

## EVALUATION CAPACITY BUILDING IN THE SCHOOLS: ADMINISTRATOR-LED AND TEACHER-LED PERSPECTIVES

Frances Lawrenz  
University of Minnesota  
Minneapolis, Minnesota, USA

Kelli Thomas  
Douglas Huffman  
University of Kansas  
Lawrence, Kansas, USA

Lesla Covington Clarkson  
University of Minnesota  
Minneapolis, Minnesota, USA

**Abstract:** The purpose of this article is to describe evaluation capacity building using an immersion approach in two schools: one with an administrator-led process and one with a teacher-led process. The descriptions delineate conceptual, developmental, and sustainability aspects of capacity building through the perspectives of the teachers, principals, and evaluation specialists. The immersion approach to evaluation capacity building highlights the distinctions between developing capacity for using evaluation information and the capacity to conduct evaluations, and provides insight into the use of direct and indirect approaches to building capacity. The two cases also showcase the roles played by ownership in developing evaluation capacity, especially sustainability.

**Résumé :** L'objectif de cet article est de décrire le renforcement des capacités d'évaluation en utilisant une méthode d'immersion dans deux écoles : l'une gérée par des administrateurs et l'autre par des professeurs. Les descriptions fournissent la délimitation d'aspects de conception, de développement, et de durabilité dans le renforcement des capacités selon les perspectives des professeurs, des directeurs d'école, et des spécialistes d'évaluation. La méthode de l'immersion dans le cadre du renforcement des capacités d'évaluation met l'accent sur les distinctions entre le développement de la capacité d'utiliser les données recueillies dans l'évaluation

---

Corresponding author: Frances Lawrenz, Associate Vice President for Research, Wallace Professor of Teaching and Learning, Department of Educational Psychology, College of Education and Human Development, University of Minnesota, 56 East River Road, Minneapolis, MN, USA 55455; <lawrenz@umn.edu>

et la capacité de conduire des évaluations, et examine l'utilisation de méthodes d'approche directe et indirecte de renforcement des capacités d'évaluation. Les deux cas montrent également les rôles joués par l'appropriation du développement des capacités d'évaluation, notamment la durabilité.

## INTRODUCTION

In the present climate of accountability and economic responsibility, organizations are becoming more and more interested in building evaluation capacity. The movement toward evaluation capacity building is supported by Torres and Preskill's (2001) suggestion that organizations are moving more toward more internal evaluation and using the results to improve their practice. Stockdill, Baizerman, and Compton (2002) defined evaluation capacity building (ECB) as "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its use routine" (p. 14). They also pointed out that much work remains to be accomplished in order to understand how best to accomplish ECB. There have been some studies of ECB. Several individual case studies have reported specific ECB attempts (Baizerman, Compton, & Stockdill, 2002; Boyle & Lemaire, 1999; King, 2002). Arnold (2006) proposed a framework for capacity building that began with developing logic models, followed by providing one-on-one help to individuals, facilitating small-team collaborative evaluation, and conducting a large-scale evaluation. Miller, Kobayashi, and Noble (2006) recommended an ECB approach called "insourcing," which is a cross between internal and external evaluation where an external evaluator works with several small agencies to develop common evaluation approaches that are then implemented by the agencies. This is reminiscent of the cluster evaluation work discussed by Jenness and Barley (1995). Cousins, Goh, Clark, and Lee (2004) linked ECB to organizational learning. Their conceptual framework suggested that organizational readiness for evaluation may be influenced directly by evaluation capacity building and indirectly by doing and using evaluation. They suggested that ECB processes should be linked to needs assessments and were supported by facilitative leadership and modelling; ongoing training and technical support; the existence of prior knowledge, skill, and facility with evaluation logic; availability of resources for evaluation; and exigencies for evidence about program and organizational performance and results.

Despite the work outlined above, more research on ECB is still needed. In particular, issues of conceptualization, development, and sus-

tainability need to be studied. In this article, the ECB experiences of two public schools are presented. Both schools implemented the collaborative immersion approach to ECB (Huffman, Lawrenz, Thomas, & Clarkson, 2006). However, at one of the schools the approach was implemented in an administrator-led manner while in the other the approach was implemented in a teacher-led manner. The manner in which the approach was implemented (administrator-led or teacher-led) was not mandated but rather evolved in a more naturalistic way in each school. The perspectives of the people involved were documented through observations and interviews. The descriptions presented here expand the existing understanding of ECB conceptual issues in terms of distinguishing between organizational capacity to use evaluation information and capacity to conduct evaluations. Additionally, the results provide insight into the strengths and weaknesses associated with direct and indirect approaches to achieving ECB. Finally, the insights gained through considering the administrator-led or teacher-led approaches shed light on the sustainability of ECB.

