

Evaluating Innovative Leadership Preparation: How What You Want Drives What (and How) You Evaluate

Frederick C. Buskey

Assistant Professor of Educational Leadership
250 Killian Building
Western Carolina University
Cullowhee, NC 28723
fbuskey@email.wcu.edu

Meagan Karvonen, Ph.D.

Associate Professor of Educational Research
239 Killian Building
Western Carolina University
Cullowhee, NC 28723
karvonen@email.wcu.edu

Abstract

Educational leadership preparation programs are expected to train graduates who change their practice and produce outcomes for teachers and students. However, programs are challenged to produce evidence of their impact while also evaluating for formative purposes. This paper describes collaboration between an educational leadership program director and a program evaluator to construct an evaluation system that incorporated program theory, processes, and outcomes. The leadership preparation program, grounded in ethical leadership practices, had a unique design with core tenets that informed choices about the evaluation design. Decisions about data sources were informed by evaluation foci, the availability of existing data sources, and resource constraints. The complexity of the evaluation design paralleled the complexity of the program itself. Leadership content expertise, evaluation design expertise, and genuine collaboration were all essential to the successful design of this evaluation plan. Several recommendations are offered for others collaborating to design evaluations of their programs.

Introduction

Building bridges is the work of education. In graduate education we often speak of the bridge from theory to practice (Hoy & Miskel, 2008; Starratt, 1991; Walker, 1993). However, if the theory rests in the university and the practice resides in K-12 schools, there is no guarantee that the bridge will be built. At our regional comprehensive university, we recently redesigned our principal licensure programs to blend practice in K-12 schools with learning in our graduate courses. Assessing whether or not we are successful in creating the theory-to-practice bridge presents complex challenges. Orr (2009) described the pressures and pitfalls of linking program preparation to leadership practice, yet called on programs to “build measures into our assessment systems that look beyond standardized test data to measure student and program accomplishment” (p. 448). Beyond outcome measures, stakeholders need additional types of information for both formative and summative evaluation purposes. While some researchers have described how individual assignments link theory to practice (Smith & Roebuck, 2010) or how a program affects adaptive skills in particular (Blackwell, Cummins, Christine, Townsend, & Cummings, 2007), this paper explores the challenges of evaluating the holistic theory-to-practice effects of a complete program aligned to the program’s intentional ethical design.

This paper describes the process of developing an evaluation system for our newly designed leadership programs. We begin with some background, including the programmatic context and the evaluation theories that guided the work. We then describe how the evaluation was conceived, including an overview of the evaluation design process, some factors that influenced the design choices, and the design elements themselves. We conclude with recommendations for a step-by-step process for constructing an evaluation system. As others grapple with the changing landscape of program evaluation and the increasing pressures to link leadership preparation with results in the schools (Duncan, 2010; Fry, O’Neil, & Bottoms, 2006; Orr & Orphanos, 2011), we hope to provide a framework for evaluating principal licensure programs. We address multiple evaluation questions, including the relationship between preparation, practice, and impact in schools, and how to tailor evaluation to the program’s intended design. The purpose of this paper is not to describe evaluation findings, but to outline an evaluation design process. By detailing our processes, we hope to help others involved in leadership education develop their own program evaluation capacity.

Background

The foundation for this exploration includes the contextual elements of the specific program being evaluated and the theoretical frameworks of several

different, yet complimentary, approaches to program evaluation. The unique program model drove many of the decisions about the evaluation design.

Context of the Programs

In the fall of 2007, seven faculty members began redesigning two programs for preparing school leaders: the Master's of School Administration (MSA) and what came to be known as the Post-Master's Certificate in Public School Leadership (PMC). The programs now consist of a common core of 24 credits including an assessment course; four integrated, consecutive leadership courses; three consecutive internships; and, three specialized administrative topics courses. The MSA includes 15 additional credits. The programs admit two cohorts of up to 22 students each fall and spring. Cohorts are integrated. PMC and MSA students work together in online learning communities. Buskey & Jacobs (2009) and Buskey & Topolka-Jorissen (2010) previously described the programs' redesign processes and features. Five aspects of the programs' innovations are of particular relevance to program evaluation. The program is based on developing *ethical leadership* as a foundation for decision-making and action. Since a significant number of participants do not intend to become line administrators, the program features an approach to developing *leaders as change agents regardless of hierarchical positioning*. Every course contains *continuous practice in the field*. The program design reflects the *complexity* of school leadership through a correspondingly complex curricular design. The core courses build upon each other, revisiting common themes in increasing layers of complexity. Finally, program participants are expected to develop into *leaders who take action* by engaging in multiple and complex tasks in their schools.

