

The Tension between Accountability and Formatively: Implications for Educational Planning

Steve Myran and Jennifer K. Clayton

ABSTRACT

Today, educators find themselves at the nexus of accountability and improvement (Earl & Fullan, 2003), contending with competing pressures to chase accountability measures and obtain short term gains, while at the same time embracing the principles of formatively (Erickson, 2007) that can promote sustained growth. While the pressures of accountability are often experienced as a push to meet standards without consideration for investments in the instructional core, the use of formative tools such as data-based decision making and formative program evaluation are geared toward iterative growth informed by ongoing data collection, analysis, and action. In our experience, working on both sides of these questions, we have found that the academy and the public schools lack much of the institutional knowledge needed to overcome these tensions. We assert that if these competing pressures are allowed to go unresolved, they will undermine educators' ability to focus on deep, substantive learning and to use assessment information to formatively point the way to program improvements. In order to better understand the sources of these tensions, we provide a synopsis of the history of program evaluation and data-based decision making in education as well as a review of our own experience with these pressures in our work with the US Doe School Leadership Program grant. Lastly, we draw a number of conclusions about the importance of these issues for educational planners and the need to understand and carefully consider these competing pressures in order to avoid potential unintended consequences.

Today, educators find themselves at the nexus of accountability and improvement (Earl & Fullan, 2003), contending with competing pressures to chase accountability measures and obtain short term gains, while at the same time embracing the principles of formativity (Erickson, 2007) present in data-based and evidence-based school improvement efforts. In this environment, dominated by the pressures of accountability and testing, both educators and program evaluators can find themselves bending to these pressures in ways that often undermine the intent of school and program improvement efforts. Curricular alignment, teaching to the test, and what some have called “black box evaluation” (Lipsey, 1987; Rossi, Freeman, & Lipsey, 1999) can undermine improvement efforts by focusing too narrowly on complying with outcome measures and overlooking what goes on inside the program itself.

Without understanding what actually is going on inside a school improvement effort, there is no way of making adjustments and refinements. In this way, the black box approach to understanding school improvement, where the system is only viewed in terms of its inputs and outputs, fails to make improvements based on information and feedback collected along the way. As Erickson (2007) points out, one of the reasons for this disconnect between accountability and the principles of formativity is that these two activities take place “on radically differing time scales” (p. 186). Accountability information is almost entirely summative and cannot inform practice in timely and responsive ways, while short-term formative information can point the way to improvements and refinements while an effort is still underway. Formative program improvement efforts do not need to wait until the final output measures are collected to understand where improvements can be made and respond in proactive and purposeful ways.

As Black and William (2005) point out, the use of summative assessments dominates the educational landscape and further highlight that interim and large-scale assessments are notable obstacles that can undermine the use of formative assessment tools. Reacting to a tendency for assessments to be used in largely summative ways, scholars and practitioners have pioneered the development and use of formative assessment strategies, which are designed to provide all stakeholders with feedback that can be used to make ongoing improvements. In the public school setting, these efforts fall under two broad categories; formative assessment, or assessment for learning (Stiggins, 1999) and data-based decision making. In the program evaluation realm, a number of formative evaluation models have been proposed going back to Scriven's (1967) original distinction between summative and formative assessment. These include *Utilization Focused Evaluation* (Patton, 1978), and the Context, Input, Process, and Product model

(Stufflebeam, 1971). Many accreditation agencies, including NCATE, have begun to more rigorously fold these formative principles into their reporting requirements focusing more heavily on evidence of continuous improvement than on compliance and outcome measures.

PURPOSE

With the above in mind, we assert that these are competing pressures, one to chase the measure to obtain short term gains on largely summative assessment instruments, and the other to embrace the principles of formativity. If these tensions at the nexus of accountability and improvement go unresolved, it will undermine educators' ability to better understand the impact of instruction utilized in schools and classrooms. Furthermore, it will undermine our ability to use these insights to better facilitate deep, substantive, and lasting learning opportunities for our students. In this paper, we provide a brief summary of the history and theories behind program evaluation and formative assessment with an eye towards what we describe as the principles of formativity. We go on to briefly describe our ongoing work on a program evaluation and assess these efforts in relationship to the principles of formativity. Lastly, we use the discussion of our own work in this area as a tool to address the theory to practice implications that educational planners should carefully consider in order to avoid the many potential unintended consequences of a failure to balance the tensions between accountability and school improvement.

