

Inside the Black Box of School Reform: Explaining the how and why of change at *Getting Results* schools

Dennis McDougall^{a*}, William M. Saunders^b and Claude Goldenberg^c

^aUniversity of Hawai'i, USA; ^bPearson Achievement Solutions, USA; ^cCalifornia State University Long Beach, USA

This article reports key findings from a process-focused external evaluation that compared a subset of *Getting Results* project schools and comparison schools in order to understand the dynamics of school-wide reform efforts at these primary schools. Findings shed light on the “black box” of school reform and illuminate the limited empirical basis for understanding the inner workings of most reform efforts. We describe how *Getting Results* Model elements—goals, indicators, assistance, leadership, and settings—worked in concert to improve teaching and learning at project schools. We also describe factors that inhibited and promoted change, as well as implications for how whole-school reform might be accomplished through purposeful manipulation of these essential change elements.

Keywords: *Change; Evaluation; Process; Reform; Schools*

Introduction

Whole-school reform proliferated in the United States (U.S.) during the 1990s and reached a high point with passage of Public Law 105-78, commonly known as “Obey-Porter”, in 1997. The bill authorised the Comprehensive School Reform Demonstration Program and triggered hundreds of whole-school reform efforts nationwide. Many promising approaches to whole-school reform have been developed over the past 15 years (American Institutes for Research, 1999; Borman, Hewes, Overman, & Brown, 2003; Desimone, 2002; Slavin & Fashola, 1998). Even before this recent

*Corresponding author. University of Hawai'i, Department of Special Education, 124 Wist Hall, 1776 University Avenue, Honolulu, HI 96822, USA. Email: mcdougal@hawaii.edu

wave of whole-school reform models, educators have for years attempted to find ways to improve entire schools (see, e.g., the “effective schools” research of the 1970s and 1980s; Bliss, Fireston, & Richards, 1991). However, the empirical basis for understanding the actual process of school reform is very limited. Few studies have examined the effects of reform models within experimental or quasi-experimental research designs that would permit clear conclusions about the effects of reform models on student outcomes. Even fewer studies have looked directly at the process of reform to examine prospectively the dynamics leading to school improvement. Fullan (2000) reviewed research on schools with effective collaborative school cultures and noted, “We know nothing about how these particular schools got that way, let alone how to go about producing more of them” (p. 582). Although the literature systematically examining the processes and outcomes of school improvement efforts has grown enormously over the past decade (e.g., Berends, Kirby, Naftel, & McKelvey, 2001; Borman et al., 2003; Desimone, 2002), the problem identified by Sarason (1972) 35 years ago is still largely with us: We have very little direct observational data to document *how* schools change from being less to more effective in educating their students. The problem is significant because without detailed knowledge of *how* schools change, we know little about *why* school reform efforts obtain the results they do (Desimone). This is the “black box” problem we address in this article.

The study we report here contributes to the literature on school change for diverse and traditionally low-achieving populations of students. We report findings from a qualitative, prospective, external evaluation (McDougall, 2002) that illuminated the internal workings of a whole-school reform project. The *Getting Results* (GR) Model that guides this school change project is comprised of five interdependent elements—goals, indicators, assistance, leadership, and settings. The evaluation study answers a set of five questions about model implementation, processes, and outcomes. We discuss findings in relation to school reform and the importance of understanding its inner workings.

Background and Basis of the GR School Change Model

In a previous project, we developed a school change model that was instrumental in producing substantial changes in teaching and learning at one pilot site—a primary school that served primarily Latino children and families, in Southern California (Goldenberg, 2004; Goldenberg, Saunders, & Gallimore, 1996; Goldenberg & Sullivan, 1994). We refer to this model as the *School Change/Getting Results* Model, or, more simply, the GR Model. Over a 6-year period, the pilot school shifted from being the lowest achieving school in the school district to surpassing district averages on both standardised tests and performance-based assessments. Fashola, Slavin, Calderón, and Durán (1996) identified the GR Model as one of only three school change models with demonstrated effectiveness in majority Latino schools.

The GR Model utilises five elements to leverage changes in educators’ instructional behaviours and attitudes, and student outcomes. These elements include: (a) goals

that are set, shared, and articulated explicitly by the school community; (b) meaningful indicators that measure progress toward goal attainment; (c) assistance by capable others from within and outside the school; and (d) focused leadership that supports and pressures goal attainment. “Settings” is a fifth element and a super-ordinate concept in the model. Implementing the GR Model requires establishing new settings and reformulating existing settings for educators to meet, collaborate, receive assistance, and do important, instructional tasks (Sarason, 1972; Tharp & Gallimore, 1988). Within the context of these settings, the other change elements work in concert to improve teaching, learning, and achievement in any targeted curriculum or subject. In working with schools, we discovered a need to focus specifically and consistently on creating and sustaining concrete settings, at each school, where change elements are operationalised in effective ways—ways that permit educators to do the very important work of improving teaching and learning.

