A comprehensive assessment of South Carolina's External Review Team Program

Paula Szulc Dominguez, Ed.D.
Craig Nicholls, Ph.D.
Barbara Storandt
Hezel Associates

Introduction

The federal No Child Left Behind Act (NCLB, P.L. 107-110) sparked a national approach to standards-based accountability that built on (and, at times, around) the actions individual states had been pursuing to varying extents for much of the previous decade (Mintrop, 2004). NCLB called for annual testing of all students in grades three through eight, a twelve-year timeline to achieve universal proficiency in reading and mathematics, identification of low-performing schools based on aggregated and disaggregated test scores, and administration of severe consequences for low-performing schools that fail to improve. Under the "new accountability" system (Fuhrman, 1999), schools are accountable to higher levels of the educational system (school districts or state education agencies), but are held accountable for producing specific levels or improvements in student learning outcomes, not for delivering designated educational inputs and processes (O'Day, 2002).

Since NCLB's passage four years ago, states have invested time, personnel, and considerable financial resources to develop and implement accountability systems that meet the law's requirements. As these accountability systems have been put into place and as increasing numbers of schools have been identified as "in need of improvement" (or some other designation to indicate a failure to meet externally mandated performance gains), the capacity of state education agencies (SEAs) to shift their responsibilities from compliance to intervention systems has been called into sharp relief. There are valid reasons to question SEAs' collective capacity to respond to increased accountability requirements. Over a decade ago, states' responses to other attempts to foster systemic reform -- the Goals 2000: Educate America Act (1994) and the 1994 Improving America's Schools Act (IASA), the reauthorization of Title I, Part A -- revealed widespread difficulties states encountered in implementing fully the accountability provisions of Title I and attaining policy cohesiveness (GAO, 2004; Riddle, 2005; Superfine, 2005). The earlier standards-based reform efforts revealed state education agencies' deficiencies in terms of the instructional expertise necessary to make a difference in troubled schools, the remedies that were put together to support low-performing schools, and the amount of assistance that states had budgeted to support the remedy or sanction side of accountability (Fuhrman, 1999).

With the new and heightened accountability requirements NCLB brings, states have the responsibility of labeling schools' (and districts') performance using a set academic standards, numeric academic and non-academic indicators. Along with this responsibility

comes the obligation of ensuring that all students have equitable opportunities to meet established performance standards, as well as of supporting teachers and administrators in struggling schools (Rennie Center, 2005).

SEAs have put into place a range of strategies and technical assistance programs that target schools identified as in need of improvement (Mintrop, 2003; U.S. Department of Education, 2004). Ideally, the reams of accountability data now available should make it possible for education leaders to determine the quality of such strategies and interventions (Doran, 2003). To shed light on how states can use school performance data to support failing schools, the present study examined a subset of underperforming schools in South Carolina in the wake of a series of state-supported interventions.[1] The study, part of a comprehensive evaluation of the South Carolina External Review Team Program, made extensive use of annual student performance data collected from schools in need of improvement, information that conceivably could be used to guide decisions for school and Program improvements.

Background: South Carolina's External Review Team Program

In South Carolina, the state's Education Accountability Act of 1998 (EAA) brought together provisions ranging from the development of detailed, grade level, specific academic standards in core content areas, the administration of annual standards-based assessments in grades 3 through 8, and steps to be taken for intervention in low-performing schools. An important outcome of the EAA was the shift in the direction of interventions: rather than focusing on *districts* (as the state had under its 1984 Education Improvement Act), EAA called for *schools* to be the target of technical assistance. This shift required that the state develop appropriate school-level technical assistance and the capacity to serve a greater number of entities. One component in South Carolina's arsenal of intervention strategies is the participation of external review teams (ERTs). In place since 2000, ERTs are small groups of skilled and experienced individuals who work closely with identified schools' staff, central offices, and local boards of trustees in the design of the school's plan, implementation strategies, and professional development training to improve student performance and increase the rate of student progress.

ERTs are assigned to all schools that receive an "unsatisfactory" academic performance rating, the lowest of the five categories on the EAA's school performance classification system. Additionally, schools that receive a "below average" performance rating (the next to the lowest category) may request an ERT visit. The ERT review is one in a series of related school improvement activities coordinated between the school and the South Carolina Department of Education, which is responsible for providing continuous follow-up and support activities to the underperforming school, including professional development, financial support, and on-site assistance.