THE TENSION BETWEEN ACCOUNTABILITY AND FORMATIVITY

History of Program Evaluation

The notion of program evaluation for educational systems and improvement projects is one with a long history of transformation and development. Program evaluations were originally intended to provide a mechanism for documentation and comparison and were used widely in government programs prior to their emergence in the field of education. Since that time, program evaluations have gone through several iterations and have often been tied to a compliance and accreditation process. Earlier evaluation efforts, which began in earnest in Great Britain in the 1800s, were designed as a mechanism for measuring student performance and then drawing comparisons between schools or systems of education (Fitzpatrick, Sanders, & Worthen, 2004). This appealed to political officials and educational theorists and during the 1950s and 1960s, schools became increasingly invested in improving and expanding the use of standardized testing as a mechanism for connecting objectives to learning outcomes. During the 1960s and 1970s, theorists such as Alkin, Stufflebeam, and Scriven advocated for more systemic models that could provide a methodical and rigorous process and be usable by educators. Stufflebeam (1971) promoted the CIPP model which examined the context, input, process, and product as a model for program effectiveness and later Guba and Lincoln (1981) expanded that work to examine naturalistic evaluation as a method of program evaluation. These models provided new frameworks that helped program evaluators begin to escape the confines of the "black box evaluation."

Traditionally, program evaluation has been shaped by an "expert" model and limited to the technical activities of data collection and analysis. Typically, under such a model, evaluators are brought in and take direction from the principal investigators, conduct the study, and report their findings to the PI's who in turn take action as they see fit and/or to the degree that the grantors require. This "black box evaluation" (Lipsey, 1987; Rossi, Freeman, & Lipsey, 1999, p. 154) and "technical activities evaluation" are problematic in that they focus on what goes into and what comes out of a program without considering what goes on inside a program, or "focus on the overall relationship between the inputs and outputs of a program without concern for the transformation process in the middle" (Chen, 1990, p. 18), neglecting that stakeholders may benefit more from what happens in the middle than the inputs and outputs of a more conventional program evaluation.

At a fairly coarse grain, Scriven (1967) distinguishes between formative and summative evaluation, while at a finer grain, others have distinguished between internal and external evaluations (Shulha & Cousins 1997; Wadsworth 2001), rational approaches (Tyler, 1942, 1966; Provus 1971; Steinmetz 1983) and constructivist approaches (Dryzek 1993; Fischer & Forester 1993; Guba & Lincoln, 1989; Majone, 1989; van der Knaap, 1995). Stufflebeam and Shinkfield (2007) identified twenty-six evaluation

approaches grouped into five categories: 1) pseudoevaluations, 2) questions and methods-oriented evaluation approaches, 3) improvement and accountability-oriented evaluation approaches, 4) social agenda and advocacy approaches, and 5) eclectic evaluation approaches (p. 137).

A Shift in Views of Program Evaluation

An important part of this history is the tension between the summative and formative aspects of program evaluation. There has been an increasing shift towards more iterative and formative means of collecting and reporting data on a program's implementation and progress. The U.S. Department of Education in 2004 indicated a "significant shift" in the process and use of program evaluation in department and department-funded projects. The new policy statement indicated, "We propose a significant shift in program evaluation, away from a compliance model and towards a system of research and evaluation focused on results and the effectiveness of specific educational interventions" (US DOE, 2004). Prior to that statement, there was a movement toward this new research approach from researchers such as Rossi, Freeman, and Lipsey (1999) who defined program evaluation as the use of social research methods to systematically investigate the effectiveness of social intervention programs. It draws on the techniques and concepts of social science disciplines and is intended to be useful for improving programs and informing social action. (p. 35)

A practical reality of program evaluation is that there are challenges and limitations to using the data in the most effective manner. Identifying these challenges often comes from understanding the experiences of stakeholders, project directors, and program evaluators in the field and navigating the issues. By making such a change to the overall structure and purpose of program evaluation, educators engaged in projects that require formal performance reports continue to struggle to combine compliance and reporting requirements with a need to embed such requirements into ongoing formative evaluation and improvement efforts.