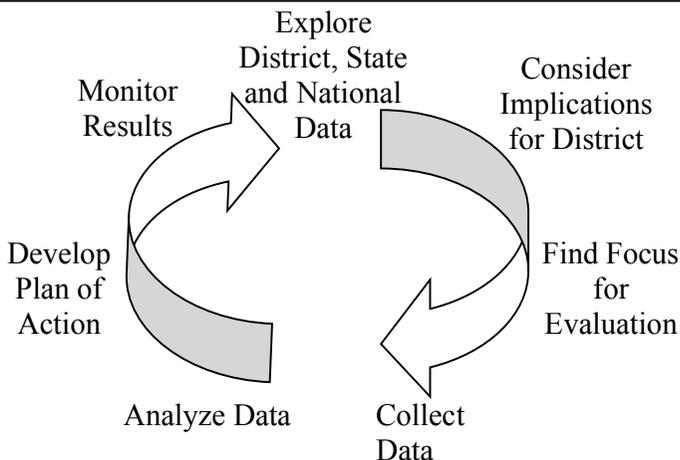
## DESCRIPTION OF THE ECB APPROACH

For ECB to be successful, the capacity of both the individuals in the organization and the organization itself need to be developed. Individual capacity building includes developing both traditional evaluation skills and knowledge and a disposition toward evaluation. Organizational capacity building includes instituting processes that support a culture of evaluation, such as strategic positioning methods, continuous improvement strategies, work plan development and assessment, staff development policies, attitude toward change, and procedures for project and individual reviews.

The collaborative immersion approach develops the evaluation capacity of individuals and organizations both directly and indirectly. In the collaborative immersion approach, individuals are immersed in a collaborative evaluation experience as a means of developing individual skills and developing the capacity of the organization. The idea behind collaborative immersion is that the individuals and the organization are collaboratively deciding how best to build their capacity in order to meet their real-world evaluation needs (Huffman, Thomas, & Lawrenz, 2008). The approach was developed by the National Science Foundation (NSF)-funded Collaborative Evaluation Communities in Urban Schools project (CEC) (Huffman et al., 2006). In this approach, teams composed of teachers, school administra-

tors, district personnel, graduate students, and university evaluators are immersed in the process of evaluation to build the evaluation capacity of both the individuals and the school. The CEC project accomplished this dual goal by engaging the team in an evaluative inquiry process focused on school programs and student achievement (Parsons, 2002). The teams examined student achievement data at the national, state, and local levels to identify broad challenges or opportunities for improvement related to either mathematics or science educational programs. Identifying aspects of the educational programs that could improve student achievement led the teams to develop problem statements. These problem statements were then translated into evaluations conducted by the teams to help address the problem. The initial evaluations produced data and findings that informed a plan of action to bring about change within the organization. Implementing the plan of action provided a context for asking new evaluation questions, more data collection, and additional analyses. Working together through a real evaluation that was directly connected to organizational issues led to a continuous improvement cycle through which evaluation was used as a tool to create change in the organization. In this way the individuals and the schools were directly and indirectly immersed in designing evaluations, collecting and using data, and setting up systems to support the continuous improvement process. The individuals were directly involved when they specifically completed tasks related to the CEC process and their school, but they were indirectly involved in the valuing and designing components. See Figure 1.

**Figure 1**  
**Inquiry Cycle**



## THE ADMINISTRATOR-LED APPROACH

In this approach, the immersion project started with schools submitting applications to participate. A district-wide call for proposals was distributed to the principals of all elementary and middle schools in the district. The intent was for principals, instructional coaches, and teachers in the schools to initiate interest and a desire to participate in the project. In the administrator-led approach, the principal of the school called for volunteers to develop a proposal in response to the call. The teachers who volunteered met with the principal and the instructional coach to complete an application that asked schools to identify aspects of the school's mathematics or science program that warranted investigation. From the beginning there were undertones of administrative influence on this group of teachers. In the application, the group described wanting to find ways to align the mathematics curriculum assessments with state mathematics standards and tests.

### Implementation

Central School is a K–5 elementary school located in an industrial section of an urban school district in the midwestern United States. The school is overcrowded, and some classes are taught in temporary structures next to the school that have become permanent fixtures. The school utilizes a “looping” policy for teachers who stay with the same group of students for two years. The student population is approximately 48% Hispanic and 44% African American; 93% are economically disadvantaged; and 64% have limited English proficiency.

When the project started, Central's team consisted of five teachers (one resource teacher, one music teacher, and three classroom teachers), the instructional coach, the principal, a graduate student, and the university evaluation specialists. The project was initiated at a meeting in which the team reviewed national and state data relating to the district's student achievement data in mathematics and science, which provided direct experience in using evaluation results. The Central team spent the next few months developing a plan for an evaluation of classroom instructional and assessment practices in mathematics based on issues identified through the initial analysis of data. This provided the school members of the team with a model and a somewhat indirect experience of designing an evaluation. The primary concern at Central was students' poor performance on annual state mathematics tests. The administrative

members of the team (the principal and instructional coach) regularly reminded the group of issues related to testing. The teachers believed that their students understood the mathematics concepts being tested but had challenges translating their understanding to a multiple-choice state test. The school district had adopted a curriculum that emphasizes learning mathematics concepts through investigation. The assessments recommended by the curriculum are based more on performance of problem-solving tasks and emphasize solution strategies rather than selecting the correct response to a multiple-choice test item.

