DEVELOPING A SPECIAL EDUCATION ACCOUNTABILITY FRAMEWORK USING PROGRAM EVALUATION

A. Sidiq Ali
Paul Favaro
Peel District School Board
Mississauga, Ontario

Abstract: Accountability has been at the forefront of standards-based educational reform. Yet, as new-age accountability is upon us, driven in the vehicle of large-scale assessment by policy decisions, research indicates that within these accountability systems are organizational and technical barriers to educational improvement. What is presented in this article is a local-level accountability framework focused on special education programs in the context of a large, predominantly urban school district in Southern Ontario. Involving principles of participatory program evaluation, this framework looks at building organizational capacity for bureaucratic and professional accountability in efforts to overcome barriers to educational improvement through accountability.

Résumé : L'imputabilité a été à l'avant-plan de la réforme éducative basée sur les normes. Cependant, étant donné que nous sommes dans un nouvel âge d'imputabilité, guidée par les règles d'action de l'évaluation à grande échelle, les recherches indiquent que ces systèmes d'imputabilité constituent des obstacles organisationnels et techniques à l'amélioration de l'éducation. Ce qui est présenté dans cet article est un plan d'imputabilité local, basé sur des programmes d'éducation spécialisée dans un grand conseil scolaire à prédominance urbaine, du sud de l'Ontario. Selon les principes de l'évaluation d'un programme participatif, ce plan vise à construire la capacité organisationnelle pour l'imputabilité bureaucratique et professionnelle, afin de surmonter les obstacles à l'amélioration de l'éducation par l'imputabilité.

Accountability has been, and will continue to be, at the centre of standards-based educational reform (National Center for Educational Outcomes, 2005). Accountability has been defined in various ways, but typically refers to an individual or group taking re-

Corresponding author: A. Sidiq Ali, c/o Assessment & Accountability, Peel District School Board, 5650 Hurontario St., Mississauga, ON, L5R 1C6; <asali@rogers.com>

sponsibility for the performance of students on educational outcomes, such as achievement, graduation rates, and attendance, for example. The success of accountability systems or frameworks (we use these words interchangeably) is measured by marked improvement in educational outcomes for all students (National Center for Educational Outcomes, 2005). The existence of accountability systems, however, is no guarantee of improved student performance, as several barriers to the success of accountability frameworks have been well documented (e.g., Baker, Linn, Herman, Koretz, & Elmore, 2001; CRESST, 2002; Earl, 1995; Linn, 2000; O'Day, 2002).

In this article, we present a case example of the application of collaborative evaluation approaches to building an accountability framework, which includes program evaluation. Through these approaches, we hope to increase the accountability and evaluation capacity of our organization, with an aim to ultimately change organizational behaviour regarding accountability-driven improvement. To set the stage for the case example, we first provide a brief description of the current state of accountability in education in general and then in the context of special education. We then highlight barriers to the success of accountability systems in improving outcomes for all students. Next, we present a conceptual framework for overcoming many of the outlined barriers through the use of both a strategic plan regarding capacity-building and some tactical ideas, including the use of collaborative evaluation approaches. We use these concepts to underpin our Special Education Accountability Framework (SEAF), a multi-faceted system of data collection, analysis, reporting, and improvement planning geared specifically to special education programs and services. Finally, we will discuss some of the initial consequences of the application of this framework in our organization—the Peel District School Board (PDSB)—a large, complex, and dynamic district school board located on the western border of Toronto, Ontario.

THE CURRENT STATE OF EDUCATIONAL ACCOUNTABILITY

Educational accountability in some form or another has been in existence since the early days of schooling when teachers were accountable for the manner in which they dressed and groomed themselves; student accountability in the form of test scores and report cards has been a facet of education for almost as long. Fiscal accountability for education came to the forefront in the 1960s with substantial increases in government funding for social programs and the emergence of program evaluation (O'Day, 2002). But when we speak of educational

accountability today, we think of the new age of accountability where student outcomes, often measured by large-scale assessments, are the indicators of educator and system performance, where the focus is on the school as the basic unit of accountability, and where public reporting of student achievement is associated with rewards and sanctions for schools (Elmore, Abelmann, & Fuhrman, 1996).

When we define this new age of educational accountability, we generally refer to an organization (like a school district, or even a school) or groups of individuals within an organization taking responsibility for the performance of students on measures of educational achievement or other types of educational outcomes such as attendance and graduation rates, for example. Outcome data collection and reporting are necessary acts for accountability but do not sufficiently describe what accountability is. Earl (1998) makes a clear distinction between data collection, or *accounting*, and *accountability*, which she describes as forward thinking, using the data "to make judgments about quality, about how good is good enough and most importantly, about how to make changes that will enhance and extend student learning" (p. 21). So, the focus of educational accountability is not so much the gathering of data, but the thought about what the data mean, and how to use the data to make improvements. Further, even a carefully managed system of accountability data reporting will not improve student achievement. The essence of new-age educational accountability is not in collecting and reporting data, but in action toward improvement (Earl & Lafleur, 1998).

New-age accountability systems generally include some similar characteristics: they (a) develop or use challenging content and performance standards that focus on student learning; (b) emphasize measurement of student achievement as the basis for accountability; (c) involve technically complex apparatus as the mechanism for evaluating schools; and (d) introduce consequences, either rewards or sanctions, as incentives for improving achievement (Linn, 2000). Yet, for all their similarities there are key differences in accountability systems from jurisdiction to jurisdiction: (a) not everyone agrees on the definition of accountability, therefore many accountability frameworks are just elaborate ways of collecting student achievement data from large-scale assessments; (b) jurisdictions differ in how they assess student achievement; (c) accountability systems differ in the manner in which they address fairness issues from a technical perspective; and (d) each jurisdiction defines its consequences attached to accountability differently (Elmore et al., 1996).

Accountability and Students with Special Education Needs

Especially problematic for stakeholders involved with programs for students with special education needs is Elmore et al.'s (1996) second point—jurisdictions differ in how they assess student achievement. Until recently, many accountability systems excluded students with special education needs. Up to the year 2001, it was common for American states to leave students with special education needs completely out of achievement assessments for accountability requirements (Thurlow, 2004). Blanket exclusion of students in special education from large-scale accountability-driven assessments initially resulted in increased special education prevalence, de facto disincentives for program improvement in special education, and spurious improvement results for schools (Thurlow & Krentz, 2001).

To ensure that educators are held accountable for the achievement of all students, and to improve achievement outcomes for all students, it is considered imperative to include all students' scores in accountability calculations (Thurlow, 2004). So, in this decade, as a result of the No Child Left Behind legislation in the United States (U.S. Congress, 2001), American state accountability systems have been more inclusive of students with special education needs, allowing these students to take accountability-driven achievement assessments through traditional methods, using accommodations or alternative methods of assessment (Thurlow, 2004). Inclusion of students with special education needs in standards-based accountability systems facilitates more instructional attention for these students and increases expectations for these students (Thurlow & Krentz, 2001). In Ontario, the Education Quality and Accountability Office's (EQAO) accountability system allows for students with special education needs to take provincial assessments with accommodations, thus increasing their participation in the system (Education Quality and Accountability Office, 2007a).

Accommodations alter the test materials or procedures so that a student's knowledge and skills, rather than a student's disabilities, are assessed (Thurlow, 2004). Common types of accommodations include extended time, repeating directions, individual settings, reading the test to the student, Braille or translated editions of the test, and verbatim scribing of written answers in tests of reading. There are dozens more accommodations that have been documented as used in accountability-driven assessments (Krentz, Thurlow, & Callender, 2000). In Ontario, for example, an accepted accommodation for a provincial assessment is *any* one that is documented on the student's

Individual Education Plan (IEP) and is consistent with regular classroom assessment practices used for the student (Education Quality and Accountability Office, 2007a).