Inclusion, Dialogue, and Deliberation: A Conceptual Framework

According to Stufflebeam and Shinkfield (2007), "inclusion, dialogue, and deliberation are considered relevant at all stages of an evaluation: inception, design, implementation, analysis, synthesis, write-up, presentation, and discussion" (p.222). Inclusion involves evaluators taking into consideration the interests, values, and views of major stakeholders involved in the program or policy under review (House & Howe, 2000). Dialogue, the second of the three principles emphasizes the importance of extensive dialogue with stakeholder groups being cautious not to organize merely symbolic interactions, but to promote honest dialogue. Lastly, deliberation should provide opportunities to involve stakeholders in weighing evidence and drawing sound conclusions. The evaluator is responsible for structuring the deliberation thereby helping to draw out valid conclusions (House & Howe, 2000). In this way, effective program evaluation can be considered that which combines the issues of context, varied data sources and analysis techniques, and extensive uses of qualitative data along with the careful consideration of communication with stakeholders groups as outlined by Stufflebeam and Shinkfield (2007).

FRAMING THIS TENSION IN CONTEXT: THE FUTURES PROGRAM EVALUATION

Background of the School/University Partnership

The authors' involvement in school/university partnership efforts provided an ideal test bed to examine the tensions between accountability and formativity and better understand the challenges in practical terms and how program evaluation can better promote school improvement. Designed to build the leadership capacity within the context of the school culture being served, the Futures Program (pseudonym) is a partnership between a small rural public school district and a large regional urban university. This program was intended to stabilize and strengthen the retention of school leaders who can successfully guide and direct instruction in this high-need LEA, ensuring a leadership pipeline of those who possess the institutional knowledge of the school division, balanced with the contemporary models of school leadership. The Futures Program was built around an integrated effort that focuses on holistic approaches to leadership preparation, developing relationships, coordinating meaningful professional development,

realism in design and experiences, and introspection as ways for participants to build stronger bridges between theory and practice (Sanzo, Myran, & Clayton, 2011).

This program was required by the funder to provide annual progress reports that focused primarily on outcomes. For example, the report was primarily a quantitative assessment of numbers of graduates, numbers of hires, and division student growth. This reporting requirement largely drove the program evaluation efforts in the early years of the project. In later years, the project directors and program evaluator made a conscious decision to focus on the formative program evaluation feedback mechanisms, making the annual report a smaller facet of the overall evaluation model. This transition provides an opportunity to examine this program evaluation as one that struggled with the tension between accountability and formativity.

Through the Lens of Inclusion, Dialogue, and Deliberation

The integration of holistic approaches to leadership preparation, developing relationships, and coordinating meaningful professional development were important aspects of the Futures Program and were developed and refined as a result of the continuous fine-tuning of the formative aspects of the evaluation model (Sanzo, Myran, & Clayton, 2011). The relationships between the diverse partners involved in the Futures Program were complex and the potential for inequitable distributions of power (real or perceived) had the capacity to trigger oversights and lead to mistakes in planning, implementation, and evaluation (Miller & Hafner, 2008). We found that the evaluation principles of inclusion, dialogue, and deliberation were essential components of developing an approach for working with future school leaders that built stronger bridges between theory and practice. These three features, inclusion, dialogue, and deliberation, we assert, represent a critical lens for how program evaluation is planned and implemented.

Through the Lens of Inclusion

When we assessed our experience with the Futures Program evaluation through the lens of inclusion, we found that including all stakeholders in the early stages of planning is extremely difficult and is not generally built into the organizational and institutional norms of the cooperating schools, nor the university. Typically, grants and projects are developed by key decision makers with the practitioners brought in at the stage of implementation. While student learning or other outcomes are always considered goals of the grant or program, students are not, however, considered as stakeholders.

