4

This case study documents two years of ECB in a large midwestern school district and exercizing ECB goals.

Building the Evaluation Capacity of a School District

Jean A. King

ECB is a growth industry in local school districts in the United States. The burgeoning use of high- and low-stakes accountability testing mandated by districts and states (and perhaps in the near future by the U.S. Department of Education) has focused attention as never before on individual schools' abilities to increase student achievement, especially their test scores, over time. Administrators and staff can read about their schools' successes or failures in the newspapers, and those that consistently perform poorly may be subject to external intervention, including takeover or "reconstitution." Advances in technology that can bring schools relevant and understandable data that are easy to manipulate, coupled with accountability and accreditation systems that require yearly school improvement plans, point to the importance of infrastructures within schools and district offices to support ongoing evaluation activities.

This case study focuses on the process of building evaluation capacity in a school district from the perspective of an internal evaluator hired explicitly to develop the evaluation function for the district, that is, to carry out ECB. The case example will show how this work was originally conceived and how and why it changed, describing activities over the course of two years and framing the issues facing one school district as it sought to build evaluation capacity. Although other district contexts will vary, the lessons learned in this district may illuminate the challenges confronting any large school system seeking to increase its long-term capacity to conduct and use program evaluations in its everyday activities.

District Context

With 40 schools and roughly 42,000 students, Anoka-Hennepin Independent School District 11, which comprises thirteen communities northwest of Minnesota's Twin Cities, is the third largest in Minnesota. Unlike the state's two largest districts (Minneapolis and St. Paul), District 11's communities do not face the challenges resulting from being an urban center such as high poverty rates, a high percentage of students who speak English as a second language, and high rates of family disruption. Students do well both on state-mandated tests (grades 3, 5, 8, and 10) and on local school-board-mandated, nationally normed tests (grades 4 and 6), suggesting that the children really *are* above average—an inside joke in Minnesota. Although the district's minority population has doubled in recent years, it is now only 8 percent, and the percentage of students receiving free and reduced-price lunches, while growing, remains comparatively small.

Garrison Keillor of National Public Radio's "Prairie Home Companion" fame graduated from Anoka High School, and Governor Jessie Ventura, whose home is nearby, has helped coach football at another of the district's four high schools. District challenges stem from a low tax base that yields low per-pupil funding, from continuing growth in the district as families move into local communities, and from a sincere commitment to help every student reach the maximum of his or her potential. The current superintendent often states publicly, "We teach all students one child at a time." It is a local school board goal, for example, that all Anoka-Hennepin children will "read independently and well" by the end of third grade, regardless of their skill level when they enter kindergarten.

Although internal program evaluation is a relatively recent addition to district practice, high-quality student assessment is not. The district's Student Assessment Department was created in the mid-1980s when a long-time superintendent hired a charismatic assessment director to expand the testing program into the areas of curriculum and classroom instruction. Under her leadership, the department's activities broadened to include standardized testing, curriculum-managed instruction (CMI), and criterion-referenced testing (CRT) at a number of grades, and the Assurance of Basic Learning (ABL)—a high-stakes test to document basic student learning prior to high school graduation. As the superintendent once stated, "If the movement toward a quantified accountability had not emerged as state public policy, it would have emerged as an Anoka-Hennepin School District effort. Throughout our district's history we have attempted to lead in the field of accountability, not follow." When two educational reforms swept the state (outcome-based education, which morphed into standards-based reform) the district was a leader in developing and piloting performance assessments in a variety of curricular areas. With the addition of state graduation standards and competency tests—including state high school graduation tests in reading, math, and writing—CMI, the CRTs, and ABL were dropped from district

practice. The Student Assessment Department remains the administrative unit responsible for increasing numbers of mandated standardized tests, which mushroomed from six different tests in 1996–97 to twelve in 2001–02. Over the years, the work of assessment has become, as one of the staff often quips, "Tests Are Us." While the standardized testing load has doubled, the number of support staff in the department has remained at three.

But as important as student assessment remains, the district superintendent and the associate superintendent for instructional support recognized an equally important role for collecting and compiling information to inform decision making. Two examples point to a central district administration attuned to the value and use of evaluation studies and committed to adding them to district functioning. When two of the district's high schools adopted four-period school days and the other two did not, the district collaborated with the Center for Applied Research and Educational Improvement at the University of Minnesota in an extensive study of this natural experiment (Maruyama and others, 1995). Based on the results, the district ultimately implemented the four-period day at all four of its high schools.