Research provides strong support for GR Model elements. Classic, and more recent, educational research documents the importance of explicit academic goals (Carter & Chatfield, 1986; Peterson & Lezotte, 1991). Goals are important for achieving long-term change and substantive improvements because they are vital for maintaining a coherent and stable student-centered vision. Indicators that measure progress toward agreed-upon goals reinforce the importance of the learning goals and help teachers and administrators gauge their goal-directed efforts. Consistent use of achievement indicators is related to improvements in student outcomes (Brophy & Good, 1986; Edmonds, 1979). Assistance from fellow professionals (Lieberman, 1988a, b; Little, 1982; Rosenholtz, 1991), including training from consultants (Joyce & Showers, 1983), is essential for successful reform (Saunders et al., 2001). Successful professional development models go well beyond “one-shot” workshops and presentations that attempt to train teachers in short order (Goldenberg & Gallimore, 1991). Professional development must be seen as long-term assistance to help educators acquire knowledge and skills essential for accomplishing agreed-upon goals (Darling-Hammond, 1997). Finally, leadership must both support and pressure changes necessary for goal attainment (Fullan, 1993). These two leadership dimensions complement each other, producing a creative tension that is perhaps the most elusive but important aspect of leadership (Blase, 1987; Bliss et al., 1991; Miles, 1983). Effective school leadership has long been recognised as fundamental to creating more effective schools, regardless of the cultural or linguistic background of students (August & Hakuta, 1997). The GR Model treats all five elements, including leadership, as highly interdependent. Strong leadership artfully combines pressure and support in a way that moves schools relentlessly toward accomplishing student achievement goals, utilising indicators, cultivating assistance and collaboration, and building productive school settings.

Current “Scale-up” Phase of the GR Project

The most current GR Model project “scaled up” our previous school reform efforts and achieved simultaneously, at nine schools, improvements in teaching and

learning similar to those achieved at the original pilot school. We examined prospectively a model of school change within a quasi-experimental research design. Fifteen elementary schools participated during this scale-up phase—nine schools that used the GR Model, all of whom entered the study and initiated GR Model implementation voluntarily, and six demographically similar schools in the same district that agreed to be comparison sites. At the study's inception, GR schools and comparison schools, as two groups, had nearly identical mean achievement scores on annual, statewide, standardised tests. These K-5 schools served predominantly poor and working class Hispanic communities. GR replication sites utilised modest local funding—US\$100,000 per school over a 3-year period—to support implementation. A research grant supported GR staff's assistance to schools, as well as data collection and analyses.

As members of the same, large, urban school district, all schools (both GR and comparison) functioned under the same reform umbrella set forth by California state legislation and local school district mandates. Four major reforms were underway in this local school district when the scale-up project began in 1997: (a) class size reduction (20 students to one teacher) at grades K-3; (b) state-established and district-implemented content standards for language arts and mathematics; (c) annual achievement accountabilities and reporting based on state-mandated, standardised tests; and (d) common curriculum, mandated training, and school-based coaches for reading/language arts and mathematics. Primary schools in this district showed steady increases in student achievement since the late 1990s, very probably related to these major district and state reform efforts. Our quantitative evaluation of achievement impact (Saunders, 2003), focused on the additive effects of the GR Model, and the extent to which outcomes and increases over time at GR schools surpassed those of other, comparable, schools in this large, urban district.

Putting GR Model elements into practice. In brief, GR Model implementation established several settings and processes designed to ensure effective application of model elements—goals, indicators, assistance, and leadership. Settings included Academic Achievement Leadership Teams (AALTs or ALTs), Grade Level Team (GLT) meetings, and GR Principals' meetings. GR also established beginning, middle, and end-of-year (BME) assessments in reading, writing, and oral language proficiency. GR staff provided on-site assistance to support school efforts to establish and maintain these settings. GR staff also provided annual, 3-day, beginning-of-year leadership training institutes for school teams, and a 1-day, mid-year follow-up. GR staff spent approximately one day every two weeks at each GR school, met monthly with principals, and facilitated many of the aforementioned GR settings. Schools also used materials developed and tested by GR staff, including rubrics and checklists to evaluate GR implementation; beginning and advanced training manuals for ALTs; a series of three videotapes that described the GR Model and illustrated various settings and processes; reading, writing, and oral proficiency

assessments for schools that did not already have such assessments; worksheets and guidelines for analysing achievement data and student work samples; and guidelines and training modules for GR coaches.

Academic achievement at GR and comparison schools. During this scale-up phase, GR schools ($N = 9$) showed significantly greater gains in academic achievement than comparison schools ($N = 6$) across 5 years. The following data are based on test scores averaged across Grades 2–5 and across reading, mathematics, language, and spelling subtests on the Stanford Achievement Test, 9th edition (Stanford 9; Harcourt Educational Measurement, 1996). Mean achievement for GR schools, in Normal Curve Equivalent (NCE) units, increased from 32.8 ($SD = 4.8$) in 1997 to 48.3 ($SD = 4.2$) in 2002, a net increase of 15.5 NCE units. Corresponding National Percentile Ranks increased from 21st to 47th. In contrast, NCE means for the six comparison schools increased from 32.0 ($SD = 6.3$) in 1997 to 43.0 ($SD = 6.0$) in 2002, a net increase of 11 NCE units. Corresponding National Percentile Ranks increased from 20th to 37th. Whereas achievement, as measured by NCE means, was nearly identical at GR and comparison schools in 1997 (32.8 versus 32.0), by 2002, achievement at GR schools exceeded that of comparable schools by more than 5 NCEs (48.3 versus 43.0)—an adjusted effect size of 0.75. An effect size of 0.75, in this case, means that GR schools scored, on average, 0.75 SD units higher than the average achieved by schools in the comparison group. This magnitude of effect is considered in the high-moderate to large range (Cohen, 1988), and it compares very favorably with effect sizes of other school reform efforts, most of which are below 0.40 (Borman et al., 2003).