After conducting a teacher survey of mathematics instructional practices in the school and through continued analysis of student achievement data, the team decided that they wanted to work toward improving students' state test scores by attempting to align classroom assessments with the state mathematics standards. In this case, designing the survey was more indirect while using the results to move forward was more direct. The team worked on creating mini-assessments for Grades 3, 4, and 5 that were aligned with units in the curriculum and designed in the same format as the state test. The Grades 3 and 4 mini-assessments were created by the classroom teachers and the resource teacher. The principal, instructional coach, and music teacher developed the Grade 5 mini-assessments. Two teachers on the team pilot-tested the mini-assessments during the school year. The team used the mini-assessment data to evaluate student learning needs and then developed a school-wide intervention in response to students' weaknesses in number sense. By the end of the second year, the team was concerned that more teachers in the school were not using the mini-assessments.

During the second year, other initiatives at the state and district level were occurring that affected the work of the Central team. The school district was piloting the use of regular student testing in mathematics and reading. The tests were developed by an outside agency, and the district needed schools to volunteer as pilot sites. The principal agreed that Central would participate. Students in the school were formally tested at three checkpoints during the year, providing teachers with additional student test results. The project team reviewed these student results and attempted to align the results with mathematics standards and the curriculum. Analysis for the purpose of modifying instruction was difficult because student scores fell into categories that represented a wide range of concepts and skills across multiple standards. In addition, the state was in the

process of developing an online resource that included a “test-builder” function for teachers. The test-builder was designed to automatically generate assessment items by mathematics standard as a formative assessment tool for teachers. By the end of the second year, Central had access to multiple sources of student data.

The team planned to begin the third year by having project teachers lead a school-wide meeting to share the evaluative inquiry processes with the teachers who had not been involved with the project. The hope was that other teachers would become more aware of the work of the project team and start using the data as part of a school-wide effort to build the evaluation capacity of the school as an organization. The teachers who had been using the mini-assessments found them to be beneficial for monitoring student progress toward mathematics standards while using the district-adopted curriculum. They used the results to inform classroom instructional practices. The mini-assessments developed by the school administrators were not being used by the Grade 5 teachers who were not part of the project team. The administrators indicated that they would take the lead in encouraging the Grade 5 teachers to use the mini-assessments. This approach was not successful, as the mini-assessments were not used consistently by Grade 5 teachers.

By the beginning of the third year, there was concern because the school had failed to meet the requirements for Annual Yearly Progress (AYP) as specified by the No Child Left Behind Act (Elementary and Secondary Education Act, 2001). This act provided that if the students in a school did not show sufficient progress on the state test scores from one year to the next (AYP), the school would be placed on “probation”; if the scores still did not improve, the school would be closed and reconstituted as a new school. The administrators decided that the school should use the state’s test-builder tool and the external test the district had purchased for formative assessment. They agreed that teachers in the school could also use the mini-assessments developed by the project team but sent clear messages that the online tool was probably better. The underlying message was that using the test-builder tool would be the best way the school could raise scores on the state test. The district also announced that a new version of the mathematics curriculum would be adopted the following year, presenting the possibility that mini-assessments would not align with the new curriculum. The teachers in the school were discouraged and questioned the extent to which district administration valued their professional expertise.

## Perceptions of Teachers

Teacher perceptions of the process were obtained by the project's external evaluator through yearly interviews with teachers and the administration. The following are excerpts from the second year evaluation report.

Those teachers were enthused about their creation of "mini-assessments" and believed that these assessments were helpful for both purposes of diagnostics and confirmation of student gains. The school principal indicated that these new assessments had been used in Grade 3–5, but also commented that teachers who participated in creating the mini-assessments seemed to be more invested in this assessment process than those who had not participated in this work.

These teachers had uniformly positive views about the impact of the project on teachers and students. They reported that they are assessing more (e.g., 4 investigation/units with assessment after each). One teacher noted that the grade levels not tested now have a better idea of what goes into the test. She stated that teachers generally have had a negative attitude toward the curriculum and this work is changing their perspectives. She noted that the project work bridges the gaps and makes teachers more comfortable in teaching the curriculum. Another teacher noted that before the curriculum and state assessment seemed different; now teachers see more of a connection. She also noted that now the curriculum and standards are integrated (embedded) so that standards don't seem "thrown in their face."

The following are excerpts from the final evaluation report.