[1] The work reported here was part of a larger evaluation study of the External Review Team Program funded by the South Carolina Education Oversight Committee (EOC) and completed in January, 2006. Further details can be found at http://www.sceoc.org/

The ERT review process is designed to diagnose areas of strength and weaknesses. A chairperson and a coordinator (a local contact person who is often the principal of the school under review) guide each ERT. Prior to conducting the on-site visit, all ERT members participate in a comprehensive training session: a one-day refresher session for veteran ERT members, with an additional follow-up training session for new ERT members. ERT members receive a stipend for their service.

ERT members spend a minimum of four consecutive days visiting the school, reviewing documents of student performance, attendance rates, and other pertinent data both before and during the on-site review, conducting interviews of parents, teachers, students, and principals, and observing every teacher. The Office of School Quality provides documents that guide the on-site data collection and analysis. Reviews are conducted in the fall and are concluded before the school's winter break.

At the conclusion of each review, the ERT compiles evidence-based recommendations, which are shared with the principal and are then assembled into a final report. After the approval of the recommendations by the State Board of Education, the Department of Education delineates the activities, support, services, and technical assistance it will provide to the school. With the approval of the state board, this assistance continues for at least three years, or as deemed necessary by the review committee to sustain improvement.

Methods

The current study took place as part of an evaluation of South Carolina's External Review Team (ERT) Program that was executed from April, 2005 to January, 2006. At the time the evaluation took place, approximately 80 schools had participated in the ERT Program since its inception in the 2001-02 school year and had put into place a variety of targeted interventions to support their academic success. The core question guiding the evaluation team centered on to what extent the ERT Program and related, subsequent interventions actually contributed to academic improvement? Or as Parrish et al. (2005) inquire, have the ERT Program's mechanisms as currently implemented been effective in promoting academic success? An ancillary question that stemmed from the evaluation activities was, to what extent can student accountability data be used effectively to assess Program impact?

From the onset of the study, the evaluation team realized that attributing the impact of the ERT Program to school improvement was, at best, a convoluted effort. As Hanushek and Raymond (2003, 2005) point out, assessing the impact of state accountability activities is anything but straightforward. For example, there is no consensus regarding how much time is required for change to become manifest in schools. Moreover, school performance does not typically improve in even and measurable increments, which frustrates attempts to detect progress (Elmore, 2003).

With the inherent limitations of the ERT Program's model and implementation in mind, the evaluation team chose to focus one of its activities on an analysis of school

improvement data on a subset of 62 schools that were originally visited in the 2001-02 school year (schools we termed "Cohort I"). The rationale underlying this decision was that this subset of schools have had the longest amount of time to develop and put into place the improvement strategies and actions that were originally identified through the ERT visit. For Cohort I schools, the evaluation team examined schools' absolute performance ratings in both mathematics and English language arts (and, for high schools, the percentage of students eligible for LIFE scholarships and graduation rates) for each school year between 2001 and 2005. We then conducted a detailed analysis of the recommendations made in ERT reports and carried out follow up interviews with ERT members and school teams at 17 of the 62 Cohort I schools to learn more about the implementation of the ERTs' recommendations.

Findings

Analysis of student performance data

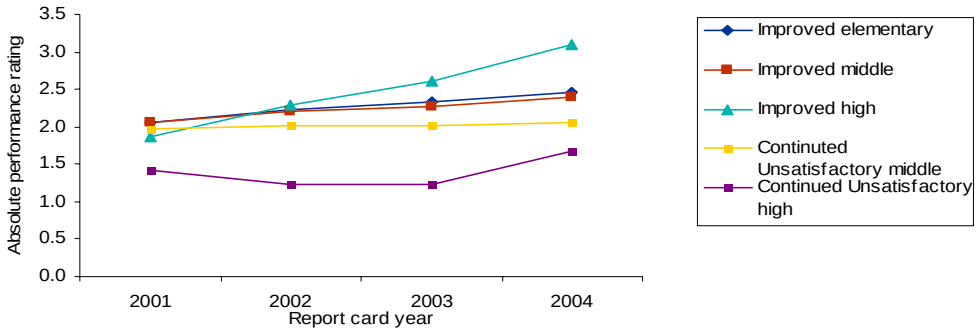
Using student performance data gathered from the 62 Cohort I schools' 2001, 2002, 2003, and 2004 school report cards, we characterized schools in one of three ways: as having *Improved* to satisfactory status and remained satisfactory, as showing *Continued Unsatisfactory* status, or as showing *No Perceived Pattern*. Using this system, we found that about 7 in 10 of the schools that were originally reviewed in the 2001-02 school year demonstrated some kind of discernable pattern (Table 1).