Achievement also increased more rapidly for GR schools than for the total population of elementary schools ($N = 600+$) in the district. Mean NCEs district-wide for Grades 2–5 increased from 36.5 in 1997 (nearly four points higher than the GR schools) to 47.3 in 2002 (one point lower than GR schools). In sum, GR schools increased an average of 15.5 points, six comparison schools increased by an average of 11.0 points, and schools district-wide increased by an average of 10.8 points. See Saunders (2003) for a more comprehensive and technical presentation of achievement results.

Methods

GR project leaders hired an external evaluator who conducted an independent, on-site, process evaluation, throughout the 2001–02 academic year (McDougall, 2002). The evaluator investigated qualitatively the how and why of school change processes and achievement gains at seven, purposefully selected, demographically similar, case-study schools—four of nine GR schools plus three of six comparison schools (see Table 1). The evaluator used qualitative methods, a comparative case-study research design, and rubric-based coding and ratings to answer the following research questions (RQs):

- RQ 1: To what extent is the GR Model implemented in GR project schools?
- RQ 2: To what extent does the GR Model establish processes discernable at GR, but not comparison, schools?
- RQ 3: What does implementation of the GR Model do to impact student achievement?
- RQ 4: What helps and hinders schools' implementation of the GR Model?
- RQ 5: What has changed, as a result of GR Model implementation, from the perspectives of participants and the external evaluator?

Data Collection and Analyses

Data collection included audio-taped and transcribed teacher focus groups and principal interviews; observations and field notes of GLT meetings, ALT meetings, school-wide faculty meetings, professional development sessions, and principals' meetings; and document retrieval. The evaluator collected additional information from spontaneous contacts with participants. In addition, the evaluator collected functionally similar types and amounts of information from each case-study school by observing and participating in over 100 events, in formal and informal settings, at GR and comparison schools.

Data analyses included qualitative analysis, plus GR Model, rubric-based, coding and rating of data from the aforementioned sources. The evaluator analysed data during and after a 9-month data collection period, and used: (a) the constant comparative method to formulate and refine findings (Glaser & Strauss, 1967); (b) triangulation to corroborate findings from multiple data sources, across individuals, time, and settings (Denzin, 1978; Miles & Huberman, 1994); and (c) member checks whereby participants provided feedback on emerging findings (Goffman, 1959; Kvale, 1996; Lincoln & Guba, 1985; Manning, 1997; Taylor & Bogden, 1998). The evaluator also used a transformative strategy (Creswell, 2003) to code and rate data.

Coding and rating episodes. The evaluator used the GR Model Rubric (Table 2) and more detailed operational charts to code and rate episodes from all data sources. Episodes were narrative data that contained information pertaining to GR elements. Episodes ranged in length from a single sentence to multiple paragraphs. The evaluator used applicable elements and corresponding descriptive criteria of the GR Model, in Table 2, to code all episodes. For example, episodes coded *L-A* contained information that pertained to descriptive criterion *A* (i.e., leadership's consistency in focusing on school-wide, academic achievement goals) of the leadership (*L*) element. Next, the evaluator rated the coded episodes 4, 3, 2, or 1 to indicate the rubric level that applied to the information contained in the episode. Some episodes included information that pertained to adjacent rubric levels. Such episodes required "in-between" ratings, such as 2 to 3. Thus, the evaluator used seven rubric-based options to rate episodes: 1, 1 to 2, 2, 2 to 3, 3, 3 to 4, and 4. Initial coding and rating

Table 1. Demographic and descriptive information for case-study schools

	GR Project schools ($N = 4$)				Comparison schools ($N = 3$)		
	Oak	Pine	Elm	Fir	Kam	Hall	Lot
Schedule	Year-round multi-track	Traditional single-track	Year-round multi-track	Year-round multi-track	Year-round multi-track	Year-round multi-track	Year-round multi-track
Students enrolled	1,204	557	1,388	1,211	1,259	1,123	1,249
Title I school ^a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students/class (M)	20.7	17.4	21.4	20.5	21.4	21.4	20.8
Students/computer	55	10	15	14	11	23	57
Classes on Internet	0	0	0	2	88	6	0
English Learner ^b (%)	79	44	72	76	61	72	76
Languages of EL	S = 77	S = 35	S = 66	S = 69	S = 61	S = 72	S = 76
Students ^c (%)	Ar = 0.6	Ar = 5	Kh = 3	Ar = 6	All other = 0	P = 0.1	All other = 0
	P = 0.5	R = 1	Ar = 2	Ko = 0.3		Ar = 0.1	
	F = 0.2	Ko = 1	T = 0.5	T = 0.2			
Free/reduced meals (%)	95	73	92	97	93	95	98
CalWorks ^d (%)	22	19	29	26	13	16	13
Compensatory Ed. ^e (%)	91	104	99	97	95	97	98
Student ethnicity ^f (%)	H = 91	H = 55	H = 83	H = 84	H = 98	H = 98	H = 99
	W = 4	W = 27	W = 5	W = 11	W = 2	W = 1	Af = 0.2
	Af = 3	Af = 14	Af = 5	Af = 3	F = 0.2	Af = 0.5	W = 0.1

