban district grants that included an aspiring leadership component were analyzed. The findings revealed that while the program evaluation proposal adhered to the intent of the grant evaluations in terms of measuring the intended outcomes related to the goals and the objectives of the proposal, (and expanded upon the USDE SLP articulated intent of the evaluations beyond the scope of project impact in many cases), the actual proposals were vague relative to the type of tools that would be used, data collected, and methodological analysis.

The evaluation plans proposed in the applications exceeded the expectations outlined by the federal government for acceptable proposals. While the use of the evaluations were necessary to effect ongoing improvement in the grants themselves, and incorporated articulated feedback loops including communication networks and responsibilities, many of the grants sought to use the findings to impact the larger field of leadership preparation. It was apparent from the review that most grant writers wanted to provide valuable data to help districts, universities, and other leadership preparation providers. It seems that at the time of the inception of most of the program proposals, methods to do this were not specific, including how this information would be disseminated (other than conference presentations).

Many of the programs reviewed in the proposals were "innovative" programs that deviated from the "norm" preparation standards expected by one of the grant partners. Several of the grants used university-based leadership programs to provide the requisite curriculum and credentialing activities for principal and assistant principal licensure. While adhering to the base requirements of the university programs, the grant-funded initiatives proposed to deviate from the traditional route to licensure articulated in the university program using such methods as an accelerated preparation timeline, in-depth and longer internships, mentoring and/or coaching methods that were not present in the traditional program, and more involved partnerships that allowed entry into the school districts' leadership initiatives. It can be argued that not all components of the proposed preparation programs for the grants had been fully fleshed

out at the time of proposal submission, and the evaluators were waiting for the final program components before being able to articulate a fully developed evaluation in place with specific tools, types of data to be collected, and methods of analysis.

This position can be supported by the fact that grants with better articulated evaluation plans, including more specific timelines and tools, were already in place prior to awarding the grant. Several of the grants were continuations of ongoing projects, either funded previously by a USDE SLP grant or by other funding sources. It seems that these grant evaluators had a clearer understanding of the programs, and could design a better developed plan (although some of the descriptions of the tools, data, and methods in these were also vague). Perhaps it was difficult to create a fully developed program evaluation at the proposal stage for those programs without a better understanding of how all of the components of the grant would be implemented.

One of the more curious findings was that no clear theory of evaluation was articulated in any of the grant proposals. While this was not a requirement of the proposal component of the grant application, evaluation still requires some type of articulated framework with theoretical underpinnings. While logic models were presented in five of the evaluations, these did not emerge from an articulated theory of evaluation. Lack of a written theory in the proposal does not mean that development of the evaluation proposal was not guided by an evaluation theory.

Leadership preparation program providers need to have well-planned program evaluations prior to the implementation of the program itself. Results from such evaluations can serve to inform the preparation provider, as well as the field at large. There is a recognized lack of research on school leadership preparation (Crow, 2006; Kottkamp & Rusch, 2009, Munoz, Winter, & Riccardi, 2006; Murphy & Vriesenga, 2006; Orr, 2010; Orr & Orphanos, 2011). Opportunities to evaluate programs such as the USDE SLP grants can be used for rich research into the nuances and impacts of the preparation of school principals and assistant principals through a well-planned and implemented program evaluation that has a clearly articulated purpose, theoretical underpinning, and research foci woven into the plans. Orr, Young, and Rorrer (2010) created a program evaluation guide for exactly such a purpose: *Developing Evaluation Evidence: A Formative and Summative Evaluation Planner for Educational Leadership Preparation Programs*. This tool, and others like this, can help program developers and evaluators guide the research development process, including rigorous research.

Future Research

Future research should be conducted on the USDE School Leadership Program grants, as it is a funding source for leadership preparation

and development in an era where little fiscal support for research on leadership preparation is available (Kottkamp & Rusch, 2009). First, the remaining grant applications need to be reviewed to determine whether the results from this research are consistent with results from studies of multi-district grants and district contexts other than urban. Second, a follow-up to this study needs to be conducted to identify changes in program evaluations between the time that a grant was awarded through the full implementation of the programs and their associated evaluations. This will identify whether there was an underpinning evaluation theory, the types of tools actually used during the evaluations, the types of data collected and their usefulness, and how these data were analyzed. While the results of the programs on the success of leaders who graduate from these programs are not generalizable, the preparation outcomes are still informative, especially if common preparation methods are used (i.e. mentors/coaches, in-depth internships, focus on theory to practice application, and case studies and problem-based learning activities), and are found to have an impact on schools.

Third, the U.S. Department of Education should consider using one external program evaluator for the general reporting requirements that measure the articulated outcomes, standardized across all programs. Internal grant evaluators are critical to these grants and should continue to be an integral component of the USDE SLP grants. The use of one central reporting agency for standardized measures (number of candidates in the program, number of candidates who secure principal or assistant principal positions, and the number of years the school leader retains an administrative position) will allow the internal evaluators to focus their evaluation efforts on formative evaluations. This enables programs to make ongoing reviews and make real-time changes to impact current grant participants, as well as make longer-term recommendations for the grant program. These evaluators can and should take on the dual-role of evaluator and researcher, producing not only evaluation reports but also helping the program partners develop research questions and implement research plans. Only through a focused, coherent effort can the field move forward in understanding what comprises a quality leadership program and the ultimate impacts these programs have on the preparation of high-quality leaders who positively impact schools, staff, and most importantly—students.