Experimental implementation of the program began in the fall of 2008. In fall 2009, the faculty determined that the old program evaluation system was inadequate for evaluating the new program because the old assessment system was designed to ensure that completers provided evidence covering minimal program competencies and was used to gather evidence for accreditation reviews. The old assessment system did not provide feedback on where the curriculum was strong or weak, nor was there any way to measure how participation in the program impacted participants' actions in their schools. The old system did not capture any of the unique features of the new program. In the fall of 2009, the authors, one the MSA and PMC program director and re-design leader and the other a program evaluator, began designing a comprehensive evaluation system.

Evaluation Theory

The new program evaluation plan is grounded in a combination of program evaluation theory and models. Drawing on Rossi, Lipsey, and Freeman's (2004)

framework, we targeted three levels for this evaluation: program theory, processes, and outcomes. Program theory is often overlooked in program evaluation but is essential to interpret evaluation findings and guide future modifications to the program's implementation (Weiss, 1998). A program's theory articulates what theoretically happens to the targets of an intervention as a result of program implementation.

Evaluation of program processes is important for several reasons. For instance, the results may be used to help monitor and adjust program delivery and improve efficiency. Findings from process evaluation may also be used to interpret program outcomes; it is important to know the nature of the "program" that created the outcomes (Rossi et al., 2004). In the context of the MSA and PMC programs, processes include everything from resources, to scheduling, to curriculum, to admissions and advising.

Because documentation of outcomes is necessary in today's educational climate, we gave careful consideration to the structure of this part of the evaluation plan. Several authors in education and organizational development (e.g., Guskey, 2000; Kirkpatrick, 1998) have provided similar models for measuring training outcomes based on stages or levels of outcome. For example, Kirkpatrick's model includes four levels of outcome: participant reactions, learning, behavior (changed actions on the part of the trainee), and results (the intended impact of training on the problem, setting, etc.). Kirkpatrick's model is not linear or purely sequential, and others have questioned whether each step depends on previous steps (cf., Schneier, Beatty, & Russell, 1994). We conceptualized these steps as sequential yet interdependent; the sequence occurs repeatedly during the program because of expectations for immediate and ongoing application of learning to the K-12 school environment. As described later, data sources are linked to these four steps.

The decision to focus on program theory, processes, and outcomes was driven by program maturity and faculty information needs. The faculty wanted to design an evaluation that examined graduates' outcomes, consistent with the national policy emphasis on outcomes. However, without articulating the program theory or understanding its processes, interpretation of outcome data would be very limited. Patton's (2008) principles of utilization-focused evaluation also guided our work. Focusing on the intended uses of the evaluation data and identifying the intended users of the findings helped us prioritize the evaluation purposes, sift through potential data sources, and develop a plan that fit our situation and resources.

Evaluation Design

The authors' areas of expertise were in different fields, one in educational leadership and the other in program evaluation and research methods. This section describes the necessity and process of sharing knowledge in order to develop a collective understanding of need and a sound working relationship. We also discuss the involvement of program faculty and the clarification of the purposes of the evaluation.

The Design Process and Evaluation Purposes

The first concrete step to developing the evaluation system required clearly identifying the reasons for evaluation. However, there were other important, preliminary steps. Although the two authors were (and remain) close colleagues, each spoke a different language when it came to evaluation. The program director had conducted an extensive qualitative study of the impacts of individual courses on professional self-image and action (Buskey, 2010) and had grandiose ideas about what program evaluation should be able to tell the faculty. He also had multiple motivations for pursuing program evaluation such as curriculum improvement, teaching feedback, support for external awards, research interest, and knowledge of impact on schools, but lacked clear priorities and a deep understanding of evaluation theory. The evaluator was experienced in designing and conducting program evaluations for a variety of purposes and settings, but lacked knowledge of educational leadership as a content area and the unique curriculum and context of the MSA and PMC programs.

The initial foundation for this work emerged from conversations that featured give and take based on what we both brought to the project. Because the evaluator did not know educational leadership literature in great depth, she asked extensive questions that required the program director to explain topics such as various leadership theories and the logic underlying the MSA/PMC course sequence. The program director's lack of background in evaluation and his expansive ideas about what he wanted to know required the evaluator to help define and reinforce limits in the scope of the evaluation.