Table 1. (Continued)

	GR Project schools (<i>N</i> = 4)				Comparison schools (<i>N</i> = 3)		
	Oak	Pine	Elm	Fir	Kam	Hall	Lot
Teacher ethnicity ^f (%)	As = 1	As = 2	As = 1	As = 0.1	As = 0.1	As = 0.3	F = 0.1
	W = 52	W = 63	W = 57	W = 62	H = 60	H = 62	H = 45
	H = 41	H = 20	H = 30	H = 25	W = 25	W = 28	W = 39
	Af = 5	Af = 9	Af = 7	As = 6	F = 5	Af = 5	As = 11
	As = 2	As = 9	As = 1	Af = 3	As, Af = 3	F = 3	Af = 3
Teachers w/o full credential ^g	26/57 = 46%	13/33 = 39%	18/67 = 27%	19/63 = 30%	22/62 = 35%	19/56 = 34%	25/65 = 31%

Notes: Percentage calculated using the total number of students enrolled (or teachers employed) at the school, rounded to the nearest whole number unless otherwise indicated. ^aQualifies for federal funds based on a high-poverty student body. ^b“English Learners” (EL) (i.e., former term used was “LEP” or students with limited English proficiency). ^cLanguage: Ar = Armenian, F = Farsi, Kh = Khmer, Ko = Korean, P = Punjabi, R = Russian, S = Spanish, T = Tagalog. ^dCalWorks = California Work Opportunity and Responsibility to Kids; formerly Aid to Families with Dependent Children. ^eProgramme for educationally disadvantaged students. ^fEthnicity: Af = African American, As = Asian, F = Filipino, H = Hispanic, W = White. ^gWithout (w/o) full credential = emergency credential/hire, waiver, pre-intern, District intern, University intern; Data reported or derived from www.ed-data.k12.ca.us/school/schoolprofile; 1999–2000 = most recent available data during evaluation study.

Table 2. GR Model Rubric

Elements	Level 1	Level 2	Level 3	Level 4
Goals <i>that are set and shared</i>	<p>No school-wide academic achievement goals are identified.</p> <p>All/nearly all teachers are unaware of goals or assume goals are unrelated to teaching.</p> <p>Teachers did not play some role in identifying goals.</p> <p>Individual teachers do not use the goals in planning instruction.</p> <p>No school-wide indicators of student achievement are used.</p>	<p>School-wide academic achievement goals identified; goals are fairly general.</p> <p>Most teachers may be unaware of the goals or assume goals are unrelated to their teaching.</p> <p>Few teachers (and others) played some role in identifying goals.</p> <p>Few individual teachers use the goals in planning instruction.</p> <p>Some school-wide indicators used to examine general trends (up/down) in student achievement.</p> <p>Most teachers may be unaware of indicators or do not view them as valid and useful sources of information.</p> <p>Few individual teachers use information from indicators to refine their instruction and curriculum.</p>	<p>School-wide academic achievement goals identified; goals are specific, concrete, and published. A</p> <p>Most teachers are aware of the goals and new teachers are usually told about goals.</p> <p>Teachers (and others) played some role in identifying goals.</p> <p>Some individual teachers use the goals in planning instruction.</p> <p>Specific school-wide indicators used to assess whether students are meeting identified goals. A</p> <p>Most teachers are aware of the indicators and they are somewhat interested in the information.</p> <p>Some individual teachers use the information to refine their instruction and curriculum. C</p>	<p>School-wide academic achievement goals identified; goals are specific, concrete, and published. A</p> <p>Teachers think the goals are important; new teachers are explicitly introduced to the goals. B</p> <p>Teachers (and others) helped develop or adapt the goals. C</p> <p>Teachers use the goals individually and collectively in planning instruction. D</p> <p>Specific school-wide indicators used to assess whether students are meeting identified goals. A</p> <p>Most teachers think the indicators are important and they view the information as valid and useful. B</p> <p>Most teachers use the indicators individually and collectively to refine instruction and curriculum. C</p>
Indicators <i>that measure success</i>				

Table 2. (Continued)