References

Ball, D. L., & Forzani, F. F. (2007). What makes educational research "educational"? *Educational Researcher, 36*(9), 530–540.

Crow, G. M. (2006). Complexity and the beginning principal in the United States: Perspectives on socialization. *Journal of Educational Administration, 44*(4), 310–325.

Darling-Hammond, L., LaPointe, M., Myerson, D., Orr, M. T., & Cohen, C. (2007). *Preparing school leaders for a changing world: Lessons from exemplary leadership development programs.* Palo Alto, CA: Standford Educational Leadership Institute.

Fitzpatrick, J. L., Sanders, J. R., & Worthan, B. R. (2011). *Program evaluation: Alternate approaches and practical guidelines* (4th ed.). Boston: Pearson.

Futrell, M. H. (2011, July). Redefining leadership: Working together to change America's education paradigm and prepare for tomorrow's world. *Journal of School Leadership, 2,* 635–648.

Hackmann, D. G., & Wanat, C. L. (2007). Licensing principals: Iowa's preparation reform initiative. *Journal of Research on Leadership Education, 2*(3), 1–36.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Kottkamp, R. B., & Rusch, E. A. (2009). The landscape of scholarship on the education of school leaders, (1985–2006). In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 23–85). New York: Routledge.

LaMagdeleine D., Maxcy B. D., Pounder D. G., Reed C. J. (2009). The context of university-based educational leadership preparation. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 129–156). New York: Routledge.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning.* New York: Wallace Foundation.

Levine, A. (2005). *Educating school leaders.* New York: Education Schools Project.

Marzano, R. J., Water, T., & McNulty, B.A. (2005). *School leadership that works: From research to results.* Alexandria, VA: Association for Supervision and Curriculum Development.

Miles, M. B., & Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Morabito, S. M. (2002). Evaluator roles and strategies for expanding evaluation process influence. *American Journal of Evaluation, 23*(3), 321–330.

Munoz, M., Winter, P., & Ricciardi, D. (2006). Inter-organizational research collaboration in education: A district-university partnership model. *ERS Spectrum, 24*(1), 13–17.

Murphy, J. (2005). Unpacking the foundations of ISLLC standards and addressing concerns in the academic community. *Educational Administration Quarterly, 41*(1), 154–191.

Murphy, J., & Vriesenga, M. (2006, April). Research on school leadership preparation in the United States: An analysis. *School Leadership and Management, 26*(2), 183–195.

Orr, M. T. (2010). Pipeline to preparation to advancement: Graduates' experiences in, through, and beyond leadership preparation. *Educational Administration Quarterly, 47*(114), 114–172.

Orr, M. T., & Barber, M. E. (2007). Collaborative leadership preparation: A comparative study of innovative programs and practices. *Journal of School Leadership, 16*(6), 709–739.

Orr, M. T., & Orphanos, S. (2011). How graduate-level preparation influences the effectiveness of school leaders: A comparison of the outcomes of exemplary and conventional leadership preparation programs for principals. *Educational Administration Quarterly, 47*(18), 18–70.

Orr, M. T., Young, M. D., & Rorrer, A. K. (2010). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs.* Retrieved from website The National Center for the Evaluation of Educational Leadership Preparation and Practice: http://www.edleaderprep.org

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Perez, L. G., Uline, C. L., Johnson, J. F., James-Ward, C., & Basom, M. R. (2010). Foregrounding fieldwork in leadership preparation: The transformative capacity of authentic inquiry. *Educational Administration Quarterly, 47*(217), 212–257.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635–674.

U.S. Department of Education USDE). (n.d.). *School leadership program.* Retrieved from http://www2.ed.gov/programs/leadership/

Wallace Foundation. (2011). *The school principal as leader: Guiding schools to better teaching and learning.* New York: Author.

Young, M. D., Petersen, G. J., & Short, P. M. (2002). The complexity of substantive reform: A call for interdependence among key stakeholders. *Educational Administrative Quarterly, 38*(2), 137–175.

Karen L. Sanzo is an Associate Professor in the Department of Educational Foundations and Leadership at Old Dominion University, Norfolk, Virginia.

Appendix A

Partners
- Lead partner
- other

Evaluation Approach
- articulated evaluation theory
- articulated methodological approach
- articulated purpose

Evaluation Tools
- empirically validated and tested
- developed during the grant by evaluators for the grant
- developed by a third party vendor
- formative use
- summative use
- local consumption use
- providing data intentionally transferable to other leadership prep programs

Data
- quantitative
- qualitative
- project participants (aspiring leaders)
- project partners
- timeline for data collection
- articulated means of analysis

Evaluator
- Internal to lead grantee
  - private company
  - university-based group
  - university-based person
  - district-based
  - private person
- Internal to one of the project partners
  - private company
  - university-based group
  - university-based person
  - district-based
  - private person
- Third party vendor/external
  - private company
  - university-based group
  - university-based person
  - district-based
  - private person