Both teachers (3rd grade and 5th grade) I interviewed use the math investigations curriculum. The 5th grade teacher "looped" from 4th grade this year. She had been involved with the project from near its beginning and participated in development of the 4th grade assessment materials. The third grade teacher came to this school last year (this was her second year at this school) and she has been participating in the project last year and this school year. She was a strong advocate for the pre/post assessments.

Likely because AYP was not achieved for math, the curriculum as taught and tested in the school was compared again with the state assessments. This review reportedly identified major gaps between what is taught and what is required/tested by the state. Those interviewed attribute the shortcomings to the persons who created the 5th grade assessments in the first year of this project and their failure to produce a “match” between curriculum focuses and the state standards.

### Perceptions of Administrators

The principal’s and instructional coach’s perceptions were obtained through a joint interview. Both requested that the interview not be taped but rather that the interviewer take notes. The notes were then reviewed and approved by the administrators. The interview was not structured to allow the direction of the interview to be shaped by the administrators. The following is a summary of the interview.

The administrators both shared that they perceived benefits from the ECB efforts at Central. They agreed that the collaborative nature of the work was beneficial. They indicated that collaboration between teachers at different grade levels probably would not have happened if the structures had not been established through the project. The school utilized collaborative planning for teachers at the same grade level but not across grade levels, and the focus was on instructional planning rather than evaluation activities. The administrators believed that increased attention to alignment among curriculum, classroom assessments, standards, and external assessments was an important achievement of the project team. They thought the project’s school-level initiatives demonstrated forward thinking and that the intentional effort to build evaluation capacity was positive. The administrators liked the structure of regular meetings because they provided opportunities for discussion about gathering and using data to determine the effectiveness of the mathematics program.

The administrators also acknowledged that the project team had challenges moving beyond plans for gathering data school-wide. A few teachers did gather student assessment data, and analysis of those data was used to make changes rather quickly during the first year. However, the administrators believed several barriers prevented the expansion of the evaluation capacity building efforts to the whole school. Four barriers reported included (a) teacher turnover, (b) dif-

faculty finding a time when all team members could attend meetings, (c) other initiatives that took priority over the work of the project team, and (d) the perception that district administration attended meetings to monitor or “check up” on the school rather than to support it.

There was also concern about various policies implemented by the district such as increased attention to test security and the threat of sanctions if a school was found to be using test items for instruction. This school was not using assessment items for instruction, but the message from the district caused the school administrators to be concerned about the consequences if outsiders misunderstood how the mini-assessments were used. They worried that if the school continued to gather student data from the mini-assessments it would be viewed by district administration as “teaching to the test” rather than using assessment to inform and evaluate the effectiveness of instructional practices. It is important to note that the mini-assessments were not used for instruction but rather as tools for gathering data that would affect instruction.

## THE TEACHER-LED APPROACH

In this approach, the immersion project started small and built up to entire school involvement over a four-year period. It was similar to the administrator-led approach in that the school and the district had agreed to participate in the project, but the similarities ended there. The school decided to participate because the teachers and principal were already familiar with one of the university team members and wanted her help in improving their school. The principal selected the primary grades for participation because these grades were affected less by other district or school initiatives.

### Implementation

Smith Elementary School is a grade K–6 magnet school in a large urban school district in the midwestern United States. Approximately 350 students attend this school each year, with a 13:1 student/teacher ratio. Approximately 94% of the students at Smith are eligible to receive free/reduced-price meals, with food service representing 3% of the total revenues and expenditure for the school. At Smith, approximately 78% of the students are African-American, 14% Asian Pacific Islanders, 4% Hispanic, and 4% White. Generally, 19% of the

students have limited English proficiency and 21% are in the special education program. Thirty-eight percent of the teachers hold master-level degrees and have more than 10 years of teaching experience.

At the beginning of the CEC project, only the nine Grade 1 and 2 teachers were involved along with the university graduate students, the professors, and the principal. The team began by examining international, national, state, and district data in elementary math during the kickoff meeting at the end of April. Focusing on the Trends in Mathematics and Science Study (TIMSS), an international study of mathematics and science student test scores in which the United States' performance was mediocre, and district level test data, the teachers reflected on their current teaching practices that included direct experience with using evaluation data. During the month of May, the CEC group met twice and generated challenge statements and proposals to do something new at their school. Examples included alignment of their mathematics curriculum with state/district standards, setting a pacing schedule, development of teaching strategies for math, assessment techniques, student motivation, and family involvement. The team decided to focus on student outcomes as they related to teaching strategies including pacing, measurement, and cooperative learning. The yearly state/district assessments were not available for the primary grades and, as a result, too far removed from the instruction to be useful evaluation devices, so the team developed unit-based student assessments. Participation of the teachers in the development of these evaluation plans was direct building of capacity to conduct evaluations. Pre and post assessments were provided quarterly so the teachers could link the results to their instruction. As the assessments were not directly designed by the teachers, this was indirect experience. Monthly meetings as well as longer workshops in the summer provided opportunity for discussion. It is worth noting that this was the first time each of the participating teachers had completely taught the entire curriculum. The meetings and discussion were focused on how to use the collected evaluation information.