The second example occurred during the 1998–99 school year, when the administration and local teachers union collaborated on a districtwide study of the implementation of graduation standards using the Concerns-Based Adoption Model Survey. The associate superintendent for instructional support presented the survey results at the opening administrative workshop in August 1999 so that building-level administrators could begin the year with current data on their teachers' concerns about implementing state-mandated graduation standards. At that same workshop, the district superintendent announced to everyone, "This district *will* become data-driven," suggesting a notion of evaluation similar to that of Preskill and Torres (1999, p. 1) that "an ongoing process for investigating and understanding critical organizational issues," with the instrumental use of data is integral to the continuing overall process.

At that time, the district had operated without a head of the Student Assessment Department for almost a year. The search for a replacement had ended unsuccessfully, and district leadership sought a creative expansion of the department's function. A teacher on special assignment assumed the role of assessment facilitator, with responsibility for district and state test administration, for biennial district involvement in the National Assessment of Educational Progress (NAEP) and for district-school liaison work on all testing matters. The district then hired as their internal evaluator an individual with more than twenty years' experience in educational evaluation. This new evaluator—a professor on leave from the University of Minnesota—was charged with conducting program evaluations and using them to build the capacity of staff at each of the district's forty schools, as well as in its central Curriculum, Instruction, and Assessment (CIA)

Department, to engage in data-based decision making, or ECB. This task was separate from the testing program (for example, the test orders, training, troubleshooting, and concerned parents). In contrast to evaluators in many districts, the internal evaluator in the Anoka-Hennepin district, given the title of coordinator of research and evaluation, was truly responsible for focusing on the district's capacity to conduct high-quality, useful instructional program evaluations that were indeed to be used at both the district and school level. Central to the coordinator's work was developing a revised "vision" and perhaps a new name for the department, one that incorporated both functions—student assessment and program evaluation and its uses.

This difficult task was made easier by the district's work climate. The professional climate in the central administration is one of mutual respect, trust, and good humor. A central coffee machine and cafeteria foster informal connections and conversations, and several times throughout the year there are buildingwide luncheons and celebrations. Although people take their work extremely seriously, they are encouraged whenever possible to collaborate and to laugh along the way.

Two Challenges of a Large System

One of the challenges facing anyone working in a large bureaucracy is how to get information about and participate meaningfully in relevant activities. In addition to informal connections, the R&E (research and evaluation) coordinator purposefully attended regular meetings of two district administrative groups: the instructional support team (IST), which consists of professional staff who report directly to the associate superintendent for instructional support in addition to appropriate others (for example, the district head of HR); and the instructional facilitators (IF), twenty or so teachers on special assignment who provide subject-specific curricular support, as well as ongoing facilitation to one of four clusters of District 11 schools (one grade 9–12 high school and its elementary and middle feeder schools). In addition, she sat on four standing district committees that met monthly:

- Graduation standards implementation committee (GSIC)—the committee the school board charged with monitoring the implementation of the Minnesota Graduation Standards
- Blueprint (for literacy) implementation committee (BIC)—a committee overseeing the district's move (K–5) to a balanced literacy instructional approach
- Curriculum advisory committee (CAC)—the committee monitoring the district's overall curriculum development and implementation process, which is important in a district with a highly centralized K–12 curriculum
- Systems accountability committee (SAC)—a state-mandated citizens' review group that prepares an annual report for the district school board and the public on the status of education in the district, including student achievement results and the results of curriculum reviews

Participation in these meetings proved invaluable because it enabled the internal evaluator to monitor what was happening in this large system and to connect with various aspects of curriculum, instruction, and assessment across the district without attempting to be a visible presence at the forty schools. Shortly after her arrival, the coordinator received good advice: Do not plan to visit each building because even one or two trips to a school during the year (a total of forty to eighty site visits) would not satisfy staff there that an outsider (especially an evaluator who is at once a frightening and powerful stranger) knew their school more than superficially—and they would be right.

By participating in these meetings, the coordinator met key district stakeholders and worked to understand potential evaluation issues affecting instructional change. During year one, these meetings helped her develop an awareness of the complexity of the district and of the status and effectiveness of the evaluation function therein—where it might be added, fostered, or otherwise changed. For example, conversations with the district's head of diversity led eventually to the planning of a developmental evaluation of the district's diversity committee's goals, which an advanced doctoral student at the University of Minnesota is conducting. This project will eventually collect annual survey data on cultural literacy to monitor its development in both students and staff. (The work is also closely aligned with the external evaluation of the district's Desegregation Plan.) Meeting participation also created opportunities and communication of ideas about ECB in the district, including evaluation studies and their use in decision making at the school and district level. For example, before the release of results of the massive Graduation Standards Implementation Study (described next), the graduation standards implementation committee spent a lengthy meeting reviewing the data and checking claims based on them—a useful activity because it brought them face-to-face with the realities of the study's outcomes, that is, the realities to which they had to respond. This was use in practice and practice in use.