However, some accommodations may alter test constructs and thus compromise the validity of interpretations associated with the test, thereby making accommodated-student test scores incomparable with those students who took the test under standardized conditions. A review of research by Sireci, Scarpati, and Li (2005) shows that two common accommodations, extended time for most tests of literacy and reading the test to students for mathematics assessments, have been shown to not alter the construct of tests, while increasing test performance of students in special education. Consistent conclusions about other common accommodations could not be made (Sireci et al., 2005). Nonetheless, carte blanche accommodations, and accommodations on their own, are not the answer to holding educators accountable for the scores of all students, and to improve achievement outcomes for all students. Accountability systems in special education must be able to inform educators, policy-makers, parents, and advocate groups about the effectiveness of programs and include students with special needs in standards-based instruction and assessment (Mertens & McLaughlin, 2004).

A Theory of Action for Educational Accountability Systems

In order for any assessment-based accountability system to be successful in improving educational outcomes for all students, a theory of action for educational accountability systems—a course from assessment to improvement—needs to be followed (Baker & Linn, 2002). Baker and Linn have described their seven-step action theory as follows:

1. Results of student assessments reported to the organization (school districts/schools) and the public (parents/community members) are accurate.
2. Results of student assessments are validly interpreted by the organization and public.
3. Organizational leaders are cognizant of both the concept of accountability and issues of data quality, are themselves willing to act based on the data, and can motivate action by organization members.
4. Alternative actions to improve the situation are known and available to the organization.

5. Organization leaders and other members (teachers and support staff) possess the requisite knowledge to apply alternative methods.
6. The selected action(s) is/are adequately implemented.
7. The action(s) selected will improve subsequent results. (Baker & Linn, 2002, p. 1)

Adherence to the aforementioned theory of action may help facilitate the success of new-age accountability systems in improving education of all students. But the flow through each step of the theory of action for accountability systems can be impeded by barriers that limit the potential for accountability-driven improvement.

BARRIERS TO ACCOUNTABILITY-DRIVEN IMPROVEMENT

O'Day (2002) has described three overarching barriers to accountability-driven improvement: (a) the school is the unit of intervention, yet the individual is the unit of action; (b) external (central administration) control seeks to influence internal (school-level) operations; and (c) information, and its flow within the organization, while essential to improving schools, is problematic. These barriers highlight the complexities involved in accountability-driven improvement within large organizations like district school boards. O'Day (2002) elaborates on the first barrier, stating that individual teachers, administrators, and parents must in some way change what they are doing in efforts to change what students do. The second barrier points to the hierarchical structure of school boards, as individual team members (teachers) operate within schools, and schools operate within districts, each with its own level of administrators. She argues that new-age accountability approaches seek to influence from the outside (the central administration level) what goes on inside schools. For the third barrier, O'Day (2002) stresses that bureaucracy from the central agency often does not affect the core, as Elmore (1996) has termed it, of teaching and learning. For example, central administration policies often do not penetrate the privacy of classroom practice and normative structures within schools, which are determining factors not only in the implementation of accountability-based actions (O'Day, 2002), but, more importantly, in the school's improvement (Fullan, 1990) and overall effectiveness (Stoll & Fink, 1994).

Even if organizational barriers to accountability-driven improvement can be overcome, barriers involving principles of psychometrics also require overcoming if accountability systems are to be successful.

For instance, the research on accountability-driven school reform provides little evidence that test-based accountability of schools produces better student achievement (Mehrens, 1998). O'Day (2002) suggests that the lack of accountability systems' positive effect on student achievement is largely due to the validity, periodicity, and specificity of student outcome measures.

Validity in accountability systems affects two interrelated aspects of the system: (a) the validity of the outcome measure, most likely a measure of student achievement on which improvement is based; and (b) the validity of the measures of goals (or standards) for the system that are crucial to the success or lack thereof for account-ability systems. For example, if an assessment does not measure what it purports to measure, it could divert the organization's atten-tion and efforts away from the goals of the accountability system, which are tied to the assessment results (O'Day, 2002). By specificity, O'Day (2002) refers to the level of information offered by the student achievement outcome measure. If, for example, an annual test of achievement is used as the outcome measure, she postulates that such an assessment is too distant from the complexities of instructional inputs for the teacher to make reasonable attributions of causality, and improvement planning is hindered as a result. As for periodic-ity, O'Day (2002) suggests that the annual achievement test should be supported with more focused and periodic assessments, closely aligned with the goals of the accountability system to enhance its desired effect on the improvement process.

Still, student outcomes in accountability frameworks tend to be measured by large-scale (district, state/provincial level) assessments that generally are not well-aligned to curriculum expectations (Linn, 1998) or instructional strategies (O'Day, 2002). Further, even when assessments are closely linked to the curriculum and instructional practices, valid interpretation of the data for improvement planning may not come easily. It has been suggested that many key decision makers at the upper echelons of educational organizations, who through the theory of action in accountability systems are required to act based on data, while eager for accountability data, in fact lack sophistication in data interpretation (Linn, 1998, 2003). This lack of sophistication or data illiteracy hinders educational organization leaders' ability to comprehend what the data are telling them, and provides them with little or no basis by which to judge the quality of the data (Earl, 1995, 1999), erecting another barrier to the account-ability-driven improvement process.

It seems that attempts at overcoming organizational and technical barriers to accountability-driven improvement can be a daunting task. Proposed ideas for making accountability systems successful for all students require careful implementation and rigorous study. Our organization is in the midst of carefully implementing an accountability framework with specific focus on special education programs. We have studied the relevant literature and have planned approaches to overcoming barriers. We have chosen to start with a strategic focus on building organizational capacity to understand what accountability for all is, and how we as an organization can make better data-driven decisions, because the research also suggests that, despite a lack of technical cognition among senior administration and others in school boards, there is a real appetite in educational organizations, like ours, to know how accountability data can lead to improvement (Linn, 1998). Next, we present a conceptual framework, which serves as the basis of our approach to overcoming barriers to accountability-driven improvement.

OVERCOMING BARRIERS—THE STRATEGIC PLAN: BUILDING ORGANIZATIONAL CAPACITY FOR ACCOUNTABILITY-DRIVEN IMPROVEMENT

Building organizational capacity for accountability-driven improvement is the goal of a strategic plan that we trust will allow the organization and accountable units within to accept, participate, and develop in the improvement process. Part of building this capacity is establishing trust within the organization. For accountability to be successful in improving student outcomes, it must be seen by stakeholder groups within the organization as a course for improvement, not as an agenda for surveillance (Earl, 1999). Organization members must view themselves as participants in the accountability process with a common goal—the improvement of student achievement. Further, the relevance of accountability systems to the educational organization they serve depends on the extent to which stakeholders' interests and values are reflected in them (Nagy, Demeris, & van Barneveld, 2000). Additionally, *accountability theory* (Tetlock, 1998) suggests that organization members are more likely to strive to meet high standards of accountability systems when the standards are seen as attainable, when the procedures for setting them are judged to be fair, and when those standards are justifiable in terms of the organization's survival. Also, in building organizational capacity for accountability-based improvement, the organization's data synthesizers must have the collective ear of the senior leadership, the deci-

sion-makers, to aid these leaders in the development of their ability to judge and use data (Patton, 1997).