Elements	Level 1	Level 2	Level 3	Level 4
<i>Assistance from capable others</i>	No/very few opportunities for teachers to receive assistance.	Intermittent opportunities for assistance.	Regular, ongoing, school-wide opportunities for assistance.	Regular, ongoing, school-wide opportunities for assistance. A
	Provided by someone who does not have relevant experience, expertise, respect.	Provided by someone who may or may not have relevant experience and expertise.	Provided by someone with relevant experience and expertise.	Provided by someone with relevant experience and expertise who has or gains the respect of most teachers. B
	Unrelated to meeting school-wide achievement goals.	Mostly unrelated to meeting school-wide achievement goals.	Related to meeting identified school-wide achievement goals.	Focused directly on meeting identified school-wide achievement goals. C
	Unavailable/available briefly.	Available over short time period.	Available over extended period of time.	Available over extended period of time. D
	Leadership often distracted or does not focus on academic achievement.	Leadership generally focuses on academic achievement on school-wide.	Leadership focuses on meeting identified achievement goals school-wide.	Leadership focuses consistently on meeting identified ach. goals school-wide. A
<i>Leadership that supports and pressures</i>	Does not participate in establishing and revising plans.	Identifies general areas of need, recommends possibilities.	Helps establish a concrete plan to achieve goals.	Helps establish concrete plan, and over time continues to identify next steps. B
	Provides little if any encouragement and resources.	Provides some encouragement and available resources.	Provides encouragement and needed resources.	Provides ample encouragement and resources. C
	Rarely if ever encourages individual and collective responsibility	Usually does not encourage individual and collective responsibility.	Encourages individual and collective responsibility.	Clarifies and holds people accountable for their individual and collective responsibilities. D

Table 2. (Continued)

Elements	Level 1	Level 2	Level 3	Level 4
<i>Settings that allow staff to get important things done</i>	Time outside of classroom is used to clarify procedures and policies, and deliver required trainings.	Time outside of the classroom is regularly used to deliver necessary information, and discuss academic achievements.	Time outside of the classroom is regularly scheduled and generally used to work on things related to achieving academic goals. A	Time outside of the classroom is dependably scheduled and consistently used to get done important things related to achieving academic goals. A
	Teachers and administrators view meetings as an intrusion on their time.	Most teachers/administrators view meetings as part of their work, but do not value it as important.	Most teachers and administrators value time to meet, collaborate and get things done as an important part of their work. B	Teachers and administrators value time to meet, collaborate and get things done as an important part of their work. B
	Meetings are rarely effective and almost never efficient.	Some meetings are efficient and effective.	Most meetings are scheduled, planned, and carried out efficiently and effectively. C	Nearly all meetings consistently scheduled, planned, and carried out efficiently and effectively. C

Note: Bold letters A, B, C, and D identify descriptive criteria (scan horizontally) for each element of the rubric.

of all data sources produced 2,940 episodes. The evaluator subsequently reviewed all episodes, eliminated episodes that lacked sufficient content or clarity, and identified episodes that were overlooked during the initial coding and rating. This systematic review resulted in a net reduction of 309 episodes for a final total of 2,631 coded-rated episodes.

Assigning ratings to GR elements and schools. As was the case for individual episodes, the evaluator utilised seven rubric-based options to rate GR elements: 1, 1 to 2, 2, 2 to 3, 3, 3 to 4, and 4. The evaluator assigned element ratings, for each school, based on the mean and median ratings for that element's distribution of episode ratings (see Table 3). The evaluator assigned a single rating to an element if its episode rating distribution had a similar mean and median. However, if these measures of central tendency were not consistent (e.g., if the mean suggested a rating of 3 and the median suggested a rating of 3 to 4), then the evaluator assigned a dual rating for the element. The evaluator also assigned an overall rating of school change, to each school, based on the pattern of element ratings for each of the five GR Model elements. The bold numbers in Table 3 highlights rating patterns.

Reliability of episode ratings. Secondary evaluators conducted reliability checks of rubric-based episode ratings. Overall inter-rater agreement for *exact* agreements equaled 88%, for 104 episodes selected at random from 2,631 total episodes, using the formula "agreements divided by agreements plus disagreements multiplied by 100%". Overall inter-rater agreement for exact plus adjacent agreements equaled 100% using the same formula.

Results

RQ 1: To what extent is the GR Model implemented at GR schools?

GR Model implementation was fairly strong at three of the four GR schools, including Oak, Pine, and Elm, with overall school ratings of 3 to 4, and comparatively weaker at Fir, which was rated 2 to 3. As seen in Table 3, ratings for each of the five GR elements at GR schools varied somewhat between and within schools. In addition, ratings for the *settings* element, based on direct observations of GLT meetings at GR schools, indicated that K-3 teachers tended to implement GR procedures more effectively than their colleagues in Grades 4 and 5.

RQ 2: To what extent does the GR Model establish processes observable at GR schools but not at comparison schools?