Table 1. Cohort I Schools' Improvement Patterns

	Improved	Continued Unsatisfactory	No Pattern	Total
Elementary Schools	6	0	1	**7**
Middle Schools	9	11	5	**25**
High Schools	12	5	13	**30**
Total	**27**	**16**	**19**	**62**

We then focused on those 43 schools for which we could distinguish a performance pattern. Beginning with data gathered during 2001, we examined the performance of the 27 Improved and 16 Continued Unsatisfactory schools in the 3 years that followed their initial ERT review. We found that the greatest gains in absolute performance ratings took place between 2003 and 2004, two years following the initial review (Figure 1).

Figure 1: Absolute performance ratings in schools following ERT visits



Some Continued Unsatisfactory schools showed greater improvements in their absolute performance ratings between 2001 and 2004 than some of the Improved schools, though they were unable to meet the threshold for satisfactory school performance. Since different methods are used to determine absolute performance ratings by school level, we break out comparisons by elementary, middle, and high schools.

1. Improved elementary and middle schools

All 6 Improved elementary and middle schools entered the ERT program with absolute performance ratings that closely bordered the threshold for a satisfactory designation. With only one or two tenths of a point away from a satisfactory designation in most cases, all Improved elementary schools achieved a satisfactory absolute performance rating in the year immediately following the initial ERT visit. One of these elementary schools improved beyond the level of "below average" to "average." By contrast, none of the 9 Improved middle schools achieved beyond the rating of "below average," despite sizeable gains in performance. The greatest gains in elementary and middle schools' absolute performance ratings took place between the 2003 and 2004 report card years (Figures 2 and 3).

Figure 2: Absolute performance ratings in Improved elementary schools

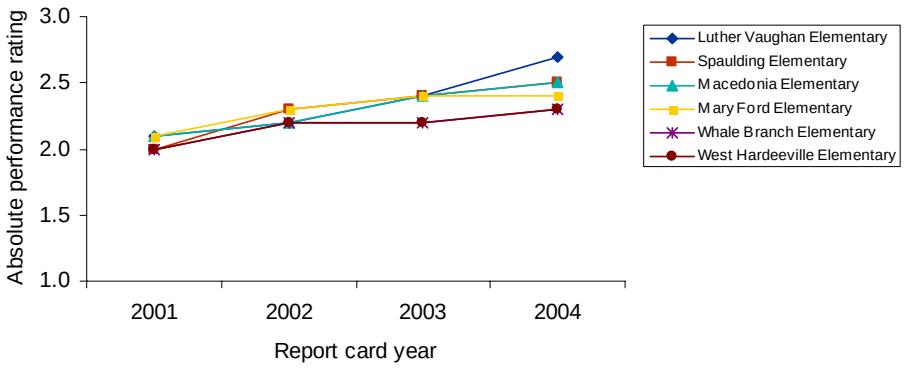
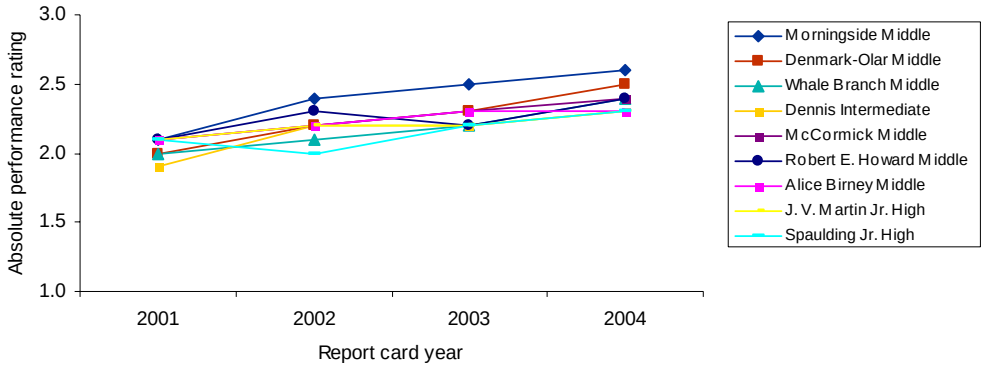