Capacity-building within the organization to use accountability as an effective vehicle for improvement stems from the goals and values of the organization, personified in the senior leadership. Accountability can be successful only in a results-oriented organizational culture fostered by leaders, which model for other members the serious use of accountability-based data to make decisions that lead the organization toward its goals (Patton, 1999). Further, in educational accountability systems, the goals of the organization are tied to improved student achievement. Improving teacher practice is directly associated with improving student outcomes. Accountability systems will foster improvement in achievement to the extent that they generate and focus attention on data relevant to teaching and learning, motivate organization members to use those data in decision-making, and build the knowledge base necessary for interpreting data to improve practice and allocate resources to improve practice (O'Day, 2002). We intend to achieve our strategic goal, to build organizational capacity for accountability-driven program improvement, through the execution of specific tactics, which include the application of collaborative approaches of program evaluation.

OVERCOMING BARRIERS—TACTICAL IDEAS

The Use of Collaborative Approaches of Program Evaluation

While the primary goal of accountability systems is to improve education, these systems have not traditionally provided information to the school community about *what to do and how to improve* performance; program evaluation can help in this matter. Often, accountability systems are put in place with specific student outcomes in mind, but programs purported to affect the outcomes have not been defined and the logic of the program not articulated (Ryan, 2002). Wholey (1994) and Weiss (1998) emphasize that until program definition and theory are clearly articulated, programs are not ready to be evaluated and indicators for accountability purposes cannot be specified. Developing program theory helps provide a mechanism for reflecting about what to do to improve student outcomes. Further, once program logic is specified, then accountability outcomes can be defined and broadened through the use of mixed method evaluation (Mark, Henry, & Julnes, 2000). This proposed use of program evaluation helps fulfill needs with the technical aspects of defining contexts, inputs, pro-

cesses, products/outcomes, and outputs for improvement planning (see Scheerens, 1990; Stufflebeam, 1983). But these technical issues could be addressed without collaborative evaluation approaches. Our rationale for employing collaborative evaluation is that while evaluation helps define accountability outcomes and a course for improvement, collaborative approaches to evaluation help fulfill the greater need within our organization, which is building capacity for accountability-driven program improvement.

We subscribe to Patton's (1999) view that programs are embedded in larger organizational contexts, and improving programs may be linked to and even dependent on changing the organization of which they are a part. The process of program evaluation supports changes in organizations by getting stakeholders involved in what Patton (1999) describes as "reality testing"—getting people to think empirically, focusing on "specificity and clarity" while teaching them about the methodology and utility of data-based decision-making (p. 94). Using a participatory approach to evaluation (see Cousins & Earl, 1992, 1995) engages key stakeholder groups within an educational organization, such as decision-makers (senior administration) and program managers (school administrators and teachers), in the evaluation process. Increased participation can lead to empowerment of participants (Morris, 2002) and greater utilization of evaluation results (Patton, 1997).

Empowerment of stakeholders in program evaluation is a collaborative approach in its own right. Typically, empowerment evaluation is the use of evaluation methodology and findings to foster improvement self-determination (Fetterman, 1999). Empowerment evaluation is designed to help people help themselves and improve their programs using forms of self-evaluation (Zimmerman, 2000). But since the program evaluations proposed here will take place in a school board, a naturalistic setting where all the ideals for participatory (see Cousins & Earl, 1992) and empowerment evaluation (see Fetterman, 1999) may not exist, we as professional evaluators must continue to foster the organization's capacity for accountability-driven improvement by acting as educators, mediators, and facilitators (Morris, 2002), engaging stakeholder groups in evaluation participation and helping them set their own courses for improvement. The transition between participatory and empowerment evaluation can be a difficult one, but studies do show that stakeholder participation in the evaluation process can lead to stakeholder empowerment and self-determination (Earl, Ali, & Lee, 2005; Hunter & Gambell, 2000; Lee, 1999), thus

minimizing the need for or reliance on central/senior administration in the improvement process.

We hypothesize that increased participation and empowerment in the accountability-driven improvement process while addressing O'Day's second barrier to the success of accountability systems—external (central/senior administration) control seeks to influence internal (school) operations—will help motivate program managers (school administrators and teachers) into action, addressing the first of O'Day's (2002) barriers to the success of accountability systems focusing on the individual as the unit of action. We recognize that so much of any school improvement, based on accountability or not, requires the motivation of teachers. Simply, the teachers must buy in, or be internally motivated to commit, to any improvement initiative (Fullan, 2001; Stoll & Fink, 1992). For accountability-based improvement, we think the use of collaborative evaluation approaches will promote a professional accountability model, one that O'Day (2002) describes as focusing on incentives for improvement, with particular emphasis on the intrinsic motivators that bring teachers into the profession in the first place—a commitment to students and an identity as an educator. Too often, accountability is based solely on a bureaucratic model, which often focuses on negative incentives, undermining natural innovation and risk-taking by teachers to meet the needs of their students. Bureaucratic accountability in schools with disadvantaged populations diverts attention to organizational survival rather than teaching and learning (O'Day, 2002).

If greater professional accountability motivates improvement, it means the accountable units, such as program managers (teachers), take responsibility not just for giving an account and justifications of practice, but also for reducing any discrepancies between current and expected performance. But, in order to make up any shortfalls in performance, teachers and schools must have the capacity (skills, knowledge, and resources) to develop more effective practices (Robinson & Timperley, 2000). Ultimately, though, it is the central administration or the bureaucracy that provides resources to build capacity for teachers and schools to improve. So, in order to motivate accountable units to improve, there must be a melding of both professional and bureaucratic accountability models (O'Day, 2002). We believe that, in our organization, the use of participatory and empowerment evaluation with the professional evaluator serving as a link between central administration and schools and acting as facilitator and mediator allows such a melding to transpire.

Program evaluation, however, plays another role in the success of accountability systems, as once accountability systems are in place and data-collection cycles have been completed, evaluation of the systems are necessary. Using program implementation theory, treating the accountability system as a program, we can test the notion that if program activities (the accountability system) are conducted as planned, the desired results—improved student performance—will be achieved. The use of program evaluation and specifically implementation theory allows us to study intended and unintended consequences of the accountability system to refine and improve it (Weiss, 1998).

Within the proposed use of collaborative approaches of program evaluation as an enabler of accountability-driven improvement, the notion is that O'Day's (2002) first two barriers to improvement through accountability—the first one focusing on the individual as the unit of action and the second one focusing on the amount of central administration control—will be addressed by increased participation and empowerment, especially of program managers. The third barrier—focusing on information and its flow within the organization—will be addressed through the use of another tactical idea: the use of multilevel statistical models.

The Use of Multilevel Statistical Models and the Value Added Approach

Currently, local education authorities in the United Kingdom (U.K.) and various jurisdictions in the United States, such as the states of Tennessee, North Carolina, and South Carolina, and the Dallas Public School District use multilevel statistical models with contextual, input, and process variables to provide less biased estimates of student outcomes. These models are better able, when compared to single-level models, to adjust for factors that schools do not influence over a given school year, such as students' socio-economic status (SES) and prior achievement. This type of statistical adjustment, originating in the U.K., is referred to as the value added approach (e.g., Gray, Jesson, & Jones, 1986; Willms, 1987) and is deemed a more fair way of presenting accountability data (Goldstein, Huiqi, Rath, & Hill, 2000; Willms, 1992).