By the second year of the project, the beginning of a “buzz” could be felt in the school, which could be considered as indicative of indirect ECB. This year the new Grade 1 and 2 teachers joined the team. Two former Grade 1 and 2 teachers who had moved into more administrative positions were retained—one as a mathematics tutor and the other as a parent liaison. It was also apparent that the Grade 3 teachers were interested in joining the project (because one teacher

had looped with her Grade 2 class); they were also added to the team. During this year the teachers continued to focus on instruction but expanded the data sources to include videotaping of lessons and consideration of parental involvement. The nuances of the instructional strategies revealed by the videos were discussed among the team and linked to student outcomes. These discussions facilitated direct capacity building related to use of evaluation information.

By the third year, teacher leaders were beginning to emerge, which would be an example of developing the capacity to conduct evaluation. These teachers were encouraged to lead presentations at the monthly meetings and were sent to national education meetings. Attempts were made to transfer the development of exams to the teachers by providing two workshops related to using assessment disks (from the curriculum) with items related to the instructional units. The principal had become convinced that the whole school would benefit from participation in the project and suggested that the project work with all the teachers—an example of indirect school-level ECB. After discussion with the members of the team and the potential new teachers, the project was embraced by the entire school. The principal's more active involvement allowed the project to be part of the official school schedule, making the school-wide ECB more direct.

With the expanded operation, unit-based tests had to be designed for all the grade levels. Based on the school's performance on the state/district tests, it was additionally decided to focus on teaching problem-solving strategies, so additional problem-solving tools were developed and school-wide professional development was provided for the staff. Because of the expansion, it was impossible for the university personnel to phase out of the school in the third year, so their phase-out was relegated to the following year. Despite this unanticipated outcome, efforts for identifying teacher leaders intensified and now five teacher leaders have been identified who have volunteered to continue ECB in the school.

### Perceptions of Teachers

Teacher perceptions were obtained by the project external evaluator through yearly interviews. The following are excerpts from the final evaluation report.

Several project impacts were noted. These include an emphasis on the ideas provided, such as "The partnership

brought different perspectives—other ideas, identified weak points and helped with their teaching.” A couple [of teachers] spoke of a greater student excitement for learning. One noted a greater focus on data, “look at data more closely.” One noted a greater attention to student needs—where students struggle, and one noted a greater focus on math than previous to the project. Consistently teachers mentioned that they can now complete the entire math curriculum for the year, something they had not been able to do prior to the project.

In terms of meeting or exceeding project expectations, the teachers listed a variety of points including: the assistance provided to them to learn about and better teach mathematics, the injection of new ideas—interventions for working with students, and the opportunity to go to Atlanta and present to others about their project work. One teacher noted that the “project exceeded expectations by not producing more work for us, just required our sharing.”

All the teachers noted personal changes as a result of the project. Most talked about improved understanding, teaching skills, and greater knowledge of how to teach the curriculum. Plus, they can now cover the full curriculum in the year. While only one teacher noted changes in assessment practices, several noted that the assessments have had an impact on their judgments of students and how they think about and use assessment information.

When asked to grade the project, the grades ranged from C to A+. Three of the teachers rated the project an A or A+, and two gave the project a B or B+. Overall, the project is viewed as meeting its obligations and doing a wonderful service to the school. The university’s assistance in helping them to teach math better, the university’s putting tests together and crunching the numbers, and the interventions created by the university were noted by those who gave positive grades. The testing problems, the fact that results were not obtained in a timely way for use in parent-teacher conferences, and difficulties in using the new chart intervention were noted by those awarding lower grades.

## Perceptions of Principal

The principals' perceptions were obtained through conducting an interview with the principal and then having the principal read through a written summary of interview. The current principal of Smith Elementary inherited the position after the death of the first principal just a week before the beginning of the academic year. However, because the initial principal had been ill, the new principal had extensive opportunity to discuss the CEC project with the initial principal before she assumed the role of principal. Because of these circumstances, the new principal's perspectives presented here echo the voices of both individuals.