We determined that the program evaluation in the first two years of the grant failed to include the voice of the various stakeholders within the district and university. While not intentional, this oversight narrowed the scope of the program evaluation to only include the perspective of the project directors and participants, while not including the input of university administration, district administration, teachers, parents, and students. One formative change made to the program evaluation model in year three is an ongoing effort to develop instruments and build trust with these groups to initiate their involvement in providing data to the evaluative piece.

Through the Lens of Dialogue

Regarding the principle of dialogue, we found that communication among the various stakeholders, including that of the evaluators and the program directors is critical. However, inclusion was a prerequisite to dialogue. Without the various stakeholders' authentic inclusion in the various aspects of the program, there limited content for meaningful and productive dialogue. This may go without saying, but it highlights the interrelated nature of inclusion and dialogue. Specifically, the mutual understanding about the nature and purpose of the program and the evaluation, as well as clear communication about roles and responsibilities was found to be crucial. We found that in the first two years of the project, the lack of clarity on these issues limited the evaluator-project directors' interaction as well as the evaluator, stakeholder interactions creating a kind of role-limitation thereby impoverishing the quality and accuracy of the evaluation and program itself.

In contrast to these earlier failures of inclusion and dialogue, in our third year of the project with a new evaluation, we found that building mutual trust and understanding about the nature of evalua-

tion, particularly the formative purposes of evaluation, resulted in refinement and implementation of the evaluation. We also found that when stakeholder feedback was included, power imbalances often associated with stakeholder disenfranchisement were mitigated. For example, when teachers are asked for feedback, but do not see their concerns addressed or represented in program activities, their motivation can be undermined causing a decrease in participation.

Through the Lens of Deliberation

Not surprisingly, we found deliberation to be the critical kingpin of the minimizing the tensions between accountability and formativity. In the same way that inclusion was found to be a prerequisite of dialogue, we found that without inclusion and dialogue, there would be little substance about which to deliberate. In years one and two of the Futures Program evaluation, there was limited inclusion and associated dialogue, hence, there was far less substantive program information to deliberate and make formative improvements from. While there were data being collected that met the technical requirements of the funder, these data did not represent robust patterns of inclusion, dialogue, and deliberation. Instead, we found that we were meeting the basic criteria of the funded proposal, but lacked the more authentic, iterative, and meaningful formative data that would have promoted ongoing understanding of the program's strengths and weaknesses pointing the way to program improvements.

We also found that the format of deliberation needs to vary in response to the nature of the issues being addressed. For example, in the assessment of the increases of cohort members' leadership disposition, deliberation needed to be done carefully as not to expose them to a threatening environment or to make them feel vulnerable. The climate created to encourage deliberation must be consistent with a safe space to openly dialogue.

Implications for Educational Planning

The shift from evaluation models that emphasized compliance to outcome measures to those that capitalize on formative feedback loops about improvement is one that needs to be well understood by educational planners. While there has been a lack of agreement among scholars about just what educational planning it is, Farrell (1997) does outline a broad definition, drawn from Anderson, Brady & Bullock's (1978) "policy cycle", that is useful for considering the implications of the nexus of accountability and improvement for educational planning. This definition has six components:

1. Identification of a social phenomenon as a policy problem,
2. Placement of the problem high on political agendas,
3. Identification and evaluation of a range of possible "solutions,"
4. Selection of one solution (the policy),
5. Implementation, and
6. Evaluation, feeding back into the cycle wherever appropriate (p. 280).

The Farrell model captures a broad definition of education planning which includes policy, process, administration, and management and provides a framework for considering the relationships between educational planning and the tensions between formativity and accountability.