As we worked through many discussions and consulted MSA faculty, potential intended evaluation purposes emerged. Was the work of developing and teaching the new program worth the effort? Would the future program completers be ethical change agents in their schools? What elements of the program were working and which were not? In turn, these generalized questions led to more specific questions, such as how do program participants' roles change within their social networks during the program? In addition to questions about linking unique

aspects of the program to impact on participants and their schools, we also retained the more common concerns about institutional, state, and national accreditation and review (Orr, 2009). Eventually the scope of the program was narrowed down to evaluation of three distinct purposes:

1. Document the implementation of the revised MSA/PMC programs, including elements that were implemented as intended and those that evolved or changed to meet unforeseen needs.
2. Evaluate the growth of students as leaders *while in the program*.
3. Evaluate the growth of students *after successful completion of the program*.

For clarity, the program evaluation is described as a linear process. In reality, the process was less linear and more iterative.

Program Theory and Logic Model

A firm understanding of program theory was an essential foundation for the evaluation. The MSA and PMC programs drew from two different theoretical models. One mirrored the new state standards for school administrators and included traditional theories of hierarchical leadership and management, such as standards on human resource leadership and managerial leadership (North Carolina Department of Public Instruction [NCDPI], 2006). The program faculty also created a set of standards (Buskey & Topolka-Jorissen, 2010) inspired by theories in ethical leadership, professional learning communities, and organizational change. These program standards included a focus on K-12 students; servant leadership, regardless of hierarchical position; change as a complex and ongoing opportunity; ethics as a foundation for action; and, continuous self-improvement. Elements of the theory had been articulated during the program redesign, and the program theory was thoroughly reflected in the design of courses. However, the theory was not yet reflected in the evaluation.

Multiple conversations were held about the program during which the evaluator asked the program director questions to elicit information about how the faculty theorized the program would operate. We were able to articulate what was intended during the program and what was hypothesized to happen to students during and after the program. With the program's focus on enacting ethical leadership (Gross, 2006; Hurley, 2009; Starratt, 1991; Starratt, 2004), the program theory incorporated elements of the reciprocal interactions between candidates and their environments (Orr, 2009). Based on the way leadership was conceptualized in the program, we also captured the idea that various leadership

traits and behaviors would grow at different times. Through those conversations, we created a logic model (Knowlton & Phillips, 2009) (see Figure 1) to illustrate the program theory. We validated the model with the rest of the program faculty. The model then served as the foundation for developing process and outcome evaluation questions and methods.

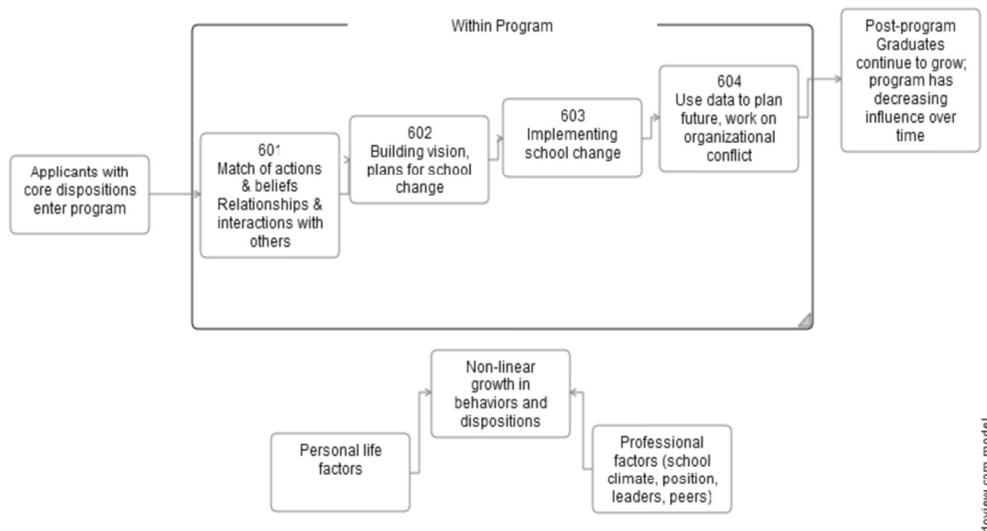


Figure 1. Logic Model for the MSA/PMC Programs.