Finding 1. GR schools utilised processes associated with the five elements of the GR Model more frequently and more effectively than comparison schools. Ratings for the five change elements at GR schools, except Fir, exceeded ratings for the comparison

Table 3. Frequency distribution of episode ratings for GR elements at GR and comparison schools (ratings based on GR Model Rubric levels)

School	Element	Rating						
		1	1 to 2	2	2 to 3	3	3 to 4	4
Oak (GR)	Goals	1	2	9	5	11	4	15
	Indicators	0	2	5	0	9	33	20
	Assistance	1	0	17	21	14	100	101
	Leadership	4	11	2	18	16	45	60
	Settings	0	1	0	6	7	16	6
Pine (GR)	Goals	0	0	4	8	12	56	22
	Indicators	2	4	8	9	32	29	13
	Assistance	0	3	7	5	26	61	18
	Leadership	0	12	7	22	27	56	41
	Settings	0	0	0	3	4	14	9
Elm (GR)	Goals	1	0	3	5	8	18	16
	Indicators	2	7	2	4	1	11	16
	Assistance	0	2	4	6	21	39	24
	Leadership	15	24	6	13	17	5	12
	Settings	0	2	0	10	6	20	3
Fir (GR)	Goals	2	3	12	21	16	0	2
	Indicators	0	7	12	14	15	1	3
	Assistance	0	31	12	31	15	21	1
	Leadership	13	39	28	33	16	8	0
	Settings	1	2	3	14	7	3	0
Kam (Comparison)	Goals	0	9	4	26	17	15	0
	Indicators	1	6	4	11	14	21	0
	Assistance	2	3	11	23	24	7	1
	Leadership	1	14	6	1	18	32	5
	Settings	0	11	2	13	5	2	0
Hall (Comparison)	Goals	0	30	17	10	9	1	0
	Indicators	12	4	5	10	10	4	0
	Assistance	8	9	11	30	19	0	0
	Leadership	24	19	5	2	24	17	3
	Settings	0	19	6	8	0	0	0
Lot (Comparison)	Goals	1	16	5	3	11	6	0
	Indicators	9	13	1	7	10	8	1
	Assistance	0	6	13	10	11	0	1
	Leadership	1	22	0	8	11	17	7
	Settings	1	2	2	13	3	1	0

Notes: Bold numbers indicate the element rating based on mean and median values for distribution of episode ratings for that element. One bold cell per row indicates the evaluator applied a *single* rating for the element based on similar mean and median values for that distribution of episode ratings. Two bold cells per row indicate the evaluator applied a *dual* rating for the element based on dissimilar mean and median values for that distribution of episode ratings

schools (see Table 3). Likewise, the overall school rating for each GR school (3 to 4) except Fir (2 to 3) exceeded that of comparison schools Kam and Lot (2 to 3) and Hall (2). With few exceptions, GLTs at GR schools used more effective processes to conduct meetings than GLTs at comparison schools. Comparative ratings for the settings element of the GR Model indicated that teachers at GR schools accomplished important instructional tasks more frequently and more effectively, during their GLT meetings, than teachers at comparisons schools (see Table 4).

Qualitative analyses indicated that teachers' behaviours and discourse, as well as meeting procedures and outcomes, differed appreciably during most team meetings at GR versus comparison schools (see Table 5). Notable examples, all suggesting more positive environments existed in GR team meetings, included: (a) more consistent focus, planning, and time for academic topics, goals, and indicators and less time discussing topics or doing tasks of a non-academic nature, such as duplicating or collating materials, and planning field trips; (b) analysis of students' products above and beyond state or district-mandated assessments (e.g., BME writing assessments at GR schools versus district-mandated, once-per-year, writing assessments at comparison schools); (c) discussing the relation between instruction, student outcomes, and the need for instructional changes; (d) modeling instructional methods for colleagues; (e) assigning and completing goal-related assignments, and using academic data with follow-up at subsequent meetings; (f) preparing and evaluating mutually agreed upon teaching strategies; (g) teachers' consistent versus sporadic attendance and participation at the meetings; (h) teachers' punctual arrivals and departures versus late arrivals and early departures; (i) principals' participation versus non-attendance at meetings; (j) teachers' use of typed agendas and prior awareness of meeting topics versus no agenda and limited, or last-minute, awareness

Table 4. Ratings for the setting element for grade-level team meetings at GR and comparison schools

		Rubric rating						
		<i>1</i>	<i>1 to 2</i>	<i>2</i>	<i>2 to 3</i>	<i>3</i>	<i>3 to 4</i>	<i>4</i>
GR schools					FFFFFFF			
	F	FF	FFF	FFPPPEE	PPPOO	EEEEEE	PPPOOO	
Comparison schools	(1)	(2)	(3)	EOOOO	OO	(6)	(6)	
		HHHHH		HHKKK				
		HHHHH	HKL	KKKKL	KKL			
		HKKLL		LLLLL				
				L				
		(15)	(3)	(17)	(3)	(0)	(0)	

Note: Each letter represents one episode rating, for one school, for one descriptive criterion from the settings element of the GR Model Rubric. Numbers in parentheses indicate the total number of settings ratings assigned at rubric level. Schools: E = Elm, F = Fir, H = Hall, K = Kam, L = Lot, O = Oak, P = Pine.