In many respects, this definition of educational planning is formative in nature, emphasizing the identification of goals and needs, implementation and forward feedback cycles about program improvement. We can see the parallels of the Farrell definition to formative assessment concepts as outlined by Chappuis and Chappuis (2008). Here, formative assessment can be understood as addressing questions about strengths relative to a goal, observations of improvements, identification of weaknesses, areas of effective performance, and how one might build on these strengths and what might be done differently or better in the next iteration of learning or improvement and how one might prepare for that improvement. Because educational planning, as broadly outlined by Farrell (1997), has important formative elements, the current shifts from compliance to formatively oriented evaluation and assessment are well supported by and aligned with aspects of planning. Moreover, because of this natural alignment, educational planning is a natural ally and supporter of the principles of formativity.

Similarly, the key elements of inclusion, dialogue, and deliberation as outlined by Stufflebeam and Shinkfield (2007), provide an additional way of thinking about the implications for educational planning.

Using the Farrell (1997) definition of educational planning, we can better frame our own observations about the tensions between accountability and the principles of formativity. In the section that follows, we use this framework to link our experiences working within the tension between accountability and program improvement to the field of educational planning.

Given the centrality of accountability in today's educational climate, it is not hard to see accountability and the pressures there within, as a social phenomenon and policy problem. If policy problems are essentially a social phenomenon, they need to be understood through the perspectives of stakeholders that the effort will impact the most, as well as those who will influence the process. Along these lines, we found that including all stakeholders in the early stages of planning is extremely difficult and is not generally built into the organizational and institutional norms of the cooperating schools, nor the university. The initial oversight to include the voice of the various stakeholders within the district and university narrowed the scope of the first two years of the program evaluation to only include the perspective of the project directors and participants.

By analyzing the challenges we faced at the nexus of accountability and school improvement as a social phenomenon and policy problem, we identified a notable barrier to growth. That is the lack of organizational and institutional capacity to effectively include all stakeholders in dialogue and deliberation in the planning, implementation, and formative program improvements. We have increasingly recognized that the communications and coordination issues between public schools and universities is immensely complex and multidimensional and as such, there is a great need for better understanding of how these very different organizations can effectively plan, implement, and sustain more effective school improvement efforts.

We know from the literature the promise of better planning and implementation among school and university partnerships (Essex, 2001; Goodlad, 1991), as well as the importance of clearly defined purpose and direction, supported both with active participation and adequate resources by top leaders, trust among partners, open communication, mutual respect among partners, mechanisms to assess progress and measure outcomes, true collaboration, and school-wide representation at the beginning of an effort (Essex, 2001; Peel, Peel & Baker, 2002; Rakow & Robinson, 1997). What we found in our collaborations was consistent with Teitel's (2003) observation that, many times, there is a communication issue between entities, the university and the schools. We also found that the K-12 partners were sometimes burdened with the presumption that the university was there to "fix" public schools (LePage, Bordreau, Maier, Robinson, & Cox, 2001), and consistent with Clarken's (1999) research, that this was created across the two entities and this presumption was found to be hard to correct.

Similarly, the failure to adequately include stakeholder voice in the improvement effort could also significantly impact the placement of the problem high on political agendas given the dominant position of accountability in today's educational climate. Moreover the appropriate identification and evaluation of a range of possible "solutions" represents a key principle of formativity; the identification of strengths relative to a goal, diagnoses of possible areas of instructional weaknesses, and associated instructional improvements. Lastly the selection of one solution (the policy), implementation and evaluation, feeding back into the cycle of improvements represent the most compelling aspect of educational planning as it relates to the principles of formativity.

As we discussed, in the first two years of the project, the lack of clarity on communication limited the interactions among the different project personnel and limited roles which impoverished the quality and accuracy of both evaluation and the program itself. The lack of clarity on communication disregarded the principles of formativity and in turn emphasized compliance, oriented assessment. Overcoming this issue and building mutual trust helped us to refine the program evaluation. Similarly, including stakeholder feedback helped to improve power imbalances that had led to disenfranchising certain stakeholder groups. Lastly, we highlighted that the format of deliberation needs to vary in response to the nature of the issues being addressed in order to not expose stakeholders to a threatening environment or to make them feel vulnerable and thereby risk direct and honest feedback about the improvement effort.