Figure 3: Absolute performance ratings in Improved middle schools

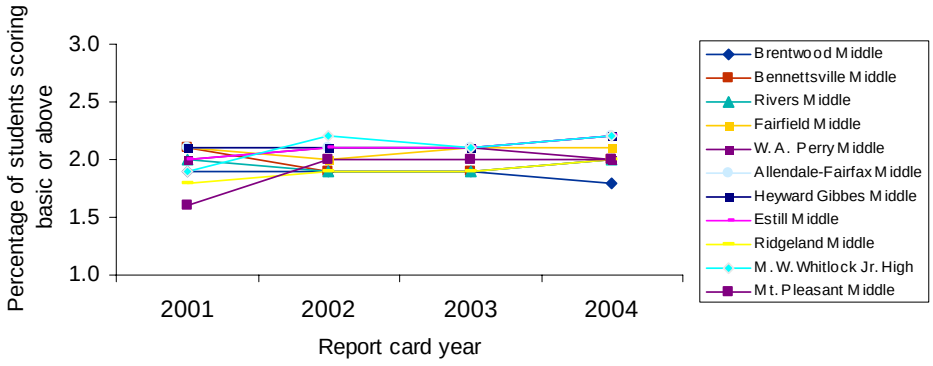


ELA performance among Improved elementary and middle schools showed the greatest gains in student performance between 2003 and 2004, whereas sizeable gains in math performance were made every year except for 2004.

2. Continued Unsatisfactory elementary and middle schools

No elementary school reviewed in 2001-02 fell into the Continued Unsatisfactory category. Eleven middle schools, however, demonstrated Continued Unsatisfactory performance following the 2001-02 ERT visit. These 11 middle schools demonstrated inconsistent patterns in student performance. For example, the greatest number and gains in absolute performance ratings took place in the year immediately following their ERT visit, when ERT recommendations had yet to be implemented. Absolute performance ratings either remained the same or showed only slight changes during subsequent years, when these middle schools had more time to implement recommendations and deploy technical assistance (Figure 4).

Figure 4: Absolute performance ratings in Continued Unsatisfactory middle schools



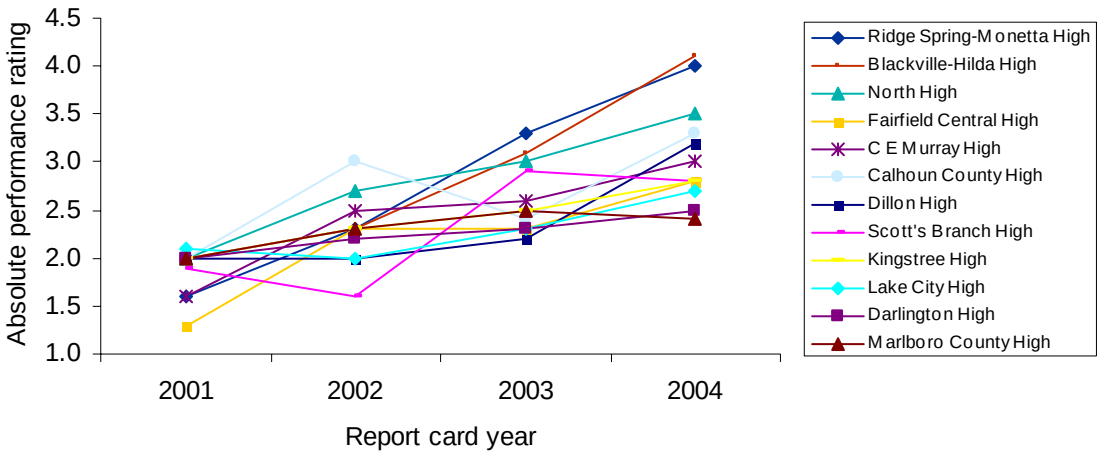
3. Improved and Continued Unsatisfactory high schools

Longitudinal changes in high school performance are more difficult to map than changes in elementary and middle schools. Whereas elementary and middle school performance is determined using student achievement on PACT scores, several performance measures determine a high school's performance rating, including student proficiency on the BSAP/HSAP exit exams, eligibility for LIFE scholarships, exit exam passage rates, and, after 2003, graduation rates. Sizeable improvements in any one of these areas, or cumulative smaller improvements, can result in considerably larger high school performance gains.

Improved high schools

More of the high schools from the 2001 cohort improved their performance ratings during the course of the ERT Program than elementary or middle schools (Figure 4 above). Further, despite earning the lowest absolute performance ratings of any school type during the first year of the ERT Program, Improved high schools showed the highest absolute performance ratings and the most improvement between 2001 and 2004. Improved high schools gained an average of 1.2 index points, from an average absolute performance rating of 1.9 ("unsatisfactory") in 2001 to 3.1 ("good") in 2004 (Figure 5).

Figure 5: Absolute performance ratings in Improved high schools



A word of caution should be inserted here. Sizeable gains in high school performance may be attributable to ERT processes and recommendations, or to comprehensive efforts to modify curricula, leadership and governance, and/or professional development. However, changes in the BSAP/HSAP exit exam administration and scoring, as well as the 2003 addition of graduation rate as a performance measure, suggest that this trend may also reflect modifications in the calculation of high school performance. Our analyses suggest that student proficiency on the ELA and math components of the BSAP/HSAP exit exams among Improved high schools showed no coherent pattern. Similarly, Improved high schools showed only marginal aggregate gains in tenth grade BSAP/HSAP passage rates, eligibility for LIFE scholarships, and graduation rates between 2001 and 2004.