A value added approach using multilevel statistical models tends to be a rather sophisticated process, and there is still much debate over what variables the basic model should employ. For accountability purposes, though, change scores when an achievement test is the outcome variable, rather than the current score, are recommended

for greater reliability. Also, longitudinal or quasi-longitudinal student data collection allows for more valid interpretation of student outcomes than analysis of successive groups of students at one particular grade (Linn, 2003, 2004). Contextual variables such as socio-economic status, input variables such as teacher qualifications, and process variables such as opportunity to learn are necessary components of these models as predictors of the outcome (Sanders & Horn, 1994; Scheerens, 1990). Articulation of program logic models will help identify data sources for the aforementioned variables, while providing direction for improvement planning (Ali, 2004; Creemers & Reezigt, 1997). Collecting these data may seem difficult, but, in Ontario for example, the EQAO Indicator Framework shows school boards where to find much of these data (Education Quality and Accountability Office, 2005). A great deal of the data necessary to employ multilevel statistical models in a value added approach are available to school boards through their student information systems (SIS). We acknowledge that much technical expertise within the organization or available to the organization is necessary to utilize these models in the improvement planning process.

Nonetheless, arguments have been made that the use of multilevel statistical models for the value added approach, because of statistical adjustments for SES and prior achievement, institutionalizes low expectations for poor, minority, low-achieving students (Elmore et al., 1996). We suggest that institutionalization of low expectations is not automatic; rather it is a choice policy makers in school districts make. For example, a school district could use the value added approach and still have the same expectations or standards for all schools on gain scores over time, which would not set lower expectations for schools with disadvantaged populations. There is, however, one caveat here. The expectations for gain scores must be reasonable, based on some "existence proof"—for example, that good teaching practice, or use of a given program will raise achievement scores a proven amount regardless of environment (Linn, 2003, p. 5).

Focus on Validity

The use of multilevel statistical models for accountability purposes may improve the validity of interpretations associated with accountability data (Linn, 2003), thus improving the quality of information that flows through the organization, helping to overcome the third barrier to accountability-driven improvement presented by O'Day (2002). Also, recall the theory of action for accountability systems

(Baker & Linn, 2002)—validity of interpretations is central to accountability. Put simply, validity is dependent on the adequacy of the congruence of what is reported and the activity it is meant to represent (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Baker and others (2001) have proposed standards for accountability systems to address validity concerns: (a) accountability data should come from multiple sources; (b) weighting of different system components, different test content, and different sources of data should be made explicit; (c) data elements are required that allow for interpretations of student, institution, and administrative performance; (d) expectations should be made public and understandable for all participants; and (e) performance of all students, including subgroups that historically have been difficult to assess (special education populations), should be included. Still, validity may be jeopardized by inaccurate reporting, attributable to technical deficiencies in assessment procedures, or by the fact that accountable units (schools and their teaching staffs) are motivated more by the desire to please or the fear of displeasing rather than by accuracy (Baker & Linn, 2002).

We believe that the conceptual framework presented will help our organization overcome the barriers to accountability-driven improvement articulated by O'Day (2002) and adhere to Baker and Linn's (2002) *theory of action for educational accountability systems*. Essentially, focused on capacity-building in our organization for improvement through accountability, the proposed uses of collaborative evaluation approaches and the use of multilevel statistical models compose the core of our Special Education Accountability Framework (SEAF). The SEAF in our organization is designed not only to overcome barriers and facilitate successful accountability action, it is designed to bring about change in the organization.

THE SPECIAL EDUCATION ACCOUNTABILITY FRAMEWORK IN THE PEEL DISTRICT SCHOOL BOARD

The Need for Conceptualizing a Special Education Accountability Framework

In our organization, the Peel District School Board (PDSB), various stakeholder groups found that the province's accountability system (EQAO assessments) did little to identify strengths and areas of need within special education programs and services. These stakeholders

found the EQAO system to be deficient in two ways. First, many special education programs are skills-based, and skill development for students in some of these programs are not congruent with the provincial assessments. For example, literacy skills may not be the main focus for particular programs. A program for students with Autism Spectrum Disorders may focus more on social skills, and a program for students with developmental challenges may focus more on independence skills. Second, special education programs are more expensive to operate than most mainstream programs, and in a time of funding cutbacks, a greater need for information at trustee, senior administration, and program manager levels regarding program effectiveness arose. A collective call for greater accountability at the program level was made.

With overcoming barriers to accountability-driven improvement in mind, we set out to design an accountability framework for the special education programs and services within our large educational organization. To make the SEAF work initially and be sustainable in the future, we envisioned that changing organizational behaviour regarding accountability would be required. Our approach, involving the application of collaborative evaluation methods to build an accountability framework (which includes program evaluation) and increase the evaluation/accountability capacity within our organization to ultimately change organizational behaviour regarding accountability, is grounded in the aforementioned conceptual framework and our previous experiences at the school board. These experiences tell us that "accountability systems are most effective when they are implemented at the local school level, issue-specific, congruent with the school's mission/goals/planning process, and designed to improve teaching and learning, rather than giving priority to narrowly defined outcomes" (Favaro, Fine, Norman, & Bell, 1998, p. 39). Such a view of accountability is different than the large-scale district-wide accountability models we see in the United States and even in Ontario with EQAO assessments. We will describe the components of our special education accountability framework, after we first set the provincial and school board contexts for it.

The Provincial Context

In the mid 1990s the government of Ontario created the Education Quality and Accountability Office to oversee a province-wide assessment program for accountability purposes. Currently, students in publicly funded schools in Ontario must take reading, writing, and

mathematics assessments in Grades 3 and 6, a mathematics assessment in Grade 9, and a literacy (reading and writing) assessment in Grade 10. In 2000, the province's Education Improvement Commission (EIC) reports took the accountability movement further, urging the provincial government and school boards to work together to develop a more comprehensive approach to accountability, stating, "We believe the development and implementation of a comprehensive accountability framework is the single factor that would have the greatest impact in improving our education system and student achievement" (Education Improvement Commission, 2000, p. 10). The Ontario EIC report goes on to state that such an accountability framework would include a system of indicators of student achievement and standards for professional practice.

In Ontario, however, the accountability system (EQAO) is not bound to legislation, as state accountability systems in the United States are (see U.S. Congress, 2001, 2004). There are no mandates in Ontario to include students with special education needs in the provincial accountability system. Further, there are no provisions in Ontario for linking student IEPs to the requirements of the EQAO assessments, or for ensuring that remedial instruction is available to allow students in special education to have reasonable opportunities to succeed (Jordan, 2001). So, in the context of a movement toward greater educational accountability in Ontario, without being mandated by legislation for inclusion of students in special education in the EQAO assessments, our organization took a bold step to help improve achievement for all students, while working with the province's accountability system.

The School Board Context

The PDSB is Canada's second largest school board, serving a predominantly suburban area west of Toronto—Canada's largest and most ethnically diverse city. In September 2007, the PDSB will serve 225 schools totaling some 153,000 students, with roughly 9% of those students requiring special education programs and services. The board's 2005–2006 expenditure budget was $1.06 billion, and the budget for special education programs and services was roughly one-tenth of the total expenditure (Peel District School Board, 2007). These programs serve K-12 students with developmental challenges, learning disabilities, behavioural issues, speech and language needs—and more. They also include enhanced learning classes for

gifted students and secondary school vocational programs. Being a publicly funded school board in Ontario, all of the PDSB's students must write the province-wide EQAO assessments administered in Grades 3, 6, 9, and 10. Schools within the board must use their EQAO assessment data as a key component of their school success (improvement) plans.