Historically, educational programs have often failed to use program evaluation formatively. Hence the iterative improvements that could be obtained through the formative use of evaluation feedback are not fully utilized. There are a number of reasons for this failure to capitalize on the potential for forma-

tive program evaluation, which includes a lack of clear purpose of program evaluation, communication issues, and sometimes competing goals. Moreover, an imbalanced notion of evaluation with a primary or even exclusive focus on summative assessment is a significant danger to school improvement and student learning. Working harder within the confines of the older notions about program evaluation quickly bumps into the functional and organizational limitations of a narrow or imbalanced notion of assessment and accountability. In the climate of high stakes tests and state-mandated standards, there is an increase of federal, state, and local demand for educators at all levels to be effective at meeting these accountability standards. Educators can help mitigate pressures to only look at inputs and outputs by fully embracing the principles of formativity so clearly present in the field of educational planning that promote more deliberative assessments of what happens inside the black box.

CONCLUSION

Because educators today find themselves at the nexus of accountability and improvement (Earl and Fullan, 2003), educational planners, as we have defined broadly, need to strongly consider the principles of formativity and the role of program evaluation. The notable pressures of accountability and testing can influence both educators and program evaluators to bend to these pressures in ways that can undermine school and program improvement efforts. Specifically related to program evaluation, what has been referred to as the “black box evaluation” can lead to narrowly complying with outcome measures and overlooking the program features that need to be strengthened or eliminated in real time. Without such an understanding, there is no way of making adjustments and refinements, and school programs risk simply replicating the same patterns of weak planning and implementation. Responding to these concerns, many accreditation agencies have updated their data collection and reporting protocols to reflect formative principles.

We maintain that the competing pressures to chase the measure to obtain short term gains while at the same time embracing the principles of formativity has created an untenable situation for educators, and if it goes unresolved, will undermine our ability to facilitate deep, substantive, and lasting learning opportunities for our students. The 2004 shift by the U.S. Department of Education in the process and use of program evaluation marks an important indicator of where the future of program evaluation is going, and one we argue educational planners should be well aware of.

We argued in this article that the Stufflebeam and Shinkfield’s (2007) model of inclusion, dialogue, and deliberation provides a valuable conceptual framework for better understanding the challenges discussed above. We used our own experiences with the Futures Program to illustrate these issues and found that including all stakeholders in the early stages of planning was important, but difficult; the lack of communication networks that are built into the organizational and institutional norms of schools and universities. We also found that inclusion was a prerequisite to communication among stakeholders. Lastly, we found that low stakes deliberation was the kingpin to promoting formative reflection and action and required inclusion and dialogue as critical starting points.

Finally, we drew from Farrell’s (1997) definition of educational planning to outline the links between planning, program evaluation and the principals of formativity. We discussed how the pressures of the accountability movement can be seen as a social phenomenon and policy problem and as such need to be better understood through the perspectives of stakeholders and how this is a needed component for placement of the problem high on political agendas. We also argued that the appropriate identification and evaluation of possible “solutions”; the selection of a solution, implementation, and evaluation that all feed back into the cycle of improvements all represent key principles of formativity. Here we argued that Farrell definition is formative in nature and as such helps to frame educational planning as a key ally in helping to foster the needed current shifts from compliance to formatively oriented evaluation.

REFERENCES

- Anderson, J., Brady, D., & Bullock, C. (1978). *Public policy and politics in America*. Monterey, CA: Brooks Cole.
- Black, P., & William, D. (2005). *Lessons from around the world: How policies, politics and cultures*