Continued Unsatisfactory high schools

Continued Unsatisfactory high schools showed no coherent pattern in absolute performance ratings. Longitudinal ELA and math proficiency among Continued Unsatisfactory high schools showed inconsistent patterns between 2001 and 2004. However, the greatest gains in ELA proficiency and the most consistent gains in both ELA and math proficiency tenth grade passage rates on the BSAP/HSAP exit exam, and graduation rates (Table 2) took place between 2003 and 2004. These trends are consistent with the notion that there is a time lag between the implementation of ERT recommendations and measurable outcomes.

Table 2. Graduation rates in Improved and Continued Unsatisfactory high schools

	Improved	Continued Unsatisfactory
2003	74%	51%
2004	75%	55%

Analysis of ERT recommendations

To supplement the analysis of student performance data, we examined the ERT reports for 17 Cohort I schools (Table 3). Our analysis focused on the indicators that ERTs used in their reporting and on the technical assistance that was identified to remedy performance gaps. We were interested in determining whether different types of identified deficiencies exist across schools that have improved relative to those that remain unsatisfactory, and whether there were different patterns in the technical assistance that was deployed to remedy the observed problems.

Table 3. Schools included in the analysis of ERT recommendations

Improved Schools	2001 Rating	Continued Unsatisfactory Schools	2001 Rating
Ridge Spring-Monetta High	1.6	Baptist Hill High	1.0
C E Murray High	1.6	Whitlock Junior High	1.9
Lake City High	2.1	Heyward Gibbes Middle	2.1
Luther Vaughan Elementary	2.1	Allendale-Fairfax Middle	2.1
Marlboro County High	2.0	Fairfield Middle	2.1
Macedonia Elementary	2.1	Eau Claire High	2.0
Mary Ford Elementary	2.1	C A Johnson High	1.4
West Hardeeville Elementary	2.0	Bennettsville Middle	2.1
J. V. Martin Jr. High	2.1		

ERTs evaluate unsatisfactory schools using indicators clustered in four focus areas, categories that also organize ERTs' findings and recommendations:

- Leadership and Governance,
- Curriculum and Instruction,
- Professional Development, and
- Performance.

According to our analysis, ERTs consistently identified more schools with deficiencies in Curriculum and Instruction (Table 4). [2,3] Note that a series of statistical analyses indicated that apparent differences in the deficiencies in Improved and Continued Unsatisfactory schools are due to random variation.

Table 4. Average number of findings and recommendations, by indicator area and school improvement status

	School Type	Leadership & Governance	Curriculum & Instruction	Professional Development
Findings	Improved n = 9	2.2	3.4	1.4
	Continued Unsatisfactory n = 8	1.7	4.3	1.7
Recommendations	Improved n = 9	1.9[4]	3.4	1.9
	Continued Unsatisfactory n = 8	1.7	4.2	2.1

Table 4 lists the actual deficiency indicators most frequently cited by ERT's in 2001-02. Comparing the two groups, Improved schools and Continued Unsatisfactory schools shared only one indicator of deficiency – C&I 3.1: *There is no documentation to indicate that the curriculum development process involves staff, parents/guardians, members of the community, and students.* Although the other indicators all differ, these two groups of schools appear to share a common deficiency: a lack of involvement from key stakeholders (parents, staff, members of the community and students) in curricular and other aspects of decision making.

Despite this similarity, ERT findings in Improved and Continued Unsatisfactory schools do show substantive differences. Schools that remained Unsatisfactory appear to have deeper or more systemic problems, as evidenced by lack of procedures to support alignment with state academic standards, not using student data to inform curricula, and not engaging in planning based on research-supported practices. Improved schools tended to be cited more frequently for poorly matching instructional activities and resources with individual student needs or with state standards--specific problems that may be more readily addressed.

---

[2] The number of indicators in each focus area is: L&G = 39, C&I = 15, and PD = 8.

[3] Our comprehensive analysis of student performance, presented separately, precludes the need for analyzing the fourth focus area, Performance.

[4] Because ERTs base their recommendations on findings, ERTs also made more recommendations to Improved schools in the area of Leadership and Governance than in the areas of Curriculum and Instruction or Professional Development.