A second and more district-specific challenge came from staffing issues that affected all evaluation and ECB activities. First, although the Assessment Department workload had increased dramatically over the years, its staffing had not, resulting in frustration and anger for an extremely talented support staff. The staffing included five positions: two professional positions (the student assessment facilitator and the new coordinator of research and evaluation) and three "Class B" secretaries. These secretaries, whose positions were part of the secretarial union, were extremely competent people whose jobs required a high level of technical skill (for example, maintaining and manipulating large databases, using a number of complex software packages, preparing tables of quantitative data, interacting with schools, the public, and the state department) but whose duties prevented the reclassification of their positions to a higher level and a salary commensurate with the level of work (for example, they did not supervise anyone other than temporary clerical help).

Between 1998 and 2001, the Assessment Office earned the unfortunate reputation as a training site for other offices and agencies. Five secretaries cycled through the office: two moved to another district where they received better positions and sizable salary increases; one, frustrated after more than a year of trying to have her position upgraded, made a lateral move to a school; two took jobs in private industry. These changes left the office short one secretary for most of the 2000–01 school year, even as the workload increased due to the coordinator's evaluation and ECB activities. The constant deadlines and inevitable falling behind created a highly stressful work environment for the year.

A related staffing challenge concerned the supervision of department staff. Given the nature of their contracts, neither the assessment facilitator nor the R&E coordinator was allowed to directly supervise the support staff, which meant that the three secretaries reported to the director of curriculum, instruction, and assessment (CIA), even though the facilitator and coordinator were directing their work. Because of the personalities involved and the collaborative culture of the district office, the staff worked around these potentially confusing lines of formal authority. But the greatest success of the Fall 2000 semester—thanks to the efforts of the associate superintendent—was the reclassification of one clerical staff position to a nonbargaining-unit, office coordinator position. The person in that position now supervises the two remaining secretaries. In the proposed staffing for a reconstituted Evaluation, Assessment, and Research (EAR) Office, a full-time director of evaluation, assessment, and research would supervise both the assessment facilitator and the office coordinator. However, two factors have put that on hold for the foreseeable future: the board's traditional reluctance to add central office administrators and the grim budget situation for precollegiate education in Minnesota. A district that pink-slipped 130 teachers in May 2001 is unlikely to hire an assessment and evaluation supervisor in the months following.

Two Years of ECB Work (1999–2001)

Over the course of two school years (1999–2000 and 2000–2001), the coordinator collaborated with people both within and outside the school district to bring ECB goals to life in the district, and what follows is a progress report as the ECB effort enters its third year. During Fall 1999, the new R&E coordinator, with the support of the associate superintendent, developed a vision for the Assessment Department—an ECB plan that focused on four goals. The first was to develop staff commitment and skills in program evaluation and its use through participation in model evaluations of key district initiatives each year. This goal served two functions, one related to ECB and the other to program evaluation: (1) to engage people and teach them the logic and procedures of evaluation studies through an interactive process of evaluation (see Patton, 1997) and (2) to provide solid

data for decision making in important and visible areas of curriculum, instruction, and assessment.

The second ECB goal was to build an infrastructure for data collection, analysis, and presentation that would support program evaluation within the system. This infrastructure had to include the following capacities: (1) to generate and process surveys with quantitative and qualitative items, (2) to conduct focus groups and analyze their data, (3) to collect and compile student work, and (4) to prepare both oral and written evaluation reports. Partly related to the second, the third goal was to facilitate the existing school improvement process (SIP) by developing a system of school-level reporting to provide principals and teachers data about their schools, including existing test data and the results of SIP activities. Ideally, school staff would ultimately be able to customize SIP data collection, based on specific goals for their students. The final ECB goal—to create a network of teachers, administrators, and other staff across the district who would routinize classroom- or school-based inquiry—represented the coordinator's greatest hope: that the processes of program evaluation and its use would become integral to the way individual district personnel went about their work, making evaluation part and parcel of their ongoing activities.

Goal 1. The first goal was *to develop staff commitment and skills in program evaluation and its use.* The expanded Assessment Department's first function—to develop staff commitment and skills through participation in model evaluations of key district initiatives—in essence killed two birds with one stone. Instructional decision makers in the central office wanted solid data each year on the implementation and outcomes of specific initiatives designed to increase student achievement. At the same time, by using a participatory process to design and conduct these evaluations, district staff who took part would learn over time how to frame evaluation questions, how to put together data-collection instruments, how to make claims using data, how to use data in their everyday work, and so on. Two examples— one an evaluation of the implementation of Minnesota's graduation rule and the other a self-study of special education in the district—document how these model evaluations supported the ECB process.