Program Processes

Evaluation of program processes allows us to monitor several aspects of program design and delivery. For example, this component of the evaluation plan helps us evaluate curriculum alignment, monitor and adjust admission standards, meet documentation needs for accreditation purposes, and track resource use and needs. This information is used formatively, for continuous program improvement. Process evaluation data are also used to support interpretation of program outcome data. For example, if students demonstrate strong leadership outcomes in some areas but weaknesses in others, we can re-evaluate how certain aspects of leadership are taught in the program and redesign instruction accordingly. There are currently three key sources of process evaluation data:

1. *Program documents.* These include but are not limited to: program meeting minutes, curriculum charts, syllabi, and the program's

document submitted to NCDPI and the UNC System General Administration for approval. These documents are not formally evaluated in isolation; they are used primarily to help interpret outcome data and document the new program as it inevitably evolves over time.

2. *A comprehensive database* that combines student admission and academic records maintained by the university and a local database that contains information on changes in enrollment, life events reported by program participants, and cohort membership. This database allows faculty to track student trends across cohorts or by program (MSA and PMC). Much of the information will also be useful for future accreditation purposes. Elements of this database will eventually be linked to student outcome data through unique identifiers to examine outcomes by subgroup, student background variables, and measures of program exposure or progress.
3. *A program readiness survey* designed to be administered to applicants prior to program admission. Because the MSA/PMC programs are unique in their content, structure, and delivery model, the faculty noted that some students entered the program unprepared for the workload and faculty expectations and quickly dropped out of the program. The program readiness survey helps faculty determine whether applicants understand the program's expectations and requirements. In conjunction with the comprehensive database, the results will be used to evaluate patterns and causes of program attrition.

Outcome Evaluation

With our own questions about the program mirroring the national focus on demonstrated program outcomes, the outcome evaluation component was one of the richest and potentially most rewarding parts of the plan. However, it also had the greatest potential to be unwieldy.

We had two strategies for making the outcome evaluation plan more manageable. First, we identified two sampling strategies. Some data sources would be collected from all students and graduates, while more intensive data collection methods would be used with a small subset purposefully recruited for case studies. Second, although we initially developed a broad outcome evaluation plan, the detailed instrument development and data collection processes were separated into two phases. The first phase focused on gathering data from the larger sample during their enrollment in the program. In phase one, data sources being collected for course requirements were reviewed. Gaps were identified, new instruments

were developed to fill those gaps, and a schedule for data collection was developed.

Data sources. Program faculty recommended that the evaluation use as many existing sources as possible and avoid putting undue burden on faculty or participants. We aligned evidence to specific evaluation purposes. We made critical decisions about what to use, what not to use, and what might be usable if the format was modified. As illustrated in Figure 2, we looked for potential areas of overlap between existing course assessments, data already collected for accreditation purposes, and data sources that would answer our evaluation questions. Using the program theory, we conceptualized leadership outcomes in several ways:

- Changes in both formal and informal leadership positions and roles.
- Changes in self-described leadership characteristics aligned to state and program standards.
- Changes in specific leadership actions.
- Changes in the reciprocal relationship between leaders and their environments.

Based on the program theory, some changes were anticipated during the program, while others were expected to occur after students completed the program. Instruments were identified, adapted, or designed to meet each of these definitions. Figure 3 summarizes the instruments and timeline for both phases of outcome data collection.

Data Source	Purpose	Point in program (or years post program)																
		Pre	Semesters In				Years Post											
			1 st	2 nd	3 rd	4 th	1	3	5	10								
Demographic Survey	Description of current employment and responsibilities; background variables to interpret outcomes	X				X				X			X					X
Comprehensive Leadership Survey	Leadership behaviors aligned to program standards Leader influences on environment Environment influences on leader	X				X												X
21 st Century Standards of School Executives Rubric ^a	Leadership behaviors aligned with state leadership standards	Self				Self				Self, Prin, Fac			Self				Self	
End-of-course reflections in four core courses	Optional data for triangulation and case studies		X			X				X								
Action Research Project Impacting Student Learning	Optional data for triangulation and case studies		X			X				X								
360° Feedback	Document strengths and areas of growth, including characteristics not captured in program or state standards	X								X			CSS				CSS	CSS
Social Network Analysis	Determine how participants' influence changes	CSS	CSS							CSS			CSS				CSS	CSS
Interviews of completers, supervisors, and colleagues	Corroborate quantitative data; explain why certain data patterns emerge; obtain richer explanations of long-term program outcomes												CSS				CSS	CSS

Note. Shaded cells represent phase two of evaluation (beginning spring 2011). CSS = Case Study Sample. ^a Includes CSS.