Table 5. "Findings" indicators most frequently associated with Improved and Continued Unsatisfactory schools

School Type	Findings Indicators	Frequency
Improved	L&G 12.3 – The local board of trustees does not take appropriate actions to assist in improving parental involvement	9
	L&G 11.2 – There is no evidence to confirm that instructional activities and assistance in the homework center are tailored to the specific needs of individual students	7
	C&I 5.1 – Instructional materials, resources, and learning activities do not evidence alignment with the curricula and local and state standards.	6
	L&G 5.1 – The school does not have an academic assistance initiative to support students with academic difficulties in all grades so that they are able to progress academically with their peers	5
	L&G 12.2 – The district superintendent does not take appropriate, specific actions to assist in improving parental involvement	5
	C&I 3.1 – There is no documentation to indicate that the curriculum development process involves staff, parents/guardians, members of the community, and students.	5
Continued Unsatisfactory	C&I 3.5 – The school leadership must direct the revision of curricula using data from student performance on state assessments.	10
	C&I 4.1 – There is no documentation to confirm that the district and school annually review longitudinal student performance data.	8
	C&I 3.1 – There is no documentation to indicate that the curriculum development process involves staff, parents/guardians, members of the community, and students.	7
	C&I 5.3 – There is no evidence to confirm that instructional apparatus and equipment in all laboratories, resource centers, and classrooms are maintained in good working condition.	7
	L&G 1.1 – School policies or procedures are not designed to support the use of academic achievement standards to assist the school and students in achieving higher levels of performance by aligning school standards and assessments with state standards and assessments.	6
	PD 1.1 – There is no evidence to confirm that school professional development is planned and designed by the faculty.	6
	L&G 7.5 – Proposals and plans in the district strategic plan and the school renewal plan are not derived from strategies found by education research to be effective.	5
	L&G 7.6 – The school renewal plan does not provide for an innovation initiative designed to encourage comprehensive approaches based on effective strategies identified in research literature.	5
	L&G 12.1 - The school renewal plan does not include parental involvement goals and objectives and an evaluation component for improving these programs.	5

Identifying areas of needed improvement and recommending ways to alleviate those deficiencies are just the initial steps to improve unsatisfactory school performance. As a result of ERTs' findings and recommendations, all but one of the 17 schools included in this subset received targeted technical assistance aimed at addressing unsatisfactory schools' areas of needed improvement. [5]

Given the emphasis on curriculum and instruction in the ERTs' findings and recommendations, it is not surprising that the majority of technical assistance provided to Improved and Continued Unsatisfactory schools in the sample also focused on Curricular and Instructional issues (Table 6).

[5] Eau Claire High School was the single exception.

Table 6. Technical assistance received, by indicator area and school improvement status

	School Type	Leadership and Governance	Curriculum and Instruction	Professional Development
Technical Assistance	Improved n = 9	3.1	4.9	2.4
	Continued Unsatisfactory n = 8	2.7	4.5	2.6

Schools that later demonstrated Improved performance received different types of targeted technical assistance than Continued Unsatisfactory schools. Further, the relationships between the findings/recommendations and the types of technical assistance provided to each group of schools is not clear. For instance, technical assistance is provided in a number of areas not specifically flagged in the deficiency findings. Improved schools most frequently received technical assistance in involving parents and the community, learning best practices for homework centers, and selecting appropriate instructional materials. Continued Unsatisfactory schools most frequently received technical assistance in understanding current educational research, planning and assessing professional development, and involving stakeholders in curriculum development.

Although the quantity and type of ERT findings, ERT recommendations, and technical assistance offered to schools subsequent to their initial year in the ERT Program appear to contribute to schools' longitudinal performance, it is difficult to identify consistent patterns. For example, while Ridge Spring-Monetta, C.E. Murray, and Baptist Hill High Schools showed the greatest net gains in absolute performance ratings, the number and types of ERT findings, ERT recommendations, and instances of technical assistance they received differ substantively. Similarly, the number and type of ERT findings, ERT recommendations, and technical assistance differed substantially between Fairfield Middle School, Eau Claire High School and C.A. Johnson High School, although all of these schools showed no net gain in absolute performance ratings between 2001 and 2004 (See Table 7).