Both principals believed that the CEC program should be broadened to include every grade level in the school because of the positive results the primary teachers were experiencing in their classrooms and the positive input from the participating teachers. The teachers described in detail the challenges that they experienced in teaching mathematics and the strategies that were being implemented because of the CEC project. The school had not made adequate yearly progress in the past and particularly struggled with the teaching and learning of mathematics. Because the CEC project had facilitated mathematics instruction in the early grades, the principal felt these opportunities should be provided to the entire school. For example, a computer software program had been incorporated into the primary mathematics time; it focused on basic skills but caused teachers to struggle to complete the regular curriculum. In addition, more time during the school day had been allocated for reading and language skills development, which caused other subjects to be addressed in shorter amounts of time. The principal felt that participation in the CEC project had helped the teachers recognize and resolve these issues and wanted more evaluation to help consider the value of other school practices. The principal believed that the support of university personnel would positively affect the mathematics program in the school and welcomed the opportunity to increase participation in CEC. The new principal was so positive that she even provided time during the beginning of the year for a school workshop about the CEC project.

Even with the support of the principal, the teacher-led approach was evident in the discussions within the context of the evaluation process. Several of the original primary teachers had been reassigned to other grades, which distributed leadership to the intermediate

teams as well. In addition, the financial plan for the next academic year provided a mathematics coach to continue the work that had been introduced to the school-wide mathematics program. One of the parent meetings focused on mathematics in order to address one of the original challenges, which was to include parents as partners in their child's mathematics education. The principal perceived that the CEC relationship was strengthening their overall mathematics program through the identification of the challenges that inhibited success in teaching and learning mathematics and through implementing follow-up strategies that addressed the issues. The principal stated that the additional support of the evaluation team from the university provided the necessary data collection and analysis that teachers were ill equipped to do.

#### PERCEPTIONS OF THE UNIVERSITY TEAM MEMBERS ON BOTH APPROACHES

Overall, the university team members felt the ECB process had both strengths and weaknesses but the strengths outweighed the limitations. In terms of strengths, schools as an organization appeared to view the project as providing an opportunity to improve, to address particular difficulties, to develop solutions to long-term problems, and to develop the potential for long-term relationships. However, the school sometimes viewed the project as interfering with other agendas and not necessarily addressing their highest priority problems. Teachers viewed the project's strengths as providing help in their instruction by bringing in innovative ideas and materials, by helping teachers make data-based decisions, and by providing opportunities to participate in research. However, they also thought the project included extra work and time; it forced them to question their existing practices; and it was difficult to develop a true community.

The administrator-led process worked well initially. Through direct support and direction provided by the administrators, the team was able to agree upon and develop a focus for an evaluation that was connected to organizational issues. The individual team members were immersed in designing an evaluation, creating data-collection instruments, collecting and using data, and establishing structures that would support the organization in conducting and using evaluation for a continuous process of improvement. The classroom teachers who participated in the project were active participants in evaluation processes but did not take on leadership roles in expanding the work beyond the small group on the team. The administrators

took responsibility for seeking involvement by other teachers in the school, and there seemed to be an undertone of resistance to what might have been perceived as a top-down mandate from the principal and instructional coach. The school administrators were also sensitive to the policies and initiatives of the district, particularly those related to student assessment. These external factors contributed to some hesitation by the school administrators about using the mini-assessments for evaluation purposes. The organization does have structures in place that can be used for evaluative inquiry processes that were not established prior to the project. Unfortunately, external pressures to demonstrate AYP have overshadowed the ECB processes at this time. The administrators acknowledge that aspects of evaluation developed through the project will provide a mechanism for continuous improvement processes in the future, but whether they will be able to sustain such efforts is unknown.

The teacher-led process appeared to work well. Beginning with a small group of teachers and building trust, then using this trust to move more broadly into the organization provided an effective pathway. Teachers who initially participated were able to reassure new teachers. The movement of participating teachers in and out of administrative positions allowed the project to become embedded in the organization. Concrete demonstration of effectiveness encouraged more administrative support resulting in institutionalization. The sustainability of the project is still in question, but the prognosis is positive. Lead teachers have been identified, and a few of the teachers have gained capacity in the traditional evaluation skills. All teachers learned more about how to use evaluation data. The culture in the school has become one supportive of data-based decision making. Despite the positive outcomes, the project has also taken almost four years and has been supported by substantial external funding. The organization, however, does have a much more data- and evidence-centred focus. The tendency has become one of linking initiatives with some type of evaluation that will help to identify its effectiveness.

## DISCUSSION

The two case studies presented here expand upon the existing understanding of conceptual, development, and sustainability issues surrounding ECB. In terms of conceptual issues, the immersion approach to ECB helped the schools develop the capacity both to do and to use evaluation, although it was probably most successful in developing skills in using evaluation. In both schools the personnel were

presented with a variety of evaluation information and were directly helped to understand how to use that information to improve their school. It was clear that all personnel became better able to make decisions based on data. Additionally, the administrators and teachers in both schools learned how to do evaluation through immersion in multiple evaluation processes that were both direct and indirect. The teacher-led vs. administrator-led approach was an important distinction that naturally occurred in the two schools. The two approaches are important because they provide information about the different ways school personnel may engage in evaluation, and how each approach can develop ECB.