During the 1999–2000 school year, the Graduation Standards Study produced extensive evaluation data using multiple methods, including two surveys (all administrators and Graduation Standards representatives in each building) and two rounds of focus groups (teachers nominated as exemplars of standards-based instruction and high school students experiencing standards-based education). Four components of the study purposefully engaged people in ways that would teach them about the evaluation process. First, the fifteen members of the graduation standards implementation committee, which was responsible for the implementation, reviewed the previous year's survey data and the concerns of the various groups they represented to collaboratively frame the study's questions. Throughout the year, they monitored the study's progress at their monthly

meetings, suggesting changes as needed. The second ECB component developed when it became clear that focus groups would help to answer one of the committee's questions and that the Assessment Department staff was unable to conduct enough groups to ensure widespread participation across the district. A group of instructional facilitators volunteered for training on how to conduct focus groups and then assisted in eighteen teacher groups and independently led sixteen student groups in the process, creating a cadre of experienced focus group facilitators for the district. Third, the committee sponsored a discussion session to which every teacher who participated in a focus group was invited. Hosted by the superintendent, the meeting content included oral and written presentations of the focus group results, followed by small-group discussion by the teachers to clarify and amplify the meaning of the data. In these discussions, teachers not only learned about others' perceptions of the Graduation Standards implementation but about the analysis of focus group data and the ways data could support district-level decisions. Fourth, as noted earlier, near the end of the year GSIC members engaged in an interactive analysis and discussion of the data, suggesting implications for the continued Graduation Standards implementation. They decided that the data collected were so extensive that no additional data would be required during the next academic year.

A self-study of the Special Education Department also supported ECB by developing two processes that are now integral to the district's evaluation function: a viable way to organize large-scale participatory studies and an interactive, cost-effective means of collecting qualitative data. Owing to the highly political nature of special education in the district, the self-study team had an initial membership of over one hundred people (representatives from different grade levels, roles, disabilities, and so on), fifty of whom eventually attended monthly meetings for over a year. The district coordinator worked with two external consultant-facilitators to design a participatory process for all phases of the evaluation study. In so doing, they developed a two-part structure that has become the model for all large studies in the district: (1) a six- to eight-person data collection team (DCT), made up of district personnel most closely involved with the study's focus, and (2) a much larger study committee (thirty to fifty people) that frames questions, analyzes data, and makes recommendations. The DCT was responsible for the nuts and bolts of carrying out the study—inviting people to meetings, following up on absences, compiling existing data, arranging logistics, and so forth. Relying on methods borrowed from cooperative learning, the monthly meetings of the study committee were devoted to short presentations, table teamwork on data analysis, and the development of claims based on the data, and eventual creation and prioritizing of recommendations. Those who participated in the study committee reported that they enjoyed both the special education content of the process and what they had learned about evaluation through their participation.

A second capacity-building process emerged of necessity during the course of the study. Given budget constraints, parental focus groups were

simply not feasible. But whether they were costly or not, the study committee believed they were needed in order to allow parents' voices to be heard. In response, the evaluation consultants devised a cost-effective adaptation of focus group methodology, first labeled "dialogue groups" (since renamed "data dialogues"), in which small groups of two or three participants discussed questions related to the study, then recorded their own data. In this way, large numbers of people were able to give data and simultaneously have a good conversation with other parents. Other evaluations in the district (for example, the teacher induction program evaluation, a study of middle school implementation) have since used the data-dialogue process both to foster meaningful discussion of issues and to generate qualitative data in a manner the district can afford.

After two years of studies, then, the first ECB goal of developing evaluation capacity in the district through participation in the evaluations of key initiatives had resulted in increased capacity in several ways. The members of certain committees accepted responsibility for working on evaluations of their activities, and over three hundred district staff, students, and parents had helped to manage or participate actively in one or more evaluations, learning about the process en route. In addition, a viable structure emerged for conducting effective evaluation studies with limited resources. During the 2001–02 school year, potential topics of inquiry (for example, the implementation of balanced literacy, the secondary advanced learner, and the "disengaged" learner) created additional opportunities for such capacity building.

Goal 2. The second goal was *to build an infrastructure for data collection, analysis, and presentation that would support program evaluation and its use within the district.* As noted earlier, this infrastructure needed to include the capacity to create and analyze high-quality data on any number of topics and to craft effective reports for intended users. Two components added to the infrastructure have been discussed—the instructional facilitators trained to conduct and analyze focus groups and the "data dialogue process"—both of which increased the district's capacity to collect qualitative data. The true challenge stemmed from the survey work that had been the department's responsibility for a number of years in support of the state-mandated curriculum review and other surveys. Historically, instructional facilitators had compiled teacher and student survey items, which Assessment Department clerical staff then copied onto computer forms. Once respondents returned the forms to the office, they were scanned and data summaries given to the appropriate facilitators. Although it was not a primary responsibility, department staff also facilitated the administration of other surveys on occasion.