Table 7. Findings, recommendations, and technical assistance, by indicator area and school improvement status

		Improved Schools									Continued Unsatisfactory Schools							
		Ridge Spring-Monetta High	CE Murray High	Lake City High	Luther Vaughan Elementary	Marlboro County High	Macedonia Elementary	West Hardeeville Elementary	Mary Ford Elementary	J.V. Martin Junior High	Baptist Hill High	Whitlock Jr. High	Heyward Gibbes Middle	Allendale Fairfax Middle	Fairfield Middle	Eau Claire High	CA Johnson High	Bennettsville Middle
Net Change in Absolute Perf. Rating 2001-2004		**2.4**	**1.4**	**0.6**	**0.6**	**0.4**	**0.4**	**0.3**	**0.3**	**0.2**	**1.0**	**0.3**	**0.1**	**0.1**	**0.0**	**0.0**	**0.0**	**-0.1**
Findings	L&G	4	21	11	3	11	9	10	4	6	13	11	7	6	17	0	6	10
	C&I	3	5	5	3	2	11	7	0	15	15	7	8	8	13	0	8	1
	PD	0	0	2	0	0	4	4	1	3	6	1	0	5	4	0	0	3
Recommendations	L&G	5	22	9	4	8	10	6	11	7	19	9	4	2	20	0	10	10
	C&I	4	6	5	5	2	10	1	7	13	17	6	10	2	8	0	10	2
	PD	0	0	2	0	0	3	3	4	3	6	1	0	1	2	0	0	3
Technical Assistance	L&G	8	38	13	7	15	6	5	19	8	22	1	6	14	39	0	12	13
	C&I	6	12	9	10	3	3	0	20	11	14	1	8	14	25	0	20	6
	PD	0	0	5	0	0	0	2	9	3	8	0	0	7	6	0	0	0

Discussion

One of the most difficult issues to emerge from state accountability systems is the question of how to best intervene in the most academically vulnerable schools, those schools whose student achievement data reveal that they dramatically and persistently fail their students (McRobbie, 1998). This situation represents an opportunity for researchers to step in and use their skills to address and answer core questions policy makers face. What distinguishes those underperforming schools that successfully improve from those that fail to move forward? Why do some underperforming schools remain "stuck" while others manage to improve? What kind of information allows us to understand what schools need to move forward?

If nothing else, state accountability systems generate a wealth of annual student performance data, information that may shed light on how to refine state improvement strategies. So how can student achievement data like the kind stemming from the External Review Program in South Carolina inform state education leaders? We see four main considerations for education policy. First of all, the data here suggest that all unsatisfactory schools are not the same and that large-scale improvement efforts should not treat them as though they are. Unsatisfactory schools that appear to share a common set of deficiencies do not, in fact, react uniformly to the technical assistance that is put into place. Some improve, others simply do not, and some vacillate without apparent direction. These findings are consistent with an emerging body of research that relies on analysis of student achievement data (e.g., Kannapel et al., 2005; Parrish et al., 2005) and suggests that even large-scale improvement strategies that may work in some underperforming schools may not exert a measurable affect in other underperforming settings. The optimum mix, range, and levels of state-supported technical assistance for schools identified as in need of improvement are still unknown (Goertz et al., 2001). Given the ambiguity, states have opted for undifferentiated strategies and programs. Our South Carolina study underscores the notion that such undifferentiated approaches do not work. Unfortunately, our work stops short of suggesting how programs such as the External Review Team Program can be modified to suit particular kinds of unsatisfactory schools.

A second message for state education policy leaders is that time will tell, and that accountability plans should accommodate a lag in measurable outcomes. Our data suggest that schools in need of improvement may require more than a year following the implementation of technical assistance to demonstrate measurable change in performance. Elmore (2003), too, contends that performance often lags behind practice – schools are improving just as much when they are in the process of changing practices, even though no measurable impact can be discerned.

A third message for policy makers is that changes to non-academic indicators used as part of an accountability system may frustrate attempts to track improvements in academic outcomes. In the case of the data we saw, new high school indicators added to the mix affected schoolwide performance ratings, even in the absence of real improvements in core curricular areas.

Finally, education policy makers should ensure that a variety of data are used to build on and help clarify findings emerging from analyses of student and school achievement data. Our study of the ERT Program included, for example, analyses of ERT reports and recommendations, as well as site visits. These additional activities suggested that curriculum and instruction may assume an especially strong role in school improvement, information that could not be gleaned easily from performance data. Note, however, that the ERT Program itself was structured to provide limited classroom and instructional data, which may have prevented a clearer portrait to emerge regarding consistent differences across schools in their instructional strengths and weaknesses. Others (e.g., Elmore & McLaughlin, 1988) have also noted that schools' core instructional routines typically remain safe from external scrutiny. Instead, schools react to demands for reform with incremental add-ons to services, rather than substantial and meaningful transformation (Mintrop & MacLellan, 2002; Tyack & Cuban, 1995). Our interviews and site visits also yielded data that underscored how the mobility of school principals and good teachers can undermine school improvement. As Mintrop (2003) points out based on his research in Maryland and Kentucky, widespread organizational instability undercuts schools' ability to achieve continuous improvement and should be addressed even prior to instructional improvements. That is, these schools require "baseline stabilization first before they [can] embark on ambitious instructional reforms" (p. 23). The South Carolina underperforming schools we focused on were similarly characterized by teacher instability, principal change, and student mobility, factors that may have undercut their attempts at instructional improvement.

The continued lack of performance of those students at the very bottom of academic achievement represented one of the primary reasons for the federal government's requiring each state to develop a test-based accountability system under NCLBA (Hanushek & Raymond, 2005). With all of the resources that states have invested in developing, training, and launching large-scale improvement and intervention efforts like the ERTs we studied, it stands to reason that education researchers should spend time considering whether and how such efforts have contributed to schools' performance. Each study that makes use of accountability-derived data has the potential to contribute to our understanding of how to turn around failing schools, and the chance to help policy makers at the state level develop, test, and refine large-scale improvements.

Bibliography

Elmore, R. (2000). *Building a new structure for school leadership*. The Albert Shanker Institute. Retrieved online February 23, 2006 from http://www.shankerinstitute.org/Downloads/building.pdf.

Elmore, R. (2003). A plea for strong practice. *Educational leadership*, vol. 61, no. 3. pp. 6-10.

Elmore, R. & McLaughlin, M. (1988). *Steady work: Policy, practice, and the reform of American education.* Santa Monica, CA: Rand Corporation.

Fuhrman, S.H. (1999). The new accountability. CPRE Policy Brief, RB-27.

Goertz, M. E., Duffy, M.E., & Le Floch, K.C. (2001). *Assessment and accountability systems in the United States: 1999-2000*. CPRE Research Report Series RR-046. Retrieved online March 13, 2006 from http://www.cpre.org/Publications/rr46.pdf.

Government Accountability Office (2004). *Education needs to provide additional technical assistance and conduct implementation studies for school choice promotion*. GAO report to the Secretary of Education (GAO-05-7). Washington, DC: Government Accountability Office. Retrieved online February 23, 2006 from http://www.gao.gov/new.items/d057.pdf.

Hanushek, E. A. & Raymond, M.E. (2005). Does school accountability lead to improved student performance? *Journal of policy analysis and management*, vol. 24, no. 2, pp. 297-327.

Kannapel, P.J., Clements, S.K., Taylor, D., & Hibpshman, T. (2005). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: Pritchard Committee for Academic Excellence. Retrieved online February 22, 2006 from http://www.prichardcommittee.org/Ford%20Study/FordReportJE.pdf.

McRobbie, J. (1998). *Can state intervention spur academic turnaround*? Retrieved online March 3, 2006 from: http://www.wested.org/pub/dpcs/policy/intervention.htm.

Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 11 (3). Retrieved online February 23, 2006 from http://epaa.asu.edu/epaa/v11n3.html.

Mintrop, H. & MacLellan, A.M. (2002). School improvement plans in elementary and middle schools on probation. *The elementary school journal*, vol. 102, no. 4, pp. 275-300. Chicago, IL: The University of Chicago.

O'Day, J.A. (2002). Complexity, accountability, and school improvement. *Harvard educational review*, vol. 72, no. 3. Retrieved online March 13, 2006 from http://www.gse.harvard.edu/hepg/oday.html.

O'Day, J.A. & Smith, M.S. (1993). Systemic reform and education opportunity. In S.H. Furhman (ed.), *Designing coherent education policy*. San Francisco: Jossey-Bass.

Raymond, M.E. & Hanushek, E.A. (2003). Shopping for evidence against school accountability. In *Developments in school finance: 2003*, pp. 119-130. Retrieved online February 22, 2006 from http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/DevSchFin_2003.pdf

Rennie Center (2005). *Reaching capacity: A blueprint for the state role in improving low performing schools and districts*. Retrieved online February 27, 2006 from http://www.renniecenter.org/research_docs/0504_ReachingCapacity.html.

Riddle, W. (2005). *Adequate yearly progress (AYP): Implementation of the No Child Left Behind Act*. Congressional Research Service Report for Congress, Order Code RL 32495. Washington, DC: Library of Congress.

Superfine, B.M. (2005). The politics of accountability: The rise and fall of Goals 2000. *American Journal of Education*, v. 112, pp. 10-44.

U.S. Department of Education (2004). *LEA and school improvement: Non-regulatory guidance*. Washington, DC: Author. Retrieved online February 22, 2006 from http://www.ed.gov/policy/elsec/guid/schoolimprovementguid.pdf.