From the beginning the university evaluators pushed the schools to use the evaluative inquiry cycle as a framework for their investigations. In the administrator-led school, the specific involvement of the administrators seemed to facilitate using the inquiry cycle because the administrator supported school-wide data collection (i.e., instructional and assessment practices survey). In the teacher-led school, the teachers seemed to take more ownership of the data, recognizing that data availability was limited. (School-wide data were not available for the first years and grade-level data were not collected because they were not mandated for Grades 1 and 2, which were the grades of the majority of the participating teachers.) Because of the complexity of the tasks and contexts surrounding the ability to conduct evaluation, *capacity building to do evaluation* will require a longer time frame than building the *capacity to use* evaluation information even though each experience increased the schools' capacity to conduct evaluation.

Developmentally, the university evaluators utilized an integration of both direct and indirect methods for ECB. The participants at both schools were directly engaged in personal capacity building, which enhanced the organizational capacity as structures were developed and implemented to support evaluation. At the administrator-led school, teachers who taught at multiple grade levels began meeting to generate evaluation-oriented questions, gather data to answer those questions, and make changes based on the data. The participants learned directly through the tasks they were specifically involved in and indirectly through the tasks the university people completed. These cross-grade-level meetings helped build organizational capacity both directly in terms of the specific meetings and structures put in place and indirectly in terms of understanding and acceptance of evaluation throughout the school. One drawback to this in the administrator-led school was the difficulty in attracting

more teachers, as only a few participated regularly. One strategy that was implemented at the teacher-led school was videotaping lessons as a data-collection tool, which was an indirect approach to ECB. The teacher leadership seemed to establish an environment in which teachers felt safe gathering concrete and specific evidence of their instructional behaviours. Both schools were also directly influenced to be more evaluation-oriented because the university evaluators interacted with the teachers and administrators, which pushed the schools to move forward. For example, at the administrator-led school the teachers and administrators had ideas about changes to be made related to the mathematics program, but they did not have evidence to support their proposal. The university evaluators helped them generate evaluation questions and develop data-collection instruments to gather evidence. Likewise, at the teacher-led school the teachers were assisted in identifying curricular challenges that originally appeared to be unique to the participating teachers but upon further examination of data and expansion of the number of participants, it was found to be a school-wide challenge. The university evaluators' presence at both schools was direct and specific to involve the schools in building evaluation capacity. By gathering data and using it to make decisions at both the individual classroom level and the broader school level, the participants were building evaluation capacity more indirectly. It appears that the more direct approach to ECB produced more in-depth and more focused, but more limited, results, while the indirect approaches produced more widespread understanding and attitudinal support. This fits with the finding above about this type of ECB being more likely to lead to capacity to use evaluation rather than to do it. There were many direct opportunities to use evaluation findings, and this use was aligned with the more established expertise of the participants—improving curriculum and instruction. The opportunity for doing evaluation—at least with all of the many components involved—was overall more indirect, as the participants were directly involved with some aspects but not all. Additionally, the capacity necessary to do an evaluation was less aligned with the pre-existing capacity of the participants.

Issues of sustainability were less stable at both schools, and results will not be fully known until the schools are in a position to continue the work without the involvement of the university evaluators. At the administrator-led school, the two administrators have taken clear leadership roles related to evaluation processes but may not have developed sufficient evaluation skills to sustain future evaluative

inquiry. Likewise, whether or not the teachers in the school will follow their lead is still not apparent. It may be that when administrators utilize a directive leadership style, teachers resist the traditional “top-down” model. Also, school-level initiatives were often trumped by district-level mandates that can have an effect on the ability to sustain school-level evaluation efforts. At the teacher-led school, capacity for sustainability proceeded slowly. This year, however, the university evaluators are transitioning out of the school, and five of the teachers stepped up to become teacher leaders. They will be working with the evaluators to plan the summer workshops and will conduct the inquiries in the upcoming year. It remains to be seen, however, just how successful these new “evaluation leaders” will be. It appears that the best type of leadership would be a combination of the two approaches. An administrator might begin the process to give it a jump start but might then have to “back off” and let the teachers really develop their approach for a time. Then the administrator might be invited back into the picture to help ensure institutionalization and the necessary resources to continue the process.

Cousins et al.’s (2004) organizational learning framework suggests that ECB processes are supported by several things: needs assessment, facilitative leadership, and modelling; ongoing training and technical support; the existence of prior knowledge, skill, and facility with evaluation logic; availability of resources for evaluation; and exigencies for evidence about program and organizational performance and results. Preskill (2008) also suggested several ECB supports, including stakeholder involvement in evaluation, training, technical assistance, written materials, communities of practice, internship and apprenticeship, coaching and mentoring, appreciative inquiry, meetings, and technology. She suggests that effective ECB should also include clear goals and objectives; reflect adult learning principles; be designed to foster transfer; include dialog reflection and feedback; consider different learning styles; take into account the cultural context and be culturally relevant; employ facilitators who are trusted, respected, and humble; and provide adequate time for learning.