Figure 3. Outcome Evaluation Data Sources and Timeline

Item construction. Within the first phase of data collection, we developed a demographic questionnaire and a leadership survey. The demographic survey served several purposes, two of which were relevant for the outcome evaluation. First, it included items about current job title and leadership responsibilities outside of what might be reflected in the job title. Second, it incorporated background characteristics (e.g., career goals, reasons for pursuing degree, first generation college student status) that might be used to analyze subgroup differences in outcomes.

Because we wanted to measure outcomes of a leadership program, one central question we addressed was how to operationally define leadership. Although a number of published leadership measures exist, such as the School Leadership Preparation and Practice Survey (SLPPS), the NASSP Assessment Center, and VAL-Ed (Orr, Young, & Rorrer, 2010), instruments for our evaluation were locally developed for several reasons. First, most published instruments measure only a limited range of the leadership skills we expect of our graduates (e.g., Interstate School Leaders Licensure Consortium [ISLLC] standards; Council of Chief State School Officers, 2008) and assume leaders are in specific hierarchical roles (i.e., principalship). In contrast, our program emphasizes the application of ethics and caring to a broad range of leadership practices regardless of position. A secondary concern was the cost of published instruments.

We decided to use the two sets of standards that underlie the program's definition of leadership for reference points in defining leadership characteristics we wished to measure. North Carolina's 21st Century Standards for School Executives (NCDPI, 2006) consists of seven standards which are typically assessed on a four-point rubric. NCDPI's standards are aligned to ISLLC standards. To overcome social desirability problems with self-ratings on the existing rubric, we converted the anchors within each domain to checklists and removed information that would cue the reviewer to choose certain response options.

We also have a set of program-specific standards that are less defined but that form the basis for many aspects of the program structure and instructional practices. Features of the program theory specific to ethical leadership largely drove items construction. For example, the program emphasizes an action orientation and a student focus. Thus, several items ask respondents to report the frequency and conditions in which they have advocated for students within the past year.

The first full draft of the leadership survey incorporates items related to formal leadership roles, specific leadership behaviors, roles played in situations that required leadership, and factors that facilitate or inhibit respondents' abilities to enact ethical leadership in their schools. Specific items are from a combination of

program curricular elements, faculty input on likely response options, and current student feedback, with items designed using principles described by Dillman, Smythe, & Christian, (2008). Analysis methods are planned to investigate the quality of these measures, and revisions may be made depending on the outcomes of those analyses. Samples of the various items are provided in Table 1.

Table 1
Sample Outcome Assessment Items

Type of Outcome	Sample Item
Formal leadership role	<p>What is your current position (or, your most recent position if you are not currently employed)?</p> <ul style="list-style-type: none"> • Classroom teacher • Assistant principal • Other school staff (guidance counselor, school psychologist, etc.) • Central office staff • NC principal fellow • Other (describe):
Informal leadership responsibility	<p>Which of the following leadership responsibilities/roles fall within your current job responsibilities, even if you don't hold the formal title? Select all that apply.</p> <ul style="list-style-type: none"> • School Improvement Team leader • Grade level coordinator / Department head • Teaching team leader • Student services / student support team leader • Member of ad-hoc committee (e.g., curriculum or textbook review, policy review) • PLC group leader • Other committee leader • Formal (assigned) mentor • Informal mentor • Head athletic coach • Other (describe):
Self-rated leadership traits – NC standards	<p>Please select from the list below all items that that reflect behaviors or characteristics the candidate <i>currently possesses</i>.</p> <ul style="list-style-type: none"> • Understands the attributes, characteristics, and importance of school vision, mission, and strategic goals; and can apply this understanding to the analysis and critique of existing school plans. • Develops his/her own vision of the changing world in the 21st century that schools are preparing children to enter. • Works with others to develop a shared vision and strategic goals for student achievement that reflect high expectations for students and staff. • Maintains a focus on the vision and strategic goals throughout the school year. • Designs and implements collaborative processes to collect and analyze data, from the <i>North Carolina Teacher Working Conditions Survey</i> and other data sources, about the school's progress for the periodic review and revision of the school's vision, mission, and