In the context of ECB in a large district, the need for an effective way to routinely conduct surveys was heightened, but during these two years, computer hardware and software failure led to great frustration on the part of many people—Assessment Department staff primary among them. For over a decade, the Assessment Department had used the same software to

scan and analyze bubbled data sheets, successfully keeping the survey system alive even when newer programs became available. In fact, by the year 2000 the company that originally wrote the scanning software no longer provided technical support for it, so there was no external help when something went wrong. During the course of the first summer, it became clear that this system was on its last legs—staff who had previously been able to coax the machine to cooperate were no longer able to do so, and scanning became a highly uninviting and time-consuming experience for those responsible. At times, the equipment simply wouldn't work and projects sat unfinished. Unfortunately, the cost of a new system was far beyond the department budget, so for about a month during the summer it seemed there was no solution. The hero of this story was the district's head of technology, who had a continuous improvement mind-set and fully understood the importance of having an efficient survey process in the Assessment Department. With support from the associate superintendent, authorization was given for the purchase of state-of-the art scanning equipment and software; it was installed within the month.

Unfortunately, as is often the case, this good news was accompanied by some bad news: the expensive new equipment and software had equally expensive (and non-negotiable) training—over $2,000 per person, plus expenses. Such an expenditure for even one person was unimaginable in the department's existing budget. And even if the department had had the money, staff turnover made such an investment seem risky. So during the 2000–01 school year, the exciting new possibilities became limited to what staff were able to figure out by reading the manual and calling the company's less-than-responsive help line. They quickly figured out how to do straightforward scanning, but any subtleties or complexities created impossible challenges. In July 2001, CIA leadership paid for one secretary to attend four days of training, and, not surprisingly, the department's survey productivity increased dramatically. Cross-training of other Assessment Department staff in the near future will provide appropriate technical back-up.

Goal 3. The third goal was *to facilitate the existing school improvement process (SIP)*. This would be done by developing and implementing a system of school-based reporting to provide principals and teachers data about their schools and the results of SIP activities. Over time, the data would be customized; the principals and teachers would help create it. In Anoka-Hennepin, the central administration required building-level school improvement teams to submit school improvement plans by October 15 each year, which the appropriate area superintendents then reviewed. The plans included goals, indicators, strategies, a time line, and so on, and site teams were encouraged to review and reflect on outcomes at the end of the year. In practice, however, the plans differed widely, with little formal accountability. Some schools took on multiple goals and listed fifteen to

twenty strategies for reaching them; others wrote lists of ongoing activities (for example, staff attending workshops, meetings being held to discuss certain topics) with little visible pay-off in terms of outcomes. Most buildings listed increased reading or math test scores as indicators of progress, but the link between the strategy and the outcome was rarely explicit, that is, no explicit logic model was in place. In short, the district school improvement process was in essence a hollow evaluation process—goals were set, strategies implemented, and data collected but with little support or accountability for the results. An ECB assumption was that providing all available data to a school on an annual basis might prove useful for both building and central administrators.

As the assessment staff set about developing reports, however, they came to understand why these had never been done in the district. They quickly learned that student databases were purged over the summer, and data not stored prior to the purge might be lost. Other data were inaccessible without time-consuming work (for example, they were in teachers' grade books locked in school vaults). Even data the staff knew existed weren't necessarily readily accessible because one of the hard drives in the Assessment Department had crashed a year earlier and lost all of its memory, some of which was not backed up. Other data needed for the reports (for example, on attendance and dropouts) were available only in other departments in the central district office. Because staff in those offices were already fully occupied with their own jobs, they had little incentive to generate data for the Assessment Department, often putting the request in an "eventual" pile with other such requests ("We'll add it to our list of things to do"). A final and obvious challenge related to the packaging and use of data for school improvement was extremely practical: many of the staff in the district had old Apple computers and software, making it difficult to send them information electronically in any form that they could easily access or use. These problems, coupled with the department context of understaffing and overwork, meant that the building reports unavoidably received low priority.

Despite the many challenges, however, department staff (assisted by temporary clerical help) did succeed in creating data reports for all elementary schools in Fall 2000 and again in Summer 2001. During the 2000–01 school year, a technology facilitator began work on an electronic form for the district's school improvement plans. As the 2001–02 school year began, two forces were likely to drive the development of a workable school improvement planning process: (1) the state had instituted a process of mandatory school improvement for schools receiving Title I funds and not meeting the criteria of "Annual Yearly Progress," and (2) the associate superintendents instituted changes that would provide support to schools and then hold them more accountable for the results of their annual school improvement activities.