The collaborative immersion approach to ECB described here incorporated most of these elements and appears to be an effective means of improving educational organizations. The model offers significant strengths in terms of involving teachers in data-based decision making and in providing the data necessary to address school-based problems. The model can also affect school culture and

facilitate the development of structures to support the use of evaluation within the organization. Needs assessment was initially provided through the use of national, state, and district data. These data were then supplemented by more locally relevant assessments and processes. Clear goals and objectives were provided for each inquiry and the use of direct and indirect approaches to ECB were designed to foster transfer. Participants were directly exposed to some components related to conducting evaluation (e.g., developing mini-assessments, using evaluation data to make decisions) and indirectly exposed to others (e.g., designing an evaluation plan, building supportive structures). The collaborative communities provided facilitative leadership including coaching and mentoring by trusted facilitators who were respectful of cultural differences and adult learning styles.

The skills and facility with evaluation logic and resources for evaluation were provided by the university members of the team; however, how those resources will be provided in the future is unknown. Because the administrator-led approach had more initial emphasis on leadership facilitation and modelling, it moved more quickly than the teacher-led approach to a point where internal sustainability would be possible, although the teacher-led approach also achieved a measure of internal sustainability over time. The administrator-led approach produced more school members with specific evaluation expertise than did the teacher-led approach, but both approaches resulted in organizations with cultures more supportive of evaluation. Both administrator-led and teacher-led approaches also resulted in the consideration of evaluation findings and how to integrate findings into organizational operations. It appears that, conceptually, initial efforts at ECB are more likely to produce organizational or school capacity in how to use evaluation rather than capacity in conducting evaluation. It also appears that a blending of administrator-led and teacher-led efforts, carefully utilized to launch and sustain evaluation capacity building, would help ensure the integration of more aspects for both conducting and using evaluation into the organization. Furthermore, it appears that both direct and indirect approaches to ECB may be necessary to move a school forward and that broad ownership in the ECB process promotes the potential for sustainability. Ideally, the increased evaluation capacity of the schools will help sustain their efforts to engage in continuous improvement and make them into true evaluation-based organizations.

## REFERENCES

- Arnold, M. (2006). Developing evaluation capacity in extension 4-H field faculty: A framework for success. *American Journal of Evaluation, 27*(2), 257–269.
- Baizerman, M., Compton, D. W., & Stockdill, S. (2002). New directions for ECB. In D. W. Compton, M. Baizerman, & S. Stockdill (Eds.), *The art, craft and science of evaluation capacity building. New Directions for Evaluation, 93*, 109–116.
- Boyle, R., & Lemaire, D. (1999). *Building effective evaluation capacity: Lessons from practice*. New Brunswick, NJ: Transaction.
- Cousins, B., Goh, S., Clark, S., & Lee L. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation, 19*(2), 99–141.
- Elementary and Secondary Education Act (The No Child Left Behind Act of 2001). (2001). Pub. L. 107-110, 115 Statute 1425 (enacted January 8, 2002).
- Huffman, D., Lawrenz, F., Thomas, K., & Clarkson, L. (2006). Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. *New Directions for Evaluation, 109*, 73–85.
- Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29*(3), 358–368.
- Jenness, M., & Barley, Z. (1995). Using cluster evaluation in the context of science education reform. In R. O'Sullivan (Ed.), *Emerging roles of evaluation in science education reform. New Directions for Evaluation, 65*, 53–69.
- King, J. A. (2002). Building the evaluation capacity in a school district. In D. W. Compton, M. Baizerman, & S. Stockdill (Eds.), *The art, craft and science of evaluation capacity building. New Directions for Evaluation, 93*, 63–80.
- Miller, T., Kobayashi, M., & Noble, P. (2006). Insourcing, not capacity building, a better model for sustained program evaluation. *American Journal of Evaluation, 27*(1), 83–94.

- Parsons, B. A. (2002). *Evaluative inquiry: Using evaluation to promote student success*. Thousand Oaks, CA: Corwin.
- Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation*, 29(2), 127–138.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In D. W. Compton, M. Baizerman, & S. Stockdill (Eds.), *The art, craft and science of evaluation capacity building. New Directions for Evaluation*, 93, 7–25.
- Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22, 387–395.

**Frances Lawrenz**, Ph.D., is the Associate Vice President for Research at the University of Minnesota, and the Wallace Professor of Teaching and Learning in the Department of Educational Psychology at the University of Minnesota.

**Kelli Thomas**, Ph.D., is an Associate Professor of Mathematics Education at the University of Kansas.

**Douglas Huffman**, Ph.D., is a Professor of Science Education at the University of Kansas.

**Lesla Covington Clarkson**, Ph.D., is an Assistant Professor of Mathematics Education at the University of Minnesota.