	strategic goals.
Specific ethical leadership behaviors	<p>In the past 12 months, I enacted ethical leadership in the following areas:</p> <p>Response options: Not at all, once in the past year, quarterly, monthly, bi-weekly, weekly, daily</p> <ul style="list-style-type: none"> • Advocated for a single student or small group of students • Advocated for a large group of students • Advocated for all students school-wide • Worked with another individual on curricular or pedagogical initiative to improve learning
Change in leader-environment interaction	<p>Sometimes there are things in our work environments that facilitate our leadership behavior. Other things may be barriers to our leadership. To what extent do each of the following represent <u>facilitators or barriers</u> to your leadership actions?</p> <p>Response scale: -5 (significant barrier) to +5 (significant facilitator)</p> <ul style="list-style-type: none"> • Holding a formal position with power • Being recognized as a leader in the school (regardless of position) • Having input during decision-making processes • Working in a school where the culture promotes collaboration and cooperation • Not having to fear negative consequences for speaking my mind • Having formal school leaders who are open to change

Implementation. We decided to phase in implementation of the evaluation over a 12-month period due to several factors. First, the pilot group (cohort 1) would finish the program at the end of spring 2010, long before we could develop the entire evaluation system. We wanted to capture completion data from each cohort, but also wanted some basis for preliminary comparison of pre and post program measures sooner than the two years it would take a cohort to begin and end the program. Thus, we started with cross-sectional analyses based on data collected from several cohorts in spring and summer 2010. Longitudinal analysis began with cohort 6, admitted in summer 2010 and projected to graduate in spring 2012.

Recommendations and Conclusions

Designing a comprehensive program evaluation system is a complex process that should engage a variety of experts and stakeholders. Steps within the process must include identifying collaborators, determining specific purposes of evaluation, examining existing data sources, determining gaps in evaluation components, developing a plan to phase in a comprehensive system, and constructing valid instruments.

Through our experience in designing this system, we offer the following recommendations for others who wish to design evaluations of leadership preparation programs that are tailored to their unique program features:

1. Know the program. Program faculty may have long ago internalized their understanding of how the program operates. Or, maybe the program faculty never had conversations about how they believe the program operates. In either case, the faculty will need to explicate the details of how the program theoretically works, how the curriculum makes that happen, and what the students look like during and after the program.
2. Understand the needs and concerns of the constituents. The program director and faculty are not the sole consumers of the data from this evaluation system. We made adjustments to the design based on input from within the program. Further refinements may come from feedback offered from outside the institution.
3. Define what you need and want to know. It is easy to jump ahead into thinking about data sources or what is easy to collect without being careful about establishing purposes. A clear sense of purpose guides the remaining design and analysis steps. It prevents the data sources from mismatching the purposes. It also prevents the misuse of unneeded resources.
4. Give attention to the methodology – down to the details. Recognize that you do not have to collect “everything” if your evaluation purposes are not that broad. Capitalize on the data sources you already collect. Attend to the balance between quality and feasibility. Identify sources of expertise which can help you develop high-quality methods given your resource constraints.
5. Look before you leap. Pilot your instruments. Roll out the data collection process in phases. Ask for the perspectives of your current students. Establish systems for managing the recruitment, consent, data collection, and data analysis phases.

The unique collaboration between the co-authors was grounded in a friendship of several years, which made the evaluation design process easier in some ways. There was value in sharing the two perspectives, and we each initially underestimated the contributions of our own expertise. If principal licensure faculty who are in other programs do not have such existing relationships with evaluators, they may wish to look to resources on working with external evaluators (e.g., Kauffman, Guerra, & Platt, 2006). Evaluators have to feel

comfortable asking a lot of questions to understand the program and the evaluation needs. Program faculty need to be comfortable understanding why the questions are being asked, answering those questions, and not worry about having limited backgrounds in evaluation.

Although the evaluation system described here does provide for evaluation of several levels of outcomes, it stops short of evaluating the link between leader behaviors and K-12 student outcomes. Once phase two of data collection has started, we likely will return to this student outcome question and see how it fits within our evaluation framework. This link has several interim steps and it is challenging to create strong causal links for such tertiary outcomes (Leithwood, Patten, & Jantzi, 2010; Orr, 2009), especially with a relatively small data set and limited resources for tracking graduates after degree completion.