Goal 4. The fourth and final goal was *to create a network of people across the district to routinize classroom- or school-based inquiry*. This goal remained part of a vision but was not yet implemented. Some of the components to support the implementation were in place: (1) many committed professionals who were used to collaboration and were eager to make sense of their practice, (2) issues ripe for practitioner study (for example, the integration of Graduation Standards into the curriculum, (3) the shift to a balanced literacy program in grades K–2, (4) the status of gifted and talented students in the district, (5) technical support available in the Assessment Department, and (6) leadership eager to support such activities.

But major deterrents continued to block progress. First, there were too few days of substitute teacher support available and little money available to release teachers for this purpose. Second, and more important, there were other priorities for staff development more directly related to instructional changes the district sought. Although practitioner research could support the implementation of virtually any change (it is a cost-effective form of professional development), in the district context it lacked the significance of the major studies, the importance of the survey function, or the centrality of the school improvement process.

After two years then, the Assessment Department's ECB goals remained in place, and some progress was evident:

- Several studies had made the department's evaluation activities visible across the district.
- A viable model for structuring the major annual studies was in place.
- Over three hundred people had actively taken part in at least one evaluation study, and literally thousands had completed surveys.
- Central office administrators were refocusing attention on the school improvement process.
- The Assessment Department (once again fully staffed) was able to prepare student data for building use.
- New technology had upgraded the department's survey function.

In the words of the assessment facilitator who sometimes concluded presentations about the department: "Yesterday we stood at the edge of a great abyss. Today we take a giant step forward."

Lessons Learned

This case has focused on two years of a district's efforts to build evaluation capacity. To frame a discussion of what has been learned through the process, however, it is critical to emphasize the Assessment Department's core function: managing the district's standardized testing program. A dozen times during the academic year, the department office was filled almost to the ceiling with boxes containing that month's test; walking in the office became difficult, owing to the number of cardboard cartons arranged in

ways to make a fire marshal cringe. These tests ranged from a standardized test of music aptitude given each year to every 4th grader in the district (approximately three thousand students), to formal district writing assessments at two grade levels (approximately six thousand students), to state tests of reading, mathematics, and writing (which over thirteen thousand students take). Department staff attended to details of test distribution and administration because errors, although sometimes unavoidable, were never acceptable. Dire consequences attended certain mistakes, so when the time for each testing approached, the staff's focus did not waiver. Testing came first. In this context, program evaluation and ECB were additions to a departmental plate already chock-full.

Early in the coordinator's tenure in the district, she assumed that the Assessment Department would sponsor formal workshops on evaluation or at least on standardized testing and how to use test data for instructional and school improvement. In the context of Graduation Standards and the implementation of other instructional strategies, however, such workshops made little sense; only the coordinator saw evaluation training as an important priority. Her initial scan found formal program evaluation activities in four places in the district office: (1) staff routinely completed evaluation requirements for federal programs and other grants; (2) instructional facilitators conducted curriculum studies, as mandated by the state, that included surveys and other data collection; (3) the Vocational Education Department commissioned a follow-up study of the graduates of one of the four district high schools each year in order to have data for future grants, and (4) on a regular basis one of the central office administrators commissioned an outside consultant or agency to conduct a study for a specific purpose (for example, the Graduation Standards implementation, areas that voters might support in a tax levy to raise district funding levels, and parents' desires for desegregation planning). In a striking parallel to evaluation in developing countries, it was often outside forces—funding agencies or the state—that either mandated or inspired formal evaluation activity in the district.

In a system where priority items competed for limited resources, this approach to evaluation was sensible. District staff conducted mandatory evaluations or evaluations thought likely to result in practical outcomes. Given competing demands, the notion of evaluation for continuous improvement or as a way of life—although extremely appealing on the conceptual level—was necessarily a harder sell in practice. It is true that the district superintendent wanted to move in this direction and that the annual SIP modeled such a process, but the informal nature of SIP accountability to date meant that building site teams were typically responsible for activities rather than outcomes, creating the possibility of their going through yearly evaluation motions. In retrospect, the prospects for ECB were clear. In the language of the Joint Committee's standards, feasibility was paramount: if program evaluation processes were not made viable in this setting, they simply would not catch on.