The PDSB does well at including close to all students in EQAO accountability assessments. Typically, just 1% to 3% of PDSB students are exempted from EQAO assessments for various reasons, including their special education needs (Peel District School Board, 2006a, 2006b, 2006c). Many students, however, receive accommodations when writing the EQAO assessments. For example, in 2005–2006, 12% of 10,122 Grade 3 students and 11% of 10,478 Grade 6 students received one or more accommodations during their EQAO assessments (Peel District School Board, 2006a). Similarly, in 2005–2006 some 13% of 9,612 Grade 9 students who wrote the mathematics assessment did so with accommodations (Peel District School Board, 2006b). Also in 2005–2006, of the 9,233 first-time eligible students writing the Grade 10 literacy test, 11% did so with one or more accommodations (Education Quality and Accountability Office, 2007b; Peel District School Board, 2006c).

Recalling that some accommodations may compromise the integrity of the assessments, and that many types of accommodations are permissible, EQAO assessment results for students with special education needs, while generally accepted within the organization with their limitations, fall short of informing senior administrators and program managers of strengths and needs of particular programs. Also, given the frequency of the EQAO assessments (in only 4 of 12 grades), and the call for greater program-level accountability, we endeavoured to supplement the province's EQAO accountability framework with our own, aiming at: (a) informing our program managers and senior administration as well as parents and advocate groups about the effectiveness of our programs; (b) including students with special education needs in standards-based instruction and assessment as much as possible; and (c) improving our special education programs and services. Our Special Education Accountability Framework is composed of three key components and incorporates the collaborative evaluation approaches discussed previously because, to make it work, we need to build the organization's capacity for accountability-driven improvement.

Three Key Components of the Special Education Accountability
Framework

The Special Education Accountability Framework for the PDSB in-
volves three components for data collection, reporting, and forward
thinking, all centred on building organizational capacity for accounta-
bility-driven improvement utilizing the tenets of participatory program
evaluation (see Cousins & Earl, 1992) and empowerment evaluation
(see Fetterman, 1999; Zimmerman, 2000). The three components are:
(1) Systemic Key Performance Indicators, (2) a Standards-based Re-
view Process, and (3) a Special Education Program Evaluation Proc-
ess. The order of each component indicates the depth of investigation,
with (1) skimming the surface and (3) delving deeper into programs
and services. All three components are designed to inform the program
improvement process, either individually or collectively, using mul-
tiple sources of data from multiple perspectives, including the EQAO
assessment data, where appropriate. Figure 1, an adaptation of Little
and Houston's (2003) research into practice models, depicts the SEAF
in the continuous program improvement process.

One basic function of the SEAF is to gather contextual, input, proc-
ess, and student outcome data on identified special education pro-
grams and services. Programs are identified by senior administration
through a needs assessment process that utilizes SEAF data. After
data collection, analysis, reporting, and improvement planning, an
instructional practice is identified to address program needs. Program
managers, content experts, and professional support staff such as
psychologists and speech-language pathologists working together in
committees investigate and select a research-based effective instruc-
tional practice to attend to the program need(s). Program staff are
selected and staff development initiatives facilitate training in the
instructional practice. However, the temporal sequence dictating in-
volvement of a particular component of the SEAF varies. For example,
indicator data could be used to identify student outcomes in a particu-
lar special education program that will be targeted for improvement
through the implementation of research-based instructional practices.
Indicator data could also be applied to monitor the effectiveness of
newly implemented instructional practices. Throughout the continu-
ous improvement cycle, the SEAF components provide data support
for each improvement initiative, and through the buttressing provided
by collaborative evaluation approaches, many program staff are al-
ready participating in the evaluation and improvement process prior
to the staff development phase. Now, each individual component of
the SEAF will be discussed in more detail.

1. Systemic Key Performance Indicators

According to the EIC, the public has a right to know how schools and school boards are performing on key indicators of success (Education Improvement Commission, 2000). The Systemic Key Performance Indicators (Systemic KPI) will help the PDSB constituency know how special education programs and services are performing. The EQAO Indicator Framework (Education Quality and Accountability Office, 2005) is used to identify sources for most of the indicator data—that is, the more coarse, general data are gathered through EQAO surveys or the Ministry of Education, and some of the finer and more specific

Figure 1
The Special Education Accountability Framework and Continuous Improvement
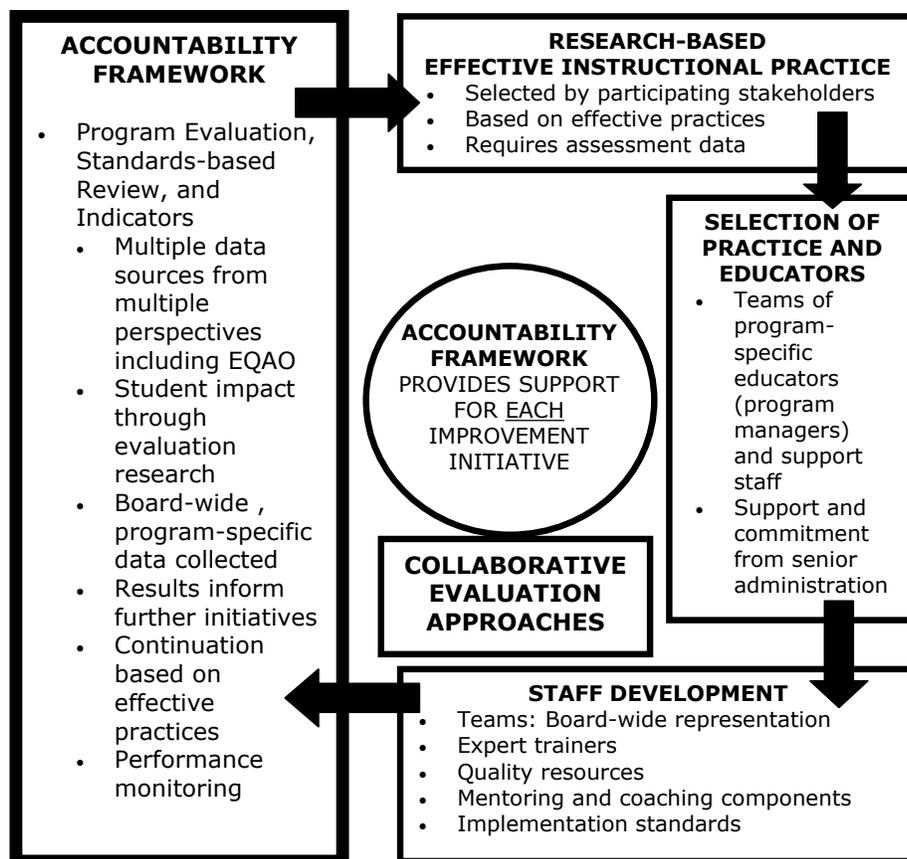
Note. Adapted from Little and Houston (2003).

data are available through board databases, including the SIS. Some of the program-specific data will be collected through instruments we develop or employ in the evaluation process. These program-specific indicators are more important than general indicators in the account-ability-driven improvement process because of their direct ties to the program (O'Day, 2002).

The SEAF's use of key performance indicators will be tied to targets or standards. Some of the standards will come from the Ministry of Education *Standards* (Ontario Ministry of Education, 2003) for spe-cial education program and service delivery. But, because some of the Systemic Key Performance Indicators will focus on student perform-ance, and the Ministry of Education *Standards* contain no standards for student performance, our organization will develop program-specific standards or targets for student performance, through the participatory and empowerment approaches outlined previously.