References

- Blackwell, C., Cummins, R., Christine, D., Townsend, C. D., & Cummings, S. (2007). Assessing perceived student leadership skill development in an academic leadership development program. *Journal of Leadership Education*, 6(1), 39-58. Retrieved from http://www.fhsu.edu/jole/issues/archive_index.html
- Buskey, F. C. (2010). Turbulence and transformation: One professor's journey into online learning. In V. Yuzer & G. Kurubaçak (Eds.), *Transformative Learning and Online Education: Aesthetics, Dimensions and Concepts*. Hershey, PA: IGI Global.
- Buskey, F. C., & Jacobs, J. E. (2009). Catalysts for redesigning principal preparation: A courageous journey. *International Journal of Educational Leadership Preparation*, 4(2). Retrieved from <http://ijelp.expressacademic.org>
- Buskey, F. C., & Topolka-Jorissen, K. (2010). No more silos: A new vision for principal preparation. *Educational Leadership Review*, 11(2), 112-122.
- Council of Chief State School Officers. (2008). *Educational leadership policy standards: ISSLC 2008*. Washington, DC: Author.
- Dillman, D. A., Smythe, J. D., & Christian, L. M. (2008). *Internet, mail and mixed-mode surveys: The tailored design method* (3rd ed.). New York: John Wiley & Sons.
- Duncan, A. (2010, February 19). *Preparing the teachers and school leaders of tomorrow*. Remarks at the American Association of Colleges for Teacher Education Conference, Washington, DC. Retrieved from <http://www.ed.gov/news/speeches/preparing-teachers-and-school-leaders-tomorrow-secretary-arne-duncans-remarks-american>
- Fry, B., O'Neil, K., & Bottoms, G. (March 2006). *Schools can't wait: Accelerating the redesign of university principal preparation programs*. Atlanta, GA: The Southern Regional Education Board.
- Gross, S. J. (2006). (Re-)Constructing a movement for social justice in our profession. *International Electronic Journal for Leadership in Learning*, 10. Retrieved from <http://www.ucalgary.ca/iejll/vol10/gross>
-

- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press.
- Hoy, W. K., & Miskel, C. G. (2008). *Educational administration: Theory, research, and practice*, (8th ed.). New York: McGraw-Hill.
- Hurley, J. C. (2009). *The six virtues of the educated person: Helping kids to learn, schools to succeed*. Lanham, MD: Rowman & Littlefield Education.
- Kaufman, R., Guerra, I., & Platt, W. A. (2006). Contracting for evaluation services. In *Practical evaluation for educators: Finding what works and what doesn't* (pp. 253–263). Thousand Oaks, CA: Corwin Press.
- Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Berrett-Koehler.
- Knowlton, L. W., & Phillips, C. C. (2009). *The logic model guidebook: Better strategies for great results*. Thousand Oaks, CA: Sage Publications.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Educational Administration Quarterly*, 46, 671-706. doi:10.1177/0013161X10377347
- North Carolina Department of Public Instruction (2006). *North Carolina standards for school executives*. Raleigh, NC: Author. Retrieved from <http://www.dpi.state.nc.us/docs/profdev/guidelines/standards.pdf>
- Orr, M. T., with Barber, M. E. (2009). Program evaluation in leadership preparation and related fields. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 457-498). New York: Routledge.
- 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, 18-70. doi: 10.1177/0011000010378615

- Orr, M. T., Young, M. D., & Rorrer, A. K. (2010). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership programs*. Salt Lake City, UT: The National Center for the Evaluation of Educational Leadership Preparation and Practice. Retrieved from <http://edleaderprepucea.squarespace.com/storage/Developing%20Evaluation%20Evidence.pdf>
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Schneier, C. E., Beatty, R. W., & Russell, C. J. (1994). *The training and development sourcebook* (2nd ed.). Amherst, MA: Human Resource Development Press.
- Smith, D. N., & Roebuck, D. B. (2010). Interviews: Linking theory to practice. *Journal of Leadership Education*, 9(2), 135-143. Retrieved from http://www.fhsu.edu/jole/issues/archive_index.html
- Starratt, R. J. (1991). Building an ethical school: A theory for practice in educational leadership. *Educational Administration Quarterly*, 27, 185-202. doi: 10.1177/0013161X91027002005
- Starratt, R. J. (2004). *Ethical leadership*. San Francisco: Jossey-Bass.
- Walker, W. G. (1993). Theory and practice in educational administration. *Journal of Educational Administration*, 3(1), 18-43.
- Weiss, C. H. (1998). *Evaluation* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.