Because program evaluation could not be the primary Assessment Department role, the tasks facing the R&E coordinator were threefold. The first related directly to program evaluation: it was the coordinator's job—and only hers—to conduct program evaluations, thereby providing evaluation data for use centrally by district leadership. A key question over time was what might motivate other district staff to voluntarily engage in evaluation activities while acknowledging that, absent a major input of resources, this would necessarily be a long-term effort. The coordinator's other tasks, therefore, related to ECB: (1) creating an infrastructure to support evaluation (ECB Goal 2, including facilitating data collection, helping staff to compile existing data, providing support for framing studies, and so on), and (2) simultaneously creating ongoing evaluation learning opportunities for district staff at all levels (ECB Goals 1, 3, and 4). For almost twenty years, the coordinator had framed her evaluation practice as an instructional activity, applying the four commonplaces of learning—teacher, students, curriculum, and context—to program evaluation (King, 1982). This framework provides a useful structure for discussing the ECB lessons learned in two years.

Teachers. In an ECB effort, the evaluator must become a teacher who purposefully structures evaluation and related activities and continuing collective reflection on these over time. Building capacity requires that the evaluator is integrally connected to people's work and alert to the programs that are potential objects for inquiry because they are key to district functioning, that is, to increasing student achievement. Interpersonal skills and the ability to identify and frame organizational issues are essential. Because relatively few districts have evaluators on staff, professional development is important, and quality program evaluation, especially when it uses participatory processes, can provide such training in situ while at the same time generating useful data.

There are other "teachers" in the ECB process: organizational leaders (administrators and opinion leaders) who make a visible commitment to evaluation, also known as the clout factor (King and Pechman, 1984). They can do this by providing verbal support in public situations (for example, when the superintendent hosted an evaluation meeting) and, more important, by serving as role models who study their own programs and practice. As examples, the superintendent commissioned the development of a districtwide parent satisfaction survey, and the associate superintendent for instructional support consistently "talked the talk and walked the walk" in her many initiatives. In the Anoka-Hennepin school district, staff knew that central office administrators expected to increase data-based decision making and that program evaluation was integral to that process.

Students. In this case, the "students" in ECB were district staff, and the long-term goal was that everyone become an evaluator to some extent, working to make sense of his or her own practice. For this to happen, the evaluator-teacher must provide support and guidance so that evaluation is

not just one more thing added to an already packed workload. Motivation and incentives are important, as staff will invariably ask (whether or not they say so out loud) what evaluation can offer them. Distrust of the evaluation process was surprisingly common, even in a district with good intentions; building-level staff nevertheless asked how the data would affect them, who would see the data, and whether these evaluation activities would be "one more thing that is dropped after a year or two." Trust can only be established over time, and in a large organization it remains a fragile commodity.

In addition, people must be able to get information they want and can use, not just what is readily available (for example, standardized test scores that may not relate to their questions of interest). They must also have the power to use evaluation results to make appropriate changes. This can be challenging in a large bureaucracy where someone may blithely propose a change in someone else's arena (for example, middle school students whose data clearly supported longer times for moving between classes and for lunch) or where certain situations are unchangeable, given existing resources (for example, expanding media center staffing in the elementary schools). To the extent that evaluation processes can be integrated seamlessly into ongoing activities (for example, setting up routine survey data collection and analysis), staff may even welcome them. The obvious danger in a system short on evaluation resources, however, is to create more demand than the Assessment Department could realistically handle. There is a delicate balance between creating demand and meeting expectations in a substantive and timely manner.

Curriculum. The curriculum in an ECB effort is the process of continuing evaluation itself, that is, teaching people the cycle and requisite skills of question framing, data collection and analysis, reflection, and planning for the next cycle. Formal training is not necessary because, for many, the evaluation process is fairly intuitive, and in school districts there are typically people familiar with it who are willing to take part. Training may well speed ECB, and targeted training of certain evaluation skills (for example, facilitation using cooperative learning techniques, focus group facilitation, involvement of minority stakeholders) could prove helpful.

The curricular goal is to institutionalize a shared commitment to program evaluation and a culture that includes the purposeful socialization of newcomers to the district. Two things became clear in the case study: (1) that every evaluation project, especially those involving people across buildings, became an instructional opportunity and (2) that participatory methods using small teams and work groups could provide nonthreatening, hands-on learning experiences. Having an established structure for these participatory studies is one indicator of district ECB. The Middle School Study used the management and participation structure that evolved during the Special Education Study, with the guidance of skilled outside facilitators but on a smaller and more cost-effective scale (thirty people rather than fifty and

using a salaried, in-house facilitator). This beta test with the same structure but using a shorter time line and no external resources demonstrated the capacity to conduct a participatory evaluation of a major initiative, although the department's being short-staffed and using new software that no one understood created overwhelming practical challenges.