It is noted that in order to develop standards from a psychometric perspective, various content experts need to become involved in the standard-setting process. To help promote this involvement, we are building capacity within our central administration, program manag-ers, and professional support staff (psychologists, speech-language pathologists, and program coordinators) for two important purposes: (a) to understand and use performance indicator data to help make programming decisions, and (b) to understand the role of standards in education. This type of participatory approach in standard-setting has led to empowerment of stakeholders to later manage the initiatives on their own (Hunter & Gambell, 2000), a goal of our SEAF. One of the key messages we deliver to stakeholder groups is that accountability-type data collection, including performance indicators, must be part of routine data collection and not seen as an add-on to the busy schedules of teachers and other in-school personnel (Linn, 2003). The board's SIS helps greatly with indicator data collection and management. Still, accountability through performance indicators alone is not a rigorous enough approach to facilitate improvement in our schools.

2. Standards-based Review Process

We realize that the use of performance indicators and their related standards or targets can only tell us what performance standards have been met or not met. The Standards-based Review Process (SRP) delves a little deeper into special education programs and services to help answer the question of why performance standards were or

were not met. With this component, a particular program or service is identified for review by senior administration after examining existing performance indicator data. Then we use case review methodology and program logic models for determining the performance of key program/service functions based on the Ontario Ministry of Education's *Standards* for special education program and service delivery. This review process is tied to the Systemic KPI and thus is also participatory, aimed at involving and eventually empowering teams of professionals associated with the program, including educators, psychologists, occupational and physical therapists, and speech-language pathologists to engage in the review and standard-setting processes. With such an approach, the resultant standards take on more interpersonal and/or interorganizational meaning (Hunter & Gambell, 2000). The focus with this component of the SEAF is on program service and delivery to identify and address any issues that may be causing shortfalls between expected and actual performance.

The collaborative approach used in these first two components of the SEAF focuses on professional accountability and moves away from the bureaucratic control model for accountability (O'Day, 2002). The methodology for this review process is modelled on what is used in the United States, especially with regard to student outcome standards (e.g., Schulte, Villwock, Whichard, & Stallings, 2001; Thurlow & Krentz, 2001) but without the pressure of legislation and the threat of sanctions, allowing motivation for improvement by accountable units to be intrinsic. Teams of program-related professionals—sometimes led by evaluators, sometimes facilitated by evaluators—will examine a particular program or service as a case study. They will focus not only on compliance with standards, but will use interviews and focus groups with parents, service providers, and others working with students at the time of the review to identify which program aspects are working well and which are not. Because of the involvement of various professionals in a participatory and empowering approach, we are better able to form action-based improvement plans and next steps in the program planning process. Yet, using the first two components of the accountability framework, we still do not have a method of investigating in depth the relationship between contexts, inputs, process, and student outcomes.

3. The Special Education Program Evaluation Process

The program evaluation process allows for the most in-depth investigation of programs and services in our SEAF. This process is also par-

ticipatory, involving key stakeholder groups in evaluation planning, student assessment development, and data reflections. Participation of stakeholder groups here and with the other components of the SEAF also helps develop organizational capacity for accountability-driven improvement. What differentiates this component from the indicator system and review process, however, is the depth and angles of inquiry and the judgemental nature of a program evaluation. Further, by utilizing a program evaluation approach, we can define and develop program logic, helping to identify how the program is supposed to work, and thus what areas of the logic require improving in order to improve student outcomes. To do this, we employ mixed methodologies, which include the use of multilevel statistical models in quasi-longitudinal designs (see Linn, 2003), looking at all aspects of program logic including educational inputs, contexts, and processes, and ultimately focusing on their effects on student outcomes. Multiple perspectives of program functioning are also sought through this program evaluation process.

Of course, following program evaluation methodology means catering assessments of students, for example, to the focus and goals of the program. This is one area in which we make a large departure from traditional accountability systems that use one size fits all large-scale assessments, mainly for comparability purposes. The focus of the SEAF is the program, not the school. So comparability here comes into play within the program only, allowing us to utilize different assessments for different programs. For example, for the first program evaluation, we employed a published test of educational achievement, focusing on reading skills with related sub-tests, to measure reading achievement of students in a program designed for extensive reading remediation. We know, however, that this assessment tool will not fit other program evaluations, and more work is necessary in assessment development for each program evaluation.

Our program evaluation process can feed off the data collection of the other two accountability components, using those data to identify a program or service in need of more in-depth investigation. Conversely as well, our program evaluations, as in business and industry, can be used to introduce performance measures that will serve annually as indicator data (see McDavid & Hawthorn, 2006). We invite key stakeholder groups to participate in the evaluation steering committee and we value their input. Every teacher, or program manager, is an important player in the program evaluation process as well; their time is respected through the use of assessment tools and strategies

that further the understanding of their students' strengths and needs, yet yield reliable and valid data for accountability purposes.

PROSPECTS OF THE FRAMEWORK

Our vision for utilizing program evaluation for accountability is focused on collaborative approaches stressing a melding of bureaucratic and professional accountability with the evaluators serving as leaders, facilitators, and supporters. We envision stakeholder empowerment through their participation. As Morris (2002) describes, "Empowerment is the increased feeling or sense of power stemming from a given action, in this case participation" in the program improvement process for students with special education needs (p. 52). We hope that meaningful inclusion of stakeholder groups in the accountability-driven improvement process will to lead to shifts in ownership of the evaluation process from the professional evaluators to the stakeholder groups, such as program administrators and program managers, and increased interest in evaluation findings, as other researchers have found (Earl et al., 2005; Lee, 1999; MacLellan-Wright, Patten, dela Cruz, & Flaherty, 2007; Weaver & Cousins, 2004). Ultimately, this is the change in organizational behaviour we strive for.

LIMITATIONS OF THE FRAMEWORK

The Special Education Accountability Framework, while locally focused, is aimed at the program, not the school, and combines the bureaucratic and professional approaches to accountability. We will assess student progress with a mix of curriculum-based and standardized assessments, employing challenging process and outcome standards. In addition, we strive to address fairness issues, technically, with the use of multilevel statistical models and the value-added approach. Yet, much work to find and develop applicable student assessments and outcome standards needs to be done, as we are still in the relatively early stages of implementing the framework. We know this is a huge task.

Further, with any new practices at the organizational level, especially in an organization as large as the PDSB, implementation issues will arise as we endeavour to penetrate the special education classroom practices and normative structures of our schools. People are resistant to change, especially when the term "accountability" is associated with it. Compounding this problem is the fact that the SEAF

currently lacks expressed consequences—positive or negative—to help motivate accountable units toward improvement. This is where we depend on increased organizational capacity for accountability-based improvement through collaborative evaluation approaches, especially involving the senior administration. Yet at present, these administrators have not articulated system-wide consequences for accountable units, further than a promise of resource allocation for program improvement. We will have to wait and see whether this lack of expressed consequences helps or hinders the effects of the PDSB's Special Education Accountability Framework on system improvement.

THE FRAMEWORK IN ACTION

The 2007–2008 school year will be the fifth year of development and implementation of the PDSB's Special Education Accountability Framework. With one three-year program evaluation on the record and the completion of another three-year program evaluation expected for the coming school year, we have initially tested the tenets of the framework. What we have learned thus far will be discussed briefly.

The SEAF has already played a key role in the continuous improvement cycle within the organization (recall Figure 1). For example, indicator data collection and consultation with senior leadership led to identification of a program for evaluation: the Communication Program is intended for students with specific learning disabilities, and at the core of the program is intensive remediation aimed at reading skills. For the Communication Program evaluation, a standardized test of achievement was selected to measure student reading achievement, a key learning outcome, at two different times during the evaluation. The evaluation steering committee, including representatives from program administration, program managers, and professional support staff such as psychologists and speech-language pathologists, helped the evaluators choose the assessment. Subsequently, expert subgroups facilitated by evaluators and composed of psychologists and speech-language pathologists then chose specific subtests of the assessment for administration. These working groups also helped administer the assessment at the first measurement time. For the second measurement time, these participant stakeholders also helped train teachers in test administration, interpretation of results, and use of results in improvement planning. After a three-year evaluation including many program components, evaluation findings were

shared with senior leadership, program administrators, and program managers. Deficit areas in students' reading achievement were identified, and members of the evaluation steering committee are currently leading working groups of program administrators, program managers, and program support staff to identify a research-based effective instructional practice to address the deficit areas. Senior administration, with the aid of evaluators, have put logistics in place to train program managers in the to-be-identified practice, and to monitor the reading performance of students.