Table 5. Grade-level team meetings at GR schools versus comparison schools

GR schools	Comparison schools
<p>TC initiates/tells other teachers to pull out Open Court reading book from the student desks at which they are sitting, open to story, "Make Way for Ducklings". Page 176 is first page of story. In describing to other teachers how she delivers instruction on the skill of new paragraphs, TC asks the teachers, and says that she asks/teaches her students to identify why did the author make a new paragraph ... what is the reason for the change ... what changed? TC has the teachers read silently, approximately the first five paragraphs and asks them to identify WHY the paragraph changed. Various reasons emerge: change in locations, change in time, change in speaker. TC makes that point that we need to instruct explicitly kids about paragraph changes and reasons why paragraphs change, via examples. She also makes the point that we need to be careful about what stories/materials we select to use for this instruction because of poor proofing and incorrect, omitted paragraphing in some books.</p> <p>L teacher writes a post-it note to self. [I think that this note, along with one or two others, is his way of recording personally useful information for himself, such as the information that TC just described about teaching paragraphing, reasons for paragraph changes, related to the writing skill of indenting] ... TC states [in a summing/confirmatory manner]: "So I will make copies of ... the run-on sentences unit for you [other teachers] and you guys make ..." (<i>Elm, grade-level team meeting</i>)</p>	<p>D off-task diad discussion with F next to her re: I need to go to the office. Response: for what? To drop off money ... while M and coach discuss issue M raised. That is, M is getting at what drives instruction. Coach says trust the framework (standards). Coach reiterates that the District is developing a pacing/instructional plan for math for the coming year, thus, next year will be easier. OC: I observe that D is sitting back, looks at watch. Coach: So does everybody kind of know what they need to work on ... Responses from teachers are on the minimal side, some yes/nods. Coach asks: So what do you ... (i.e., the coach attempt to get teachers, particularly D to vocalise some of the specific standard sub-areas/skills that they highlighted, which they will prioritize/instruct over the next few weeks. [OC: The coach presses for more specificity in this meeting than in prior upper GL meeting. Good move, but ...] D: Just review everything ... I know they'll get it [again, a non-specific response]. N says that it's comforting to know that the framework/pacing plan is being developed in time for next year; that this year is a loss and I have good, smart kids. It's really hard because of the span ... N and a couple of others they should know multiplication tables by Grade 3. D concurs and says if they don't they should be kicked out of school. D stands, saying "I've got to stand ..." Coach and one teacher talk about an acronym. D to coach, I've got to go drop off \$ to office, leaves. (<i>Kam, grade-level team meeting</i>)</p>

Table 5. (Continued)

GR schools	Comparison schools
<p>ALT T redirects the group back to task of trying to determine what portions of chapters to keep/drop ... Principal chimes in (in regard to possible ways to teach math skills, particularly other than formal scheduled math time): "... if you can [take time to teach a particular math skill] do at 8:20 [A.M.], but everyone has to do it". MT refocuses the group by summarising: So, have we decided that we will do ... ALT T begins to model a teaching activity/procedure she uses with her students (standing at her math bulletin board, coins activity). Principal chimes in: last week when I was in ___'s class, she had kids make up different ways of making 26 ... ALT T [focusing group by concluding]: "... so can we say we're going to do time and money ... one week ... that leaves ___ weeks for remaining ... (<i>Oak, grade-level team meeting</i>)</p> <p>Leader distributes "Writing Script" that grade level uses. We all use the "Bernstein Bears". She asks if it's OK if we use it again ... it's what we've used in the past. Leader reviews steps that teachers will be doing leading up to assessment (leader and others are looking at "Writing Script" which specifies these steps). In discussing how to administer the assessment, group clarifies what to do/what not to do and gets to idea that this is an initial assessment. Leader distributes "Writing Rubric" ... E describes an example from another year/teacher. OC: E seems to provide very solid examples.</p>	<p>F says that next week's collaborative planning meeting will have something guided, but that this week we're "mostly on our own". F mentions current/upcoming task that teachers need to do, indicating that this time can be used for these tasks ... Open House, Report Cards, [Implementation? and/or ELD?], cumulative folders. I ask, so you have report cards to do? F: Not us (not F and M who are on same track), but teachers on other track. M and F looking at student paintings ... discuss/ recommend use of tempura paint versus/or watercolours. The two teachers are working independently on student cumulative folders, writing information and checking information in cumulative folders ... they chat about M's wife and kids ... babysitting arrangements. (<i>Hall, grade-level team meeting</i>)</p> <p>I notice that during the initial portion of the meeting that Mrs. Z is preparing some type of materials/cutting construction paper [i.e., unrelated to meeting topic/intended tasks]. (<i>Kam, grade-level team meeting</i>)</p>

Table 5. (Continued)

GR schools	Comparison schools
Discussion of what “detail” means in this grade level’s writing. Leader reviewing “Writing Rubric”. Leader discusses upcoming meeting and that, among other activities, we’ll score writing assessments that day. Discussion of scheduling items. 2:10 Principal departs. ... Leader shifts to OCR, which is agenda item 2. Any concerns? Male asks for clarification about some writing assessment procedures time = 30 minutes; will model still be up = no; but web can stay up. E shows materials that she uses with her students; other teachers view; favourable reactions. 2:17 Meeting ends. (Pine, grade-level team meeting)	

Table 5. (Continued)