Assessment must be part of the ECB curriculum, that is, establishing and tracking indicators, which in a district environment might include the following:

- *Inputs*—fiscal support (increased staffing and resources devoted to evaluation that are built into the budget and are not the first thing to be cut, even in dire times)
- *Processes*—the in-house capacity to collect and analyze data, create usable reports, and involve people in discussion; a framework for potential studies (a structure in place to generate new studies, support the process, and use data); an accessible process and support for anyone who wants to conduct a study; the ability to capture and compile the data for use (staff and data processing capacity); and purposeful socialization into the evaluation process
- Outcomes—increased demand for existing data; widespread involvement by numerous people in different types of studies; references to data during meetings; and ongoing use of data throughout the organization.

Context. In the context of a school district in the United States, at least for the foreseeable future, program evaluation will remain secondary to administering standardized tests and reporting their results. As noted, standardized testing was the core activity for the Assessment Department, demanding resources and a staff schedule with deadlines that allow little flexibility. The one secretary who worked on survey data was pulled off evaluation work if she was needed for a testing task. In other words, the standardized testing program easily—and typically—overwhelms the program evaluation function (King, forthcoming). To build evaluation capacity in a district setting, ECB practitioners cannot ignore this context, especially as the stakes for testing increase in a context of highly visible public accountability.

School districts, however, do need access to evaluation expertise, both internal and external. Structural conditions can support the development of evaluation capacity: (1) time to collaborate including, when possible, being physically together in an environment separate from the interruptions of the school day; (2) meaningful incentives for participation; (3) opportunities for reflection; and (4) effective communication (for example, the district practice of key messages summarized from important meetings and standardized "action minutes" that document meeting content, decisions, and who will do what next). In a district context, the Joint Committee Standards' categories of utility and feasibility highlight the fact that ECB must be both useful and workable, targeting programs that are central to

improved student achievement and making visible use of data over time.[1] "Low hanging fruit"—visible, easily made changes—provides evidence that the evaluation process can lead to improvement (for example, reduced Graduation Standards requirements after teacher survey data recommended that, or heightened parent communication in Special Education when data documented such a need).

Final Reflection: To ECB or Not to ECB?

Some people may read this case study and offer condolences for two years of work with a seemingly unhappy ending. People did not deflect efforts at building evaluation capacity; they don't need to because the context itself makes it exceedingly difficult (for example, next year's budget situation). What's missing (for sure) is time, the ability to get people together, incentives for participation, and even better communication, which the size of the district impedes.

But such a reading would be incorrect. Has the district built and sustained sufficient and appropriate evaluation capacity for its needs and wants? This case provides an example of the definition by Stockdill, Baizerman, and Compton (Chapter One, this volume) of ECB as "intentional work to constantly co-create and co-sustain an overall process that makes quality evaluation and its uses routine in organizations and systems." Even in light of a dire budget situation—including the potential for a second year of budget cuts, perhaps of as much as $10 million—the ECB effort in Anoka-Hennepin has not ended; top administrators and staff are committed to its continued development. A district staff person once commented, "In a big system, you can only do big things." To commit to ECB is a big thing, and Anoka-Hennepin Independent School District 11 is happily two years closer to this goal.

Note

1. In a district setting, the JCS category of propriety is a given, the sine qua non of any program evaluation effort.

References

King, J. A. "Studying the Local Use of Evaluation: A Discussion of Theoretical Issues and an Empirical Study." *Studies in Educational Evaluation,* 1982, *8,* 175–183.

King, J. A. *Proposal for an Expanded Assessment Department.* Coon Rapids, Minn.: Anoka-Hennepin Independent School District 11, 2001.

King, J. A. "Evaluating Educational Programs and Projects in the USA. In D. S. (ed.), *International Handbook of Educational Evaluation,* forthcoming.

King, J. A., and Pechman, E. M. "Pinning a Wave to the Shore: Conceptualizing School Evaluation Use." *Educational Evaluation and Policy Analysis,* 1984, *6*(3), 241–251.

Maruyama, G., Freeman, C., Hole, D., Frederickson, J., and Springis-Doss, R. *Study of the Four Period Schedule for Anoka-Hennepin No. 11.* Minneapolis, Minn.: Center for

Applied Research and Educational Improvement, College of Education and Human Development, University of Minnesota, 1995.

Patton, M. Q. *Utilization-Focused Evaluation.* (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.

Preskill, H., and Torres, R. T. *Evaluative Inquiry for Learning in Organizations.* Thousand Oaks, Calif.: Sage, 1999.

JEAN A. KING is an associate professor in the Department of Educational Policy and Administration, College of Education and Human Development, University of Minnesota, Minneapolis.