- constrain and afford assessment practices. *Curriculum Journal*, 16(2), 249-261.
- Chappuis, S., & Chappuis, J. (2008). The Best Value in Formative Assessment. *Educational Leadership*, 65(4), 14-19.
- Chen, H. (1990). Issues in constructing program theory. *New Directions for Program Evaluation*, 47, 7-18.
- Clarken, R. (1999). *University/School collaboration: A case study*. Presentation at the American Association of Colleges of Teacher Education, Washington DC.
- Dryzek, J. (1993). Reconstructive democratic theory. *American Political Science Review*, 87(1).
- Earl, L. & Fullan, M. (2003). Using data in leadership for learning, *Cambridge Journal of Education*, 33(3).
- Erickson, F. (2007). Some thoughts on “proximal formative assessment of student learning”. *Evidence and Decision Making*. Malden, MA: Blackwell Publishing.
- Essex, N. L. (2001). Effective school-college partnerships, a key to educational renewal and instructional improvement. *Education*, 121 (4), 732-736.
- Farrell, J. P. (1997). A retrospective on educational planning in comparative education, *Comparative Education Review*, 41, 277–313.
- Fischer, F. & Forester, J. (1993). *The argumentative turn in policy analysis and planning*. Chapel Hill, NC: Duke University Press
- Fitzpatrick, J., Sanders, J., & Worthen, B. (2004). *Program evaluation: Alternative approaches and practical guidelines 3rd Edition*. Boston: Pearson Education.
- Goodlad, J. I. (1991). School-university partnerships. *Education Digest*, 56 (8), 58-61.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation*. San Francisco: Josey-Bass.
- Guba, E., & Lincoln, Y. (1989). *Forth generation evaluation*. Newbury Park, CA: Sage.
- House, E. & Howe, K. (2000). Deliberative democratic evaluation, *New Directions for Evaluation*, 85, 3-12.
- LePage, P., Boudreau, S., Maier, S., Robinson, J. & Cox, H. (2001). Exploring the complexities of the relationship between K-12 and college faculty in a nontraditional professional development program. *Teaching & Teacher Education*, 17(2), 195-211.
- Lipsey, A. (1987). *Theory as method: Small theories of treatments*. Paper presented at the National Center for Health Services Research Conference: Strengthening Causal Interpretations of Non-experimental Data, Tucson, AZ
- Majone, G. (1989). *Evidence, argument & persuasion in the policy process*. New Haven: Yale University Press.
- Miller, P. & Hafner, M. (2008). Moving toward dialogical collaboration: A critical examination of a University-School-Community Partnership. *Education Administration Quarterly*, 44(1) 66-110.
- Patton, M. Q. (1978). *Utilization-Focused Evaluation*. Beverly Hills, CA: Sage.
- Peel, H.A., Peel, B.B., & Baker, M.E. (2002). School/university partnerships: A viable model. *The International Journal of Educational Management*, 16 (7), 319-325.
- Provus, M. M. (1971). *Discrepancy evaluation: For educational program improvement*. Berkeley, CA: McCutchan Publishing Company.
- Rakow, S. J. & Robinson, L. W. (1997). School and university partnerships: Teaming for success. *Contemporary Education*, 68(2), 143-7.
- Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach*. 6th ed. Thousand Oaks: Sage.
- Sanzo, K., Myran, S., & Clayton, J. (2011). Building bridges between knowledge and practice: A university-school district leadership preparation program partnership, *Journal of Educational Administration*, 49(3).
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195-208.
- Steinmetz, A. (1983). The discrepancy evaluation model. In *Evaluation Models*, G. F. Madaus, M.

- Scriven, and D. L. Stufflebeam (Eds.), 79-100. Boston: Kluwer-Nijhoff.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191-198.
- Stufflebeam, D.L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education* 5, 19-25.
- Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, & applications*. San Francisco: John Wiley.
- Teitel, L. (2003). *The professional development schools handbook: Starting, sustaining, and assessing partnerships that improve student learning*. Thousand Oaks, CA: Corwin Press.
- Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35,492-501.
- Tyler, R. W. (1966). The objectives and plans for a national assessment of educational progress. *Journal of Educational Measurement*, 3, 1-10.
- U.S. Department of Education (2004). *New directions for program evaluation at the U.S. Department of Education*. Retrieved from http://www2.ed.gov/news/press_releases/2002/04/evaluation.html.
- van der Knaap, P. (1995). Policy evaluation and learning: Feedback, enlightenment or argumentation? *Evaluation*, 1(2): 189-216.
- Wadsworth, Y. (2001). Becoming responsive: And some consequences for evaluation as dialogue across distance. *New Directions for Evaluation*, 92, 245-58.