For the performance monitoring of students in the Communication Program, we have seen a participatory approach evolve into an empowerment of key stakeholder groups. The standardized assessment of reading used in the Communication Program evaluation is now (entering the second school year post-evaluation) being used board-wide by program managers on an annual basis as a performance measure of student reading achievement. As evaluators, we have stepped back and offer only support for this process by assisting in central data analysis and reporting of the annual performance measurement results (see Ali, 2006b). Program managers (teachers) use individual student results to monitor individual student performance, and adjust currently employed instructional strategies, whereas senior administration use aggregated results as part of the Systemic Key Performance Indicator data. Also, once a new research-based effective practice is implemented by teachers program-wide, the performance measurement system adopted by program managers will help monitor the success of the new instructional practice. Further participation of stakeholder groups will develop standards for growth in reading achievement on the standardized reading assessment.

In addition, our first few years of SEAF implementation have shown us, as Patton (1999) has suggested, that we as evaluators contributed to leadership development within the organization by "training senior people in evaluative thinking and developing their capacity to understand and use evaluation" (p. 102). Our first program evaluation within the SEAF, and the capacity building associated with that endeavour, has led to increased support for the evaluative process in our organization. For example, formation of an evaluation steering committee composed of representatives of key stakeholder groups and the adoption of an evaluation terms of reference for our second program evaluation, aimed at programs for students in enhanced placements, gained senior leadership approval much more quickly than the first program evaluation. For the second evaluation, there

was also a greater willingness of key stakeholder groups to be active participants in the evaluative process as a result of the initial implementation of the SEAF.

The application of collaborative evaluation approaches in the SEAF has also increased the utility of evaluation findings. Again drawing reference from Patton (1999), we were aware that "the use of evaluation findings and processes often depends on the support, understanding, involvement, and leadership of senior people in the organization" (p. 102). During the first few years of the SEAF, we have found that while there was an initial support for the evaluative process at the senior leadership level in the organization, the continued use of collaborative evaluation approaches has helped build a greater understanding and appreciation for the evaluative process, and has placed greater emphasis on the utility of evaluation findings for program improvement. Many key findings from the Communication Program were made, and recommendations have been acted upon (Ali, 2006a). For example, as already discussed, the evaluation of the Communication Program has led to targeted improvement efforts with regard to reading achievement. Annual performance measurement of reading achievement and the process of adopting a research-based effective instructional practice to boost reading skill development are direct uses of evaluation findings, and help fulfill accountability requirements of the board through a mix of professional and bureaucratic accountability models. Also, through the evaluative process, inconsistent identification processes for program admission were documented across the various geographic jurisdictions within the board. As a result of evaluation findings, subgroups of key stakeholders, such as program administrators and senior leadership have worked to make the admission process clearer and more consistently employed throughout the board. At this point in the development and implementation of the SEAF, it is apparent that evaluation findings and indicator data are valued and utilized by program managers, program administrators, and senior administrators.

Finally, it seems the lack of expressed consequences for accountable units beyond the promise of greater resources for program improvement has not hindered the initial operation of the SEAF. In fact, greater resources were promised and delivered in the case of the first program evaluation. More importantly, though, there is greater capacity within the organization for accountability-driven improvement, and greater willingness of stakeholders to be participants in the

evaluative process. Participatory evaluation has led to empowerment and, together with a melding of professional and bureaucratic models of accountability, some barriers to accountability-driven improvement have been overcome. More time and more development and implementation of the SEAF is needed to see if more barriers can be overcome. But, at least in the early years of the SEAF, the case can be made for the use of collaborative evaluation approaches to develop an accountability framework.

REFERENCES

Ali, A.S. (2004). *Developing and testing a school improvement-school effectiveness framework for evaluating complex school improvement initiatives*. Unpublished doctoral thesis, Ontario Institute for Studies in Education of the University of Toronto, Toronto, Ontario.

Ali, A.S. (2006a). *The Communication Program Evaluation: Final report*. Mississauga, ON: Peel District School Board.

Ali, A.S. (2006b). A special education program evaluation for accountability purposes: An in-depth case study. *Alberta Journal of Educational Research, 52*(2), 195–197.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baker, E.L., & Linn, R.L. (2002). *Validity issues for accountability systems: CSE Technical Report 585*. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing.

Baker, E.L., Linn, R.L., Herman, J.L., Koretz, D., & Elmore, R.F. (2001, April). *Holding accountability systems accountable: Research-based standards*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Cousins, J.B., & Earl, L.M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*, 397–418.

Cousins, J.B., & Earl, L.M. (Eds.). (1995). *Participatory evaluation: Studies in evaluation use and organizational learning*. London: Falmer.

Creemers, B.P., & Reezigt, G.J. (1997). School effectiveness and school improvement: Sustaining links. *School Effectiveness and School Improvement, 8*(4), 396–429.

CRESST. (2002, Winter). *Standards for educational accountability systems* (No. Policy Brief 5). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST) at UCLA. Retrieved May 12, 2003, from the CRESST website: <http://cresst96. cse.ucla.edu/products/policybriefs_set.htm>.

Earl, L.M. (1995). Assessment and accountability in education in Ontario. *Canadian Journal of Education, 20*(1), 45–55.

Earl, L.M. (1998). Developing indicators: The call for accountability. *Policy Options, 19*(6), 20–25.

Earl, L.M. (1999). Assessment and accountability in education: Improvement or surveillance? *Education Canada, 39,* 4–6, 47.

Earl, L.M., Ali, A.S., & Lee, L.E. (2005). Evaluating school improvement in Canada: A case example. *Canadian Journal of Program Evaluation, 20*(3), 151–174.

Earl, L.M., & Lafleur, C. (1998). Editor's introduction. *Orbit, 28*, 3.

Education Improvement Commission. (2000). The road ahead IV: A report on improving schools through greater accountability. Retrieved July 11, 2004, from <http://www.edu.gov.on.ca/eng/document/reports/eic/ road4/eic4.pdf>.

Education Quality and Accountability Office. (2005). *EQAO guide to school and board improvement planning: A handbook for school and board leaders*. Retrieved July 5, 2007, from <http://www.eqao.com/pdf_e/05/ 05P011e.pdf>.

Education Quality and Accountability Office. (2007a). *EQAO policy on accommodation*. Retrieved July 26, 2007, from <www.eqao.com/pdf_ e/07/07P008e.pdf>.

Education Quality and Accountability Office. (2007b). *Ontario secondary school literacy test: Board results*. Retrieved July 21, 2007, from <http://www.eqao.com/pdf_e/06/06P031e.pdf>.

Elmore, R.F. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66*(1), 1–26.

Elmore, R.F., Abelmann, C.H., & Fuhrman, S.H. (1996). The new accountability in state education reform: From process to performance. In H.F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65–98). Washington, DC: Brookings Institution.

Favaro, P., Fine, J., Norman, E., & Bell, A. (1998). Accountability and school-based planning: A focus on collaboration, data collection and communication. *Orbit, 28*, 39–41.

Fetterman, D.M. (1999). Reflections on empowerment evaluation. *Canadian Journal of Program Evaluation, 14*(Special), 5–37.

Fullan, M.G. (1990). Linking classroom and school improvement. *Educational Leadership, 47*(8), 13–19.

Fullan, M.G. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

Goldstein, H., Huiqi, P., Rath, T., & Hill, N. (2000). *The use of value added information in judging school performance*. Retrieved November 18, 2002, from <http://www.ioc.ac.uk/hgpersonal/value-added-school-performance.html>.

Gray, J., Jesson, D., & Jones, B. (1986). The search for a fairer way of comparing schools' examination results. *Research Reports in Education, 1*(2), 91–122.

Hunter, D.M., & Gambell, T.J. (2000). Incorporating stakeholders in standard setting: What's at stake? *Canadian Journal of Program Evaluation, 15*(Special), 83–109.

Jordan, A. (2001). Special education in Ontario, Canada: A case study of market-based reforms. *Cambridge Journal of Education, 31*(3), 349–371.

Krentz, J.L., Thurlow, M.L., & Callender, S. (2000). *Accountability systems and counting students with disabilities (Technical Report 29)*. Minneapolis, MN: National Centre for Educational Outcomes, University of Minnesota.

Lee, L.E. (1999). Building capacity for school improvement through evaluation: Experiences of the Manitoba School Improvement Program Inc. *Canadian Journal of Program Evaluation, 14*(Special), 155–178.

Linn, R.L. (1998). *Assessments and accountability: CSE Technical Report 490*. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing.

Linn, R.L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–16.

Linn, R.L. (2003). *Accountability: Responsibility and reasonable expectations (CSE Technical Report 601)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R.L. (2004). Accountability models. In S.H. Furhman & R.F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 73–95). New York: Teachers College Press.

Little, M.E., & Houston, D. (2003). Research into practice through professional development. *Remedial and Special Education, 24*(2), 75–87.

MacLellan-Wright, M.F., Patten, S., dela Cruz, A.M., & Flaherty, A. (2007). A participatory approach to the development of an evaluation framework: Process, pitfalls, and payoffs. *Canadian Journal of Program Evaluation, 22*(1), 99–124.

Mark, M.M., Henry, G.T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.

McDavid, J.C., & Hawthorn, L.R. (2006). *Program evaluation and performance measurement: An introduction to practice*. Thousand Oaks, CA: Sage.

Mehrens, W.A. (1998, July 14). Consequences of assessment: What is evidence? Retrieved August 19, 2002, from <http://olam.ed.asu.edu/epaa/v6n13.html>.

Mertens, D.M., & McLaughlin, J.A. (2004). *Research and evaluation methods in special education*. Thousand Oaks, CA: Corwin.

Morris, D.B. (2002). The inclusion of stakeholders in evaluation: Benefits and drawbacks. *Canadian Journal of Program Evaluation, 17*(2), 49–58.

Nagy, P., Demeris, H., & van Barneveld, C. (2000). Priorities and values in accountability programs. *Canadian Journal of Program Evaluation, 15*(Special), 67–82.

National Center for Educational Outcomes. (2005). Accountability [Electronic Version] from <http://education.umn.edu/nceo/TopicAreas/Accountability/Account_topic.htm>.

O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review, 72*(3), 293–329.

Ontario Ministry of Education. (2003). *A standards-based approach to special education programs and services (Draft)*. Toronto: Author.

Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M.Q. (1999). Organizational development and evaluation. *Canadian Journal of Program Evaluation, 14*(Special), 93–113.

Peel District School Board. (2006a). *Grade 3 and Grade 6 EQAO assessments of reading, writing, and mathematics*. Mississauga, ON: Author.

Peel District School Board. (2006b). *Grade 9 EQAO assessment of mathematics: 2005–2006 overview of results*. Mississauga, ON: Author.

Peel District School Board. (2006c). *Ontario secondary school literacy test: 2005–2006 results*. Mississauga, ON: Author.

Peel District School Board. (2007). *Just the facts*. Retrieved July 10, 2007, from <http://www.peelschools.org/facts/facts/special.htm>.

Robinson, V., & Timperley, H. (2000). The link between accountability and improvement: The case of reporting to parents. *Peabody Journal of Education, 75*(4), 66–89.

Ryan, K. (2002). Shaping educational accountability systems. *American Journal of Evaluation, 23*(4), 453–468.

Sanders, W.L., & Horn, S.P. (1994). The Tennessee Value-Added Assessment System (TVASS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299–311.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61–80.

Schulte, A.C., Villwock, D.N., Whichard, S.M., & Stallings, C.F. (2001). High stakes testing and expected progress standards for students

with learning disabilities: A five-year study of one district. *School Psychology Review, 30*(4), 487–506.

Sireci, S.G., Scarpati, S.E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457–490.

Stoll, L., & Fink, D. (1992). Effecting school change: The Halton approach. *School Effectiveness and School Improvement, 3*(1), 19–41.

Stoll, L., & Fink, D. (1994). School effectiveness and school improvement: Voices from the field. *School Effectiveness and School Improvement, 5*(2), 149–177.

Stufflebeam, D.L. (1983). The CIPP Model for program evaluation. In G.F. Maudaus, M.S. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human service evaluation* (pp. 117–142). Boston: Kluwer-Nijhoff.

Tetlock, P.E. (1998). Losing our religion: On the precariousness of precise normative standards in complex accountability systems. In R.M. Kramer & M.A. Neale (Eds.), *Power and influence in organizations* (pp. 121–144). Thousand Oaks, CA: Sage.

Thurlow, M.L. (2004). Biting the bullet: Including special-needs students in accountability systems. In S.H. Fuhrman & R.F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 115–137). New York: Teachers College Press.

Thurlow, M.L., & Krentz, J.L. (2001). A new vision for special education accountability. *State Education Standard, 2*(3), 4–7.

U.S. Congress. (2001). *The No Child Left Behind Act (PL 107-110)*. Washington, DC: U.S. Government Printing Office.

U.S. Congress. (2004). *Individual With Disabilities Education Improvement Act (PL 108-446)*. Washington, DC: U.S. Government Printing Office.

Weaver, L., & Cousins, J.B. (2004). Unpacking the participatory process. *Journal of Multidisciplinary Evaluation, 1*, 19–40.

Weiss, C.H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Wholey, J.S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J.S. Wholey, H. Hatry, & K. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15–39). San Francisco, CA: Jossey-Bass.

Willms, J.D. (1987). Differences between Scottish educational authorities in their educational achievement. *Oxford Review of Education, 13*(2), 211–232.

Willms, J.D. (1992). *Monitoring school performance: A guide for educators*. London: Falmer.

Zimmerman, M.A. (2000). Empowerment theory: Psychological, organizational and community levels of analysis. In J. Rappaport & E. Seldman (Eds.), *Handbook of community psychology* (pp. 43–64). New York: Plenum.

A. Sidiq Ali, Ph.D., is an independent evaluation consultant and was the Research Officer for Special Education in the Peel District School Board. Sid is certified to teach in Ontario with a *Specialists*' designation in Special Education. He has also taught graduate courses in educational assessment and evaluation, research methods, and statistics. His research interests include educational and community-based evaluation, policy development, and performance-based measurement.

Paul Favaro Ph.D., is Chief of Research and Evaluation at the Peel District School Board and Instructor in the School of Administrative Studies, York University. His research interests includes educational evaluation and assessment, teaching models of evaluation, empowerment and community-based evaluation, and development and implementation of performance measurement accountability systems.