GR schools	Comparison schools
N (ALT rep for this grade level) reviews last meeting:	<p>NOTE: Initially, I don't see or hear much active responses from teachers; they're mostly very quiet. The facilitator does utilise the approach/appeal in attempting to get to get teachers to provide input/decisions, that she doesn't know their students, you know your students and what they should/should not be culled. For the first few week-by-week attempts to get teacher input, I almost have that "pulling teeth" feeling and I wonder if part of the issue is whether some of the teachers are implementing any of Move It Math, or doing so on a limited basis. At one point in the meeting/before or after I hear green sweater teacher say in response to either O's (district/ national coach) or A's (school-based coach) query, "Are you on a set week?" that "yes" but subsequently "We weren't really doing as much math as we're supposed to ..." (due to, for example, not having training). O later tells me when I am alone with her that they had purposefully done three trainings in the summer, one per month, so that they could hit all tracks' teachers. I also saw a fax dated 11-08 from O to Lot/A that listed items needed for this meeting, including list/number of teachers who had not gone to/ received the Move It Math training. 11:08 One student enters room, hands paper to one teacher. The one teacher departs room for about 30 seconds. About one minute later, another teacher departs room for about 15 seconds. 30 seconds later, a student opens door and calls one teacher's name. The teacher says, "Go find Mrs. Gomez" and repeats that phrase. 11:12 O: "Do you feel comfortable ...?" (skipping) C: Regarding Monster addition: "You're not asking them to master it. You're just introducing it." A teacher had commented that this was a touchy one, that Monster math is difficult for the kids. B teacher asks: "How important is Scott-Foresman?" B teacher makes point that after doing manipulatives and ... that kids are tired of it and therefore is it necessary to do whole Move it Math page/all problems. R&B teacher: "In your school, you were managing two math problems as well?" "Because I am wondering ..." O: No, it just happened.</p>

Table 5. (Continued)

GR schools	Comparison schools
<ul style="list-style-type: none"> • Daily oral exercises ... Coach will provide • Two teachers shared strategies for sequencing • Identified new need ... new objective ... • N shifts to what this meeting will address. He places OH • Need = increase student knowledge of sequencing when writing story • Objective = By end of school year 75% of students WBAT to write 4 or more sentences to describe beginning, middle, end. <p>N leads teachers to today's task of analysing one low, one medium, one high student work samples to determine strengths and weaknesses. OC: It appears that each teacher has selected one high, one low, and one middle student summary from their own class. Good sign in terms of people actually brought the samples as agreed upon/directed in prior meeting/communications.</p> <p>N shows on OH a "1" example of student written summary of the <i>Three Little Pigs</i>. Shows a "2" example. N shows a high ("3") R asks Ts what are some of the strengths they see in the samples. Ts identify about 7 different specific strengths with multiple teachers responding, one at a time ... OC: More good signs = multiple teachers participating and teachers utilise common/efficient courtesies of speaking one person at a time. Some of the strengths that teachers vocalise-and N writes these on the OH ... (<i>Elm, grade-level team meeting</i>)</p>	<p>Sparks discussion, LB teacher says ... "feel frustrated"</p> <p>O: As Project Grad school, you're supposed to follow Move it Math and use Scott-Foresman as supplement. We don't really care if you use Scott-Foresman or ... or ... as long as you're addressing the Move it Math focus (i.e., having all your components, parts of lesson including warm-up, etc.) (<i>Lot, grade-level team meeting</i>)</p>

of meeting topics; (k) scheduled, weekly, “hands-off” (i.e., protected from competing demands and conducted as scheduled) versus “more loosely” scheduled meetings that were frequently cancelled, curtailed, rescheduled, changed at the last minute, or otherwise disrupted.

Staff at GR and comparison schools reported that the *Content Standards for California Public Schools* (California State Department of Education, n.d.) constituted their school-wide academic goals. They also stated that improving students’ scores on annual standardised tests was a school-wide goal. However, only GLTs at GR schools actually formulated, wrote, disseminated, and evaluated “more specific” academic goals based on the *Standards*. Some of these GLTs also focused their instruction on these specific goals, in a “planful” manner, by utilising a systematic process they acquired via GR training (see Table 6). GLTs at comparison schools did not have or execute such systematic processes during their meetings.

Finding 2. The GR Model established tighter linkages between teachers and administrators in their efforts to focus on academic goals and improve students’ academic achievement. First, teachers at GR schools were much more visible and regular participants in their schools’ academic leadership teams (termed ALTs at GR schools). Each of the four GR schools included at least one teacher from each grade level on their school-wide ALT. Of the three comparison schools, Hall had zero teachers on its leadership team, Lot did not have a formal leadership team, and Kam sometimes included one or two teachers on its team on an as-needed basis. Second, principals at GR schools attended and participated more consistently at GLT meetings and teachers’ professional development sessions than principals at comparison schools. Principals at most GR schools demonstrated greater awareness, focus, and participation in the day-to-day academic plans and actions of teachers at each grade level. The tighter academic linkages between teachers and administrators at GR schools facilitated more effective execution of goal-directed plans than at comparison schools, where the evaluator observed more frequent “slippage” between intended actions and actual implementation of academic initiatives.

RQ 3