The Alberta Journal of Educational Research *Vol. 56, No. 4, Winter 2010, 403-418*

John A. Ross

and

Timothy M. Sibbald
Ontario Institute for Studies in Education, University of Toronto

The Role of External Diagnosis in School Improvement in an Ontario School District

External diagnosis is recommended when schools lack the capacity to assess their needs. This qualitative study of one Ontario district compared 33 elementary schools that conducted external diagnosis with 47 schools that used internal diagnosis. External diagnosis created pressure for change, helped schools develop a plan that included previously neglected needs, promoted consistency within and between schools, contributed to the improvement culture of the school, and encouraged shared instructional leadership. It also depressed teacher efficacy and commitment to school improvement. Positive effects of external diagnosis were facilitated and negative effects mitigated by principals who adopted shared instructional leadership strategies.

L'on recommande une analyse externe quand les écoles n'ont pas la capacité d'évaluer leurs besoins. Cette étude qualitative d'un district en Ontario porte sur une comparaison entre 33 écoles élémentaires qui ont entrepris un diagnostic externe et 47 écoles qui ont eu recours à un diagnostic interne. Les diagnostics externes ont créé une pression pour le changement, aidé les écoles à développer un projet qui incluait des besoins auparavant négligés, favorisé la cohérence à l'intérieur des écoles et entre elles, contribué à la culture d'amélioration de l'école et encouragé le partage du leadership de l'enseignement. Ils ont également diminué l'efficacité des enseignants et leur engagement face à l'amélioration de l'école. Les effets positifs des diagnostics externes ont été facilités et les effets négatifs mitigés par les directeurs qui ont adopté des stratégies de partage du leadership de l'enseignement.

Introduction

In the first waves of reform, provinces and reform agencies did not intervene directly into instructional improvement and bypassed districts to stimulate change in individual schools (Elmore & Burney, 1998). Inattention to the role of the district dissipated these efforts. Districts inadvertently frustrated individual school progress when they failed to set priorities among alternate agendas for reforming instruction, exerted pressure for immediate improvement that conflicted with the time required for teachers to learn new methods, and failed to provide strategies for scaling up successful instructional innovations (Corcoran, Fuhrman, & Belcher, 2001; Hajnal, Walker & Sackney, 1998).

John Ross is a professor emeritus of curriculum, teaching, and learning at OISE/UT and Head of the Institute's field centre in Peterborough. His research interests are school change, mathematics education, student assessment, and program evaluation.
Timothy Sibbald is a secondary school mathematics teacher in Thames Valley District School Board (London, ON) and a research associate at the OISE/UT Trent Valley Centre. His research interests are mathematics education and school improvement.

Researchers have identified functional district strategies that contribute to school improvement such as district monitoring of school progress (Harris, Chapman, Muijs, Russ, & Stoll, 2004), providing probation managers for failing schools and external partners to guide curriculum renewal (Finnigan & Stewart, 2008), coordination of school and district goals, including bilateral negotiation of priorities between individual principals and district leaders (Elmore & Burney; Hajnal et al.; Zavdasky, 2007), clearing away dysfunctional policies (Fullan, 1999), and designing district professional development that integrates individual and collective capacity development (Watson, Fullan, & Kilcher, 2000).

Studies of effective schools and comparisons of successful and failing schools led to clusters of prescriptions describing how districts can support school improvement[1] (Elmore, 2004). But there has been little research on the effects of specific district strategies, with the exception of studies of the district's role in supporting schools' use of data to diagnose student needs, design instruction, and monitor progress of learners. District facilitators of effective data use include the provision of teacher and school leader inservice on how to interpret data in the context of state performance standards, technical support to schools, the establishment of structures for storing and accessing data, and ongoing and explicit endorsement of data-based instructional actions by district leaders (Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Wayman & Stringfield, 2006). Firestone and Gonzales (2007) found that district guidance to schools on how to handle data was mediated by district culture: For example, in accountability cultures, districts treated test scores as ends in themselves and required schools raise scores, whereas in learning cultures, districts expected schools to use test data to improve instruction.

We investigated a district strategy that focused on the needs assessment phase of the school improvement process. The district provided an external diagnostician to schools that were perceived by district leaders as lacking the capacity to identify their own needs. The remaining schools were permitted to conduct their needs assessments internally. In this article, we contrast the experiences of staff in both types of school, identifying the benefits and challenges of this district strategy.

We define an external diagnostician as an expert external to the school who uses his or her knowledge of school effectiveness factors, school improvement processes, and local conditions to identify a school's needs. The argument for external diagnosis rests on a deficit model that assumes that underachieving schools lack the internal capacity to launch improvement efforts. A diagnostician can provide this missing capacity, thereby building the foundation for a well-targeted improvement plan. There is little research on the use of the strategy despite its popularity in countries as diverse as Canada (Ross, Sibbald, Gray, & Scott, 2006) and the Netherlands (Jong, Houtveen, & Westerhof, 2002). In this article we identify the perspectives of multiple stakeholders on external diagnosis in order to identify the benefits and deficits of the model.

Theoretical Framework

School Improvement Mechanisms

We framed our inquiry in Fullan's (2002, 2005) theory of action, which highlights a small set of principles and strategies, particularly *capacity building*,

partnerships, and *accountability* (Fullan & Campbell, 2007). In Fullan's theory, capacity building consists of three dimensions: (a) the acquisition of research-based technical skills to improve student learning, for example, Expert Panel on Early Reading (2003); (b) the creation and maintenance of a supportive organizational structure (such as the professional learning communities of Louis & Marks, 1998); and (c) moral leadership that raises the values of members, motivating them to go beyond self-interest to embrace organizational goals, and redefining their needs to align with organizational preferences. Partnerships refer to the development of productive relationships between schools and external agencies such as national project staff (Chapman & Allen, 2006) or school district agents (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). Accountability refers to the process of setting school targets, measuring performance, and identifying data-based, ameliorative strategies.

Potential Contribution of External Diagnosis to School Improvement

Earlier studies identified four ways in which the needs assessments undertaken by low achieving schools are deficient. First, internal assessments are characterized by shallow compliance. The goal of the staff is to appease external authorities, not identify authentic needs (Burch, 2007; Mintrop & MacLelland, 2002). Internal diagnosis identifies safe areas, ignoring contentious issues such as the quality of school leadership and the appropriateness of instructional practices (Vincent, Patterson, Buehler, & Gearity, 2006), that is, these schools focus on the wrong issues. Second, internal diagnoses suffer from lack of knowledge of school effectiveness factors and school improvement strategies: deficits based on the belief that academic research has little practical value (Burkhardt & Schoenfeld, 2003; Coburn & Talbert, 2006; Kennedy, 1997). Underachieving schools focus on needs that may not be strongly associated with outcomes. Third, evidence about needs is overly reliant on locally developed measures of student achievement and shallow analysis of standardized tools (Vincent et al., 2006), creating school improvement goals founded on invalid assumptions. Fourth, underachieving schools may lack the resources to conduct rigorous needs assessment because non-academic demands consume so much staff time (Harris et al., 2006), and such schools may have no history of prior success, a key element in developing an improvement culture (Reezigt & Creemers, 2005).

An external diagnostician might have the expertise to generate a more accurate estimate of school needs and the independence to address sensitive areas without fear of retribution. By identifying substantial gaps between desired and actual performance, the diagnostician could lower the comfort level of school staff and create a state of positive urgency that motivates change (Donaldson, 1999; Reezigt & Creemers, 2005).

The Risk of External Diagnosis

Social cognition theory (Bandura, 1997) suggests that in attempting to supplement a school's capacity, external diagnosis could depress it further. A strong sense of group capability (i.e., collective teacher efficacy[2]) establishes expectations for success that encourage teachers to adopt demanding but effective instructional strategies and persist through obstacles during implementation. Collective efficacy contributes to student achievement (Goddard, Hoy, &

Woolfolk Hoy, 2004; Ross & Gray, 2006a), the end goal of school improvement. The risk of external diagnosis is that it might depress collective teacher efficacy. Messages communicating to individuals that they are not competent to complete core tasks reduce self-efficacy beliefs (Usher & Pajares, 1995).

The negative effects of external diagnosis on collective teacher efficacy may be moderated by leadership, particularly if principals adopt a transformational style in which their actions are dedicated to fostering the growth of organizational members (Leithwood, Jantzi, & Steinbach, 1999). Ross and Gray (2006b) found that principals who engaged in transformational leadership practices contributed to higher collective teacher efficacy and commitment to organizational values.

Research Questions

We conducted a qualitative study of a district in which 32 elementary schools began their school improvement process with an external diagnosis of their needs and 48 schools began with an internal diagnosis. The research questions were: How did teachers, principals, and other school staff experience external diagnosis? To what extent did stakeholders perceive the needs identified to be authentic? What were the positive and negative effects? How did school leaders moderate the effect of external diagnosis?

Methodology

Context of the Study

The district served 35,000 students in a rural and urban area covering 7,000 square kilometers. Over 95% of the students in the district were Canadian-born, only 2% spoke a language other than English at home, 15% were identified as special needs, and average family income in the district was near the provincial median ($54,958).

Before 2006-2007, teachers and principal identified school needs without external participation, following a provincial model that required that schools identify their needs, develop improvement plans, monitor implementation, and measure outcomes. In 2006-2007 the district developed an external diagnosis component for selected schools. The district designated the bottom quartile of its 80 elementary schools as low-achieving, that is, fewer than half their students had reached the provincial achievement standard over the preceding three years. External diagnosticians were assigned to these 20 schools and to eight intermediate schools (that had grades 7-8 only) that were not included in provincial assessments. Five schools hired an external diagnostician using their own funds. The remaining 47 schools conducted internal needs assessments.

The diagnostician model was based on an Ontario program for underachieving schools. The district hired as chief diagnostician a retired official who had been a leader in the design and implementation of the Ministry's program. She had extensive senior district and school management experience. The other three diagnosticians were elementary principals in the study district (two were retired) who had demonstrated success in turning around underachieving schools. All four were considered by district managers to be experts in school improvement.

The team leader trained the other three diagnosticians. They conferred among themselves, but conducted diagnostic visits individually. The diagnos-

tician visited the school over a two-day period, spending brief periods in each classroom, reviewing student tasks, teacher daybooks, and literacy resources. Diagnosticians did not examine principals' plans or seek teacher-initiated input on school goals and practices. On the second day, the diagnostician presented a one-hour oral report to the staff. The principal received a written report within two weeks.

The oral and written reports were based on a template generated in the Ministry program. The district adopted the Ministry template because it was believed that underachieving schools across the province had similar needs. The key criteria for assessing instructional quality in the template were: (a) teaching to produce usable knowledge versus teaching knowledge for its own sake; (b) student achievement (based on four levels defined by Ministry curricula); (c) degree of curriculum integration (five levels); and (d) Bloom's taxonomy of instructional objectives (six levels). The 15-page report to schools contained a short list of strengths and a lengthy list of issues that required attention. The largest focus area addressed the need for greater fidelity to core district/Ministry policies regarding curriculum planning, instructional strategies, and assessment practices. A typical recommendation was

> Examine the use of instructional time to ensure that the majority of teaching time is not devoted to teacher talk and control. At this time the majority of instructional time is geared to student listening, single student responses, quiet and isolated follow-up activities.

Less detail was provided on the other three focus areas: collaboration among teachers, adequacy of reading resources, and the use of assessment data to set targets.

Sources of Data

In February and March 2007, we interviewed random samples of school staff (teachers, principals, literacy coaches, and special education resource teachers) in homogeneous groups, that is, six to eight staff members from the same role, same grade level, and from the same school type were interviewed as a group for 75-90 minutes. We also interviewed diagnosticians, district consultants, and the district's senior administrators. We designed the 23 focus groups to be homogeneous because participants feel less anxiety if group members have similar characteristics (Stewart, Shamadasani, & Rook, 2007). Interview guides invited interviewees to identify successful and unsuccessful episodes in the project, probed specific roles (e.g., literacy coaches and principals) and events in the change model (diagnosis, improvement plan development, etc.). Interviews were taped and transcribed. Observations were conducted in four administrative sessions (two principals' meetings, one senior planning team meeting, and an inservice on interpreting the diagnostician's report).

We also examined student achievement data, that is, scores in grades 3 and 6 reading and writing on mandated external assessments administered before the introduction of external diagnosis and after one year of the program. Prior achievement was the three-year average score on provincial assessments because comparisons from one year to the next are unstable—student cohorts and their teachers vary on characteristics that affect achievement (Kelly & Monczunski, 2007; Linn & Haug, 2002)—and when schools are grouped in

terms of prior achievement, regression to the mean effects may give the misleading appearance of improvements in low achieving schools (Chay, McEwan, & Urquiola, 2005).

Data Analysis

We began by blocking focus-group data responding to questions about needs identification, regardless of whether the diagnosis was internal or external. We coded these data using descriptive start codes consisting of a three-level category system of responses to needs assessment by role (teacher, literacy coach, special education resource teacher, principal, diagnostician, curriculum consultant, senior administrator) and by diagnosis type (with and without external diagnosis). Through constant comparison within and between these start codes, we constructed emergent codes that reflected patterns in the data. These emergent, in vivo codes were based on relationships between the central phenomenon, external diagnosis, and factors influencing the phenomenon, especially its consequences (e.g., effects on teacher morale). We selected from these emergent codes six themes that best represented the district's experience with external diagnosis. We used the six themes to describe how external diagnosis contributed to school improvement and interpreted the themes in the context of research on school effectiveness and school improvement.

The seventh theme was a comparison of student achievement scores of three groups of schools: those assigned an external diagnostician, those that hired a diagnostician out of school funds, and those that conducted an internal diagnosis. We aggregated the 2007 scores across grades and subjects and compared the results of each school group with the 2004-2006 scores.

Results

We summarized the results using the seven themes: (a) the accuracy and fairness of external diagnosis was disputed; (b) external diagnosis created powerful pressure for change; (c) external diagnosis contributed to consistency within and among schools; (d) external diagnosis contributed to teacher resistance and withdrawal; (e) external diagnosis depressed teacher confidence in their professionalism; (f) school and district agents mitigated the negative effects of external diagnosis; (g) external diagnosis was associated with improved student achievement.

The Accuracy and Fairness of External Diagnosis was Disputed

District administrators believed that the diagnostic reports accurately reflected conditions in schools. In the words of one superintendent, "It comes close to home regarding what teachers are doing." Senior managers said that the diagnosticians' reports had high credibility because they were generated by unbiased observers of school conditions who were knowledgeable about the district's expectations for schools.

This view was shared by curriculum consultants who regularly visited schools. An experienced consultant stated that the diagnostician reports were "very fair and very honest … They nailed exactly … what they were and were not doing." The consultants argued that the diagnosticians identified issues that the schools were not able to see for themselves. The theme of schools not knowing their needs was also central to the external diagnosticians' reflection on their experience. One diagnostician reported as an example: "Many of the

schools I went into told me that they didn't know what balanced literacy was, so how were they supposed to implement that?" Principals agreed that the diagnosis was accurate, albeit narrow in its data-collection.

In contrast, teachers claimed that the reports were wide of the mark. For example, a teacher in a low-achieving school argued, "Some of the data the diagnostician used to support her observations were patently false. And when challenged on that, her response was—I had to make some assumptions." Teachers believed the evaluations were unfair and the diagnosticians were not as expert as they claimed to be, as this teacher noted:

> She [the diagnostician] came early in October and [criticized us because] we didn't have guided reading set up yet. If she had [the recommended guide for teaching reading] and … had actually read it, she should know that in grades 1 and 2 [guided reading] shouldn't be started until six weeks or later when the kids actually know the routines and how to follow them independently. Then we start guided reading.

All reports followed the district template so closely that teachers questioned the authenticity of the data. Each report was said by teachers (and literacy coaches) to be a carbon copy of the others.

Teachers expressed concern that the diagnosis ignored school history. Teachers, and to a lesser extent coaches and principals, believed that low-achieving schools had made substantial progress in recent years. Student achievement was higher, and instructional practices had improved. None of this progress was acknowledged. Nor did the diagnosticians recognize the challenging context facing many schools. As one teacher in a low-achieving school stated,

> Our school is an inner-city school with very high needs … 70% of our kids live below the poverty line … Most of the people in our building are there by choice. They've worked with high risk kids for many years. They have incredible expertise—we've got specialists in reading, several people with specialists in special education, several people with master's degrees in curriculum … and we had a diagnostician who … gave us a report repeatedly using language like, the staff has a surface knowledge of everyday things.

Teachers reported making multiple changes that addressed specific needs identified in the reports. Teachers said they became more reflective about teaching practices, increased their use of higher-order questioning in lessons, raised their focus on daily lesson plans, and added cross-curricular attention to literacy. Teachers attributed these improvements to other dimensions of the district's program, not the external diagnosis. They were reluctant to acknowledge that external diagnosis had any benefits.

Although the accuracy of the reports was disputed, they focused on controversial issues about teaching that would probably not have been identified as needs by teachers (Burch, 2007; Mintrop & MacLelland, 2002; Vincent et al., 2006). In doing so, they supplemented the capacity of the school by identifying technical needs that became the focus for subsequent school improvement phases. Less attention was given to organizational structures, although specific suggestions for strengthening teacher collaboration were made. The diagnostic reports said little about leadership, even though it is a critical feature of school

improvement (Muijs, Harris, Chapman, Stoll, & Russ, 2004) and contributes indirectly to student achievement (Ross & Gray, 2006a). Nor was there a discussion of prior improvement history, another critical within-school factor predicting improvement (Reezigt & Creemers, 2005). In addition, the reports said little about the community context in which school change was to occur, even though such variables as student population demographics have a profound effect on school improvement culture (Wikeley, Stoll, Murillo, & De Jong, 2005).

External Diagnosis Created Powerful Pressure for Change

Senior managers felt that conditions in low-achieving schools required tough messages, as in this district administrator's reflection.

> [Provincial test] results for years had shown that they weren't going to progress in literacy and numeracy if they continue to do what they had always done. They needed to have a huge wakeup call and the diagnostician was a perfect way to do it. I could have talked with them about this until I was blue in the face, and it wouldn't have made any difference … they needed a little whack on the side of the head.

Principals felt that once the tough message had been delivered by the diagnostician, the principal could be a rallying point, creating cohesion by working as a team with teachers against hostile outsiders. Literacy coaches expressed similar views. Coaches in schools that did not have access to a diagnostician felt that they had to be more confrontational. As one coach said, "We have been put into a situation, without having a diagnostician's report, of being the heavy … In the schools with the diagnosticians you didn't have to break down that wall; it was broken down for them."

The diagnostic reports generated dissonance around taken-for-granted teaching practices, a key condition for change that heightens school accountability (Donaldson, 1999; Gess-Newsome, Southerland, Johnston, & Woodbury, 2003; Reezigt & Creemers, 2005). External diagnosis also contributed to school capacity by strengthening relationships among teachers and between teachers and principal.

External Diagnosis Contributed to Consistency Within and Among Schools

The needs identified in the diagnostic reports cut across classrooms. Professional conversations generated similar strategies and greater consistency among classrooms, as illustrated by a teacher in a low-achieving school who stated, "Sharing ideas, looking at the data wall and that has made us much more cohesive, instead of everybody doing their own thing in their own place."

Consistency across the district followed from each diagnostician being responsible for multiple schools; they circulated drafts of their reports within the diagnostic team; they used the same template. The diagnosticians brought a shared language for talking about literacy instruction that extended beyond the schools that received a report. Coaches and principals in schools without access to a diagnostician hastened to get copies of the template so that they could make better sense of the language used in administrative meetings and inservice sessions.

The diagnosticians framed the dissonance they evoked consistently across the district. Uniformity in needs assessment processes contributed to greater

consistency among schools and to tighter alignment among teachers, school leaders, and district leaders, a key predictor of school improvement (Zavdasky, 2007). External diagnosis strengthened organizational cohesion within and between schools.

External Diagnosis Contributed to Teacher Resistance and Withdrawal

The first response of teachers to the diagnostic report was stunned silence. Anger and resistance quickly followed, as expressed by a teacher in a low-achieving school.

> The immediate reaction was, let's circle the wagons and take care of each other and if anybody points a finger at any one of us, we're all going to jump up and down. We have a staff that is quite strong, [union] executive folks … The effect of these recommendations was, come and get us … we're ready.

Other teachers reported withdrawing from the improvement process as they "licked their wounds." The immediate result of external diagnosis was to reduce teacher commitment to school improvement.

Not involving teachers in key school improvement decisions has negative effects on improvement plan implementation (Datnow & Castellano, 2000; Smith et al., 1998). Teacher support is more likely when teachers trust and feel respected by external interveners (Stein, Hubbard, & Mehan, 2002), which was not the case with teachers who experienced external diagnosis. Although external diagnosis contributed to increased cohesion within schools, its contribution to the organizational capacity was depressed by its negative effect on teacher commitment.

External Diagnosis Depressed Teacher Confidence in Their Professionalism

Staff in schools that were assigned an external diagnostician felt stigmatized. Teachers perceived the diagnostic reports to be evaluative, even though the district emphasized they should not be viewed as personnel appraisal documents. But as one intermediate school teacher asked, "How can you go into someone's classroom, look in their day book and observe their teaching and not be evaluating them?" Teachers perceived the reports as uniformly negative. All teachers were branded with the same iron. Teachers who had been positively appraised through the district's clinical supervision policy were outraged when they were characterized as under-performers. Teachers asked, "Why was there a need to cut us down and undermine our confidence?"

The initial effect of the diagnostic reports on teacher confidence was negative. Over time the reports had an indirect, positive effect on teacher confidence in that it stimulated teacher interest in professional development offerings that upgraded their skills. Teachers who participated in these sessions reported feeling more capable as, for example, the teacher who attended the session on the district's formative assessment system: "When parents' night came I had never felt so comfortable speaking about my students' reading, speaking about their directions, about what my role was. I felt the most confident I have ever felt in interviews."

Teachers' who have confidence in their collective professional capacity are more likely than those on less confident staffs to adopt innovative instruction and improve student achievement (Goddard et al., 2004). External pressure to change instructional practice can depress efficacy beliefs (Smith, 1996). But in

this case, as in Ross, McKeiver, and Hogaboam-Gray (1997), the initially negative effects of external pressure on teacher confidence dissipated and teacher confidence returned. The direct effects of external diagnosis on school capacity were negative; the indirect effects, moderated by principals (as shown below), were positive.

School Leaders Mitigated the Negative Effects of External Diagnosis

Principals saw their tasks as reducing the emotional sting, clarifying the report, enlisting staff commitment, and sequencing efforts. Principals guided teachers through a close analysis of their report to identify specific issues on which change was required. This interpretative process eased the pain by focusing on behaviors, not people. It strengthened cohesion in the school and created work groups that had not previously existed, as indicated by the comment of this teacher in the intermediate schools focus group.

> We sat down as a group of grade 7 and 8 teachers with our principal and went through it. We tried to highlight any of the positives [group laughter]. Then we dug into all the things we did wrong. It was tough on the staff … Once that was over with, and our principal was great at doing that, then we found a place we could agree on for improvement. The greatest thing that's come out of this is collaboration.

Principals reduced the negative effects of the report by enabling teachers to acquire the skills required to meet the diagnostician's concerns. For example, schools were directed to install a "data wall," that is, a tracking space in which to record students' progress using standardized formative assessment procedures. They built teacher trust in the literacy coach who was the principal's agent for introducing and supporting new instructional practices. Other principals who had been literacy coaches earlier in their careers provided training to teachers directly. They provided formal authority to support instructional improvement. They framed their role as coordinator of internal and external resources for improvement, especially by creating networks of learning communities in their schools that enabled teachers to learn by observing peers teach and engaging in reflective dialogues about teaching.

Although principals' actions could be labeled transformational, their focus on the improvement of teaching and learning suggests that their leadership style might be better represented by a term coined by Marks and Printy (2003), "shared instructional leadership," meaning that the principal is committed to enabling teacher growth by distributing leadership functions across the school. Marks and Printy found that transformational leadership was a necessary precondition for the emergence of shared instructional leadership. They also found that the latter construct was a strong predictor of pedagogical quality and student achievement.

Transformational leadership flattens hierarchy (Leithwood et al., 1999) and is associated in earlier research with distributed leadership (Harris, 2004; Leithwood et al., 2007). Leadership was shared with literacy coaches, members of the school improvement team, and other teacher leaders. Harris found that distributed leadership required collegial norms associated with school improvement. Our data suggest that distributed leadership built collegiality as well as benefiting from it. The tension experienced by literacy coaches between

observing teachers for improvement rather than evaluation was resolved by reducing the gap between administrators and teachers. Distributed leadership was a form of collective agency that enabled the school to coordinate and focus its resources on issues cited in the diagnostic reports. The district supported distributed leaderships by making it a core value (Datnow, Lasky, Stringfield, & Teddlie, 2006) in the design of the model and by promoting it during principal training sessions. External diagnosis contributed to school capacity indirectly, that is, by encouraging shared instructional leadership.

External Diagnosis was Associated with Improved Student Achievement.
Student achievement gains from the three-year average before the implementation of external diagnosis to the year of implementation were higher in schools employing an external diagnostician than in schools that conducted an internal diagnosis. The effect sizes, Cohen's *d* with Hedges and Olkin (1985) correction for small sample size, ranged from 0.18 in schools without a diagnostician to 0.36 in schools that hired a diagnostician using their own funds, to 0.58 in schools assigned a diagnostician by the district. These differences cannot be attributed entirely to external diagnosis because each group received varying levels of district support. For example, the low-achieving schools that were assigned an external diagnostician were also assigned a full-time literacy coach, whereas the schools that conducted an internal diagnosis had access only to a part-time coach.

Epilogue: The End of the Diagnostician Role
In summer 2007, the district decided to continue to fund the diagnostician role for low-achieving and intermediate schools and to encourage other schools to hire their own diagnostician. But the Ministry revised its school improvement policy, making needs assessment an entirely internal responsibility, thereby ending external diagnosis in the district. The province did not provide a rationale for the decision, but it is likely that the concerns expressed by teachers in our study were shared by teachers in other districts.

Discussion

Contribution of the Study
External diagnosis was introduced to compensate for the perceived inability of low-achieving schools to diagnose their needs for school improvement. External diagnosis addressed many of the concerns about school self-assessment identified in earlier studies (Burkhardt & Schoenfeld, 2003; Coburn & Talbert, 2006; Kennedy, 1997; Vincent et al, 2006) and addressed the components of Fullan's (2002, 2005) theory of action. External diagnosis provided an unflinching identification of school needs that linked school conditions to standards derived from the school effectiveness and school improvement literature. The diagnosis created irresistible pressure for change (Donaldson, 1999; Gess-Newsome et al., 2003; Reezigt & Creemers, 2005). Social cognition theory (Bandura, 1997; Usher & Pajares, 1995) predicted that replacing school self-assessment with an externally controlled process would have negative effects on teachers' beliefs in their collective efficacy. There was ample evidence that this occurred and that it impeded the development of teachers' commitment to school improvement.

Not reported in the earlier literature is the positive effect of external diagnosis on consistency across the school district. By using a district template and communicating among themselves, diagnosticians increased between-school consistency in understanding the innovation, innovation language, and foci of improvement plans. Also not anticipated by earlier research was the constructive role of principals and district level agents in implementing the diagnosticians' prescriptions and moderating their messages. The principals removed much of the negative affect surrounding the message, shifting attention from teacher deficits to focus on instructional behaviors. Principals mitigated the negative effects of external diagnosis by engaging in distributed leadership practices that strengthened collaboration and beliefs about collective capacity. The effect of diagnosis on teacher reflection was mediated by other enabling factors such as principal coordination of in-school professional development and provision of professional learning materials.

Implications for Practice

External diagnosis had positive effects on low-achieving schools, but at considerable cost to the self-efficacy beliefs and commitment of teachers. External diagnosis could be improved so that the benefits (fearless and accurate identification of critical needs, increased consistency, and powerful pressure for change) could be accomplished at lower cost.

Add a self-study component to external diagnosis. External diagnosis in our site omitted self-study. School evaluation trends in England (Plowright, 2007), Belgium (Devos & Verhoeven, 2003), and Ireland (McNamara & O'Hara, 2006) emphasize the importance of self-evaluation as a precursor to external evaluation. External appraisers review the internal report, engage staff in its interpretation, verify its claims, and propose directions for improvement. Self-study increases staff understanding of appraisal criteria, stimulates the collection of credible data on organizational performance, and develops support for expert feedback.

Open the template to discussion. The diagnosis template was conceptually dense and contained challengeable assumptions. For example, the shortcomings of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwhohl, 1956) have become well known: it is limited to cognitive objectives; it does not address self-knowledge; it is based on a single dimension of growth (difficulty); its levels are not easy to distinguish empirically; and the levels may not be in the right order (Kreitzer & Madaus, 1994). Updates of Bloom's taxonomy resolving these are available (Anderson et al., 2001; Marzano & Kendall, 2007). The diagnosis might have been more palatable if there had been dialogue at the school and district levels about the meaning and validity of the criteria of the template.

Deal with more of the sensitive issues. The diagnostic reports were tough on instructional practices, raising issues that might have been neglected if the diagnosis was entirely internal. But the reports said little about other sensitive issues that impede school improvement such as the community context in which the schools worked, previous attempts at school improvement that failed, leadership deficiencies, and issues unique to each school. A major advantage of external diagnosis is the freedom that needs to be taken to examine such issues.

Recognize school strengths. The template provided limited space to summarize school strengths. Some schools made substantial progress in an improvement effort launched two years earlier, even if their current achievement was low relative to the rest of the district. Recognizing these accomplishments would send a capacity-building message to schools and would enable them to recognize successful strategies that need to be replicated in subsequent improvement plans.

Directions for Future Research

In this study, we were unable to make causal claims about the achievement effects of external diagnosis. The ideal design would involve random assignment of schools to internal and external conditions. This is unlikely given that external diagnosis may be a waste of resources for schools with sufficient capacity to diagnose their own needs. More likely is a matched design in which a district that channeled resources into hiring external diagnosticians might be compared with a district with similar characteristics that spent the resources that could have gone to external diagnosis on other forms of school support such as within-school release time for teacher skill development.

Notes

[1]School effectiveness research identifies the correlates of student achievement such as specific instructional practices; school improvement research identifies processes that lead to the enactment of these practices in specific contexts. These research strands have developed and remain quite separate, despite attempts (Reynolds, Teddlie, Hopkins, & Stringfield, 2000) to bring them together. This article is framed in the school improvement tradition and does not overtly address school effectiveness research.

[2]Research on teacher efficacy is extensive, and a review of it is beyond the scope of this article. For detailed reviews see Ross (1998) and Tschannen-Moran, Wolfolk Hoy, and Hoy (1998).

References

Anderson, L.W., Krathwohl, D.R., Airaisian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R. et al. (2001). *A taxonomy for leaning, teaching, and assessing: A revision of Bloom's taxonomy*. New York: Addison Wesley Longman.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Bloom, B.S., Engelhart, M.D., Furst, E J., Hill, W.H., & Krathwohl, D. (Eds.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay.

Burch, P. (2007). Educational policy and practice from the perspective of institutional theory: Crafting a wider lens. *Educational Researcher, 36*(2), 84-95.

Burkhardt, H., & Schoenfeld, A.H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher, 32*(9), 3-14.

Chapman, C., & Allen, T. (2006). Collaborative reform for schools in difficulty. *Improving Schools, 9*(3), 291-301.

Chay, K.Y., McEwan, P.J., & Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review, 95*, 1237-1258.

Coburn, C.E., & Talbert, J.E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*, 469-495.

Corcoran, T., Fuhrman, S.H., & Belcher, C.L. (2001). The district role in school improvement. *Phi Delta Kappan, 83*, 78-84.

Datnow, A., & Castellano, M. (2000). Teachers' responses to Success For All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37*, 775-799.

Datnow, A., Lasky, S., Stringfield, S., & Teddlie, C. (2006). *Integrating educational systems for successful reform in diverse contexts*. Cambridge, UK: Cambridge University Press.

Devos, G., & Verhoeven, J.C. (2003). School self-evaluation—conditions and caveats: The case of secondary schools. *Educational Management Administration and Leadership, 31*, 403-420.

Donaldson, L. (1999). *Performance-driven organizational change: The organizational portfolio.* Thousand Oaks, CA: Sage.
Elmore, R.F. (2004). *School reform from the inside out: Policy, practice and performance.* Cambridge, MA: Harvard Education Press.
Elmore, R., & Burney, D. (1998). *School variation and systemic instructional component in community school district #2, New York City*. Philadelphia, PA: University of Pennsylvania Consortium for Policy Research in Education.
Expert Panel on Early Reading in Ontario. (2003). *Early reading strategy*. Retrieved May 8, 2007, from: http://www.edu.gov.on.ca/eng/document/reports/reading/effective.html
Finnigan, K.S., & Stewart, T. (2008, March). *Principal leadership in low-performing schools.* Paper presented at the annual meeting of the American Educational Research Association, New York.
Firestone, W.A., & González, R.A. (2007, April). *Culture and processes affecting data use in school districts*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
Fullan, M.G. (1999). *Change forces: The sequel*. London: Falmer.
Fullan, M.G. (2002). *Leading in a culture of change.* San Francisco, CA: Jossey-Bass.
Fullan, M.G. (2005). Turnaround leadership. *Educational Forum, 69*(2), 174-181.
Fullan, M.G. (2006). The future of educational change: System thinkers in action. *Journal of Educational Change, 7*(3), 113-122.
Fullan, M.G., & Campbell, C. (2007, April). *The Ontario literacy and numeracy strategy*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
Gess-Newsome, J., Southerland, S.A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal, 40*, 731-767.
Gidney, R.D. (2000). *From Hope to Harris: The shaping of Ontario's schools.* Toronto, ON: University of Toronto Press.
Goddard, R.D., Hoy, W.K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3-13.
Hajnal, V., Walker, K., & Sackney, L. (1998). Leadership, organizational learning, and selected factors relating to the institutionalization of school improvement initiatives. *Alberta Journal of Educational Research, 44*, 70-89.
Harris, A. (2004, April). *Distributed leadership and school improvement: Leading or misleading?* Paper presented at the annual meeting of the American Educational Research Association, San Diego.
Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: Exploring the possible. *School Effectiveness and School Improvement, 17*(4), 40-424.
Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
Jong, R., Houtveen, T., & Westerhof, K. (2002). Effective Dutch school improvement projects. *Educational Research and Evaluation, 8*, 411-454.
Kelly, S., & Monczunski, L. (2007). Overcoming the volatility in school-level gain scores: A new approach to identifying value added with cross-sectional data. *Educational Researcher, 36*(5), 279-287.
Kennedy, M. (1997). The connection between research and practice. *Educational Researcher, 26*(7), 4-12.
Kerr, K.A., Marsh, J.A., Ikemoto, G.S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*, 496-520.
Kreitzer, A.E., & Madaus, G.F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L.W. Anderson & L.A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective: Ninety-third yearbook of the National Society for the Study of Education* (pp. 64-81). Chicago, IL: University of Chicago Press.
Leithwood, K.A., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times.* Milton Keynes, UK: Open University Press.
Leithwood, K., Mascall, B., Strauss, T., Sacks, R., Memon, N., & Yashkina, G. (2007). Distributing leadership to make schools smarter: Taking the ego out of the system. *Leadership and Policy in Schools, 6*(1), 37-67.
Linn, R.L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(4), 29-36.

analogies to learn science is commonplace and that teachers should work to promote children's abilities for analogical reasoning.

In their study of grade 5 students, Abell and Roth (1995) found that children's conversations with the teacher and their discussions among themselves showed students using analogies in their verbal reasoning and a few in their written models. This evidence led to speculation that young students could probably move toward more sophisticated perceptions of models. The understanding exhibited by the students was probably facilitated by guided experience with students' own models and the opportunity to compare the strengths and limitations of students' models before progressing to scientific models.

Gobert (2000) and Gobert and Clement (1999) studied grade 5 children as they read, drew diagrams, and talked about causal and dynamic processes related to plate tectonics. The children's drawings showed a progression in understanding about volcanic eruptions from Local Models (including simplistic causal mechanisms) to Mixed Models (including heat and movement) and Integrated Models (including many heat- and movement-related mechanisms). The researchers speculated that developing an understanding of integrated causal models (e.g., models of plate tectonics) could be facilitated through an initial emphasis on drawing static components of the model and progressing to causal and dynamic aspects of the phenomena (Gobert).

In general, researchers working with young children have been intrigued with what appears to be at least some children's abilities to think analogically. Analogical thinking is a critical component of analyzing and interpreting models, and researchers working with young children agree that more research is needed into how to build from these promising beginnings.

Significance of the Study

In this study, we seek to explore whether the recommendations for teaching students about models first and scaffolding their analogical thinking can result in young children engaging in more sophisticated thinking about and with models. Implications from this study could stand to influence current teaching strategies focused on how to support young children to think critically about models and how to select models that are pedagogically appropriate.

Study Questions

- What visual aspects of a globe do children use during analog-target mapping?
- Are there some visual aspects in particular that facilitate this mapping?
- To what extent do grade 5 children supported by instruction about scientific models engage in thinking beyond a naïve realist level about a globe?

Purpose of the Study

Recognizing that children's responses can allow insight into their analytical thinking, in this study we explored grade 5 (ages 10-12) children's ideas about models as they participated in lessons drawn from a researcher-developed resource entitled *Understanding Models in Science*. Specifically in this article, we focus on a subset of these data from the initial lesson and discuss how thinking about scale models (e.g., a globe) represents productive groundwork for under-

standing how children interact with models and develop meta-conceptual knowledge about models.

Methods

This exploratory study was conducted with 87 grade 5 (ages 10-12) children (47 boys, 40 girls). We contacted grade 5 teachers who had registered to attend a science program at a local informal science center, and four teachers volunteered to participate in the study. We provided the teachers (4 women) with the *Understanding Models in Science* teaching resource that we had developed. The teaching resource was intended (a) to teach children *about* models (e.g., how models are defined, why models are important in science, and how every model has strengths and limitations); and (b) to introduce the children to some fundamental ideas about the small, unseen particles that comprise matter. The resource incorporated the view that models can be organized according to intellectual demands that they place on children and that children should be progressively challenged to gain experience with simpler models (e.g., scale models) before moving on to more complex models (e.g., pictorial models of processes, Harrison & Treagust, 2000a). The data consisted of an open-ended worksheet completed by each child during the initial lesson on thinking about the globe as an example of a scale model of the Earth (see Appendix A).

The Understanding Models in Science Teaching Resource

The *Understanding Models in Science* teaching resource began with information for the teacher. This included a definition of the word *model* (models are visualizations in our minds, drawn on paper, or created in three dimensions that help us understand a real object, event, or idea. A model is a representation of something else). It also included information about how models are used in science (to make predictions, to communicate, and to explain), and ideas about the need to teach about models before teaching scientific concepts *through* models (because all models have strengths and limitations that need to be recognized).

The teaching resource began with teaching children about models (e.g., scale models, pictorial models). Four anchoring questions were continually revisited in the lesson contexts to assist children to develop an understanding about models (see Table 1). These anchoring questions were asked orally by the teacher and were included on the students' worksheets.

Features of the Globe

The globe was selected to fit with the children's anticipated prior knowledge. The mandated provincial curriculum provides many contexts in which children have direct, practical experiences with features of the Earth (e.g., rural and urban areas, land and water features) and the use of globes (both in social studies and in science).

We judged it essential for children to be familiar with the target and the analog before being asked to analyze correspondence between the two. This same assumption underpins other teaching strategies designed to help children understand models (Glynn & Takahashi, 1998; Harrison & Treagust, 2000b; Justi & Gilbert, 2002a, 2000b; Treagust et al., 1998). Further, Mathewson (1999) cautions that a barrier to some students' ability to analyze a model may come

Louis, K.S., & Marks, H.M. (1998). Does professional community affect the classroom? Teachers' work and student experience in restructuring schools. *American Journal of Education, 106*, 532-575.

Marks, H.M., & Printy, S.M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly, 39*, 370-397.

Marzano, R.J., & Kendall, J.S. (2007). *The new taxonomy of educational objectives. 2/E*. Thousand Oaks, CA: Corwin Press.

McNamara, G., & O'Hara, J. (2006). Workable compromise or pointless exercise? School-based evaluation in the Irish context. *Educational Management Administration and Leadership, 35*, 564-582.

Mintrop, H., & MacLellan, A. M. (2002). School improvement plans in elementary and middle schools on probation. *Elementary School Journal, 102*, 275-300.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socio-economically disadvantaged areas: A review of research evidence. *School Effectiveness and School Improvement, 15*(2), 149-175.

Plowright, D. (2007). Self-evaluation and Ofsted inspection: Developing an integrative model of school improvement. *Educational Management Administration and Leadership, 35*, 373-393.

Reezigt, G., & Creemers, B. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement, 16*, 407-424.

Reynolds, D., Teddlie, C., Hopkins, D., & Stringfield, S. (2000). Linking school effectiveness and school improvement. In In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 206-231). London: Falmer.

Ross, J.A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), *Research on teaching* (vol. 7, pp. 49-74). Greenwich, CT: JAI Press.

Ross, J.A., & Gray, P. (2006a). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education, 29*(3), 1-25.

Ross, J.A., & Gray, P. (2006b). Transformational leadership and teacher commitment to organizational values: The mediating effects of collective teacher efficacy. *School Effectiveness and School Improvement, 17*(2), 179-199.

Ross, J.A., McKeiver, S., Hogaboam-Gray, A. (1997). Fluctuations in teacher efficacy during the implementation of destreaming. *Canadian Journal of Education, 22*, 283-296.

Ross, J.A., Sibbald, T., Gray, P., & Scott, G. (2006). *Framework for a value for money evaluation of the Turnaround Teams program*. Final report to the Ontario Ministry of Education. Peterborough, ON: OISE/UT Trent Valley Centre.

Smith, J.P. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education, 27*, 387-402.

Smith, L., Ross, S., McNelis, M., Squires, M., Wasson, R., Maxwell, S., et al. (1998). The Memphis restructuring initiative: Analysis of activities and outcomes that affect implementation success. *Education and Urban Society, 30*, 296-325.

Stein, M.K., Hubbard, L., & Mehan, H. (2002, April). *Reform ideas that travel far afield: The two cultures of reform in New York City's District #2 and San Diego*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Stewart, D.W., Shamadasani, P.N., & Rook, D.W. (2007). *Focus groups: Theory and practice 2/E*. Thousand Oaks, CA: Sage.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*, 221-258.

Tschannen-Moran, M., Wolfolk Hoy, A., & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202-248.

Usher, E.L., & Pajares, F. (2005, April). *Inviting confidence in school: Sources and effects of academic and self-regulatory efficacy beliefs*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Vincent, A.A., Patterson, F., Buehler, A., & Gearity, B. (2006). School improvement planning in east Tennessee middle schools: A content analysis and perceptions study. *NASSP Bulletin, 90*, 277-300.

Watson, N., Fullan, M., & Kilcher, A. (2000, April). *The role of the district: Professional learning and district reform*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Wayman, J.C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112*, 549-571.

Wikeley, F., Stoll, L., Murillo, J., & De Jong, R. (2005). Evaluating effective school improvement: Case studies of programmes in eight European countries and their contribution to the effective school improvement model. *School Effectiveness and School Improvement, 16*, 387-405.

Zavdasky, H. (2007, April). *Improving education through a district-wide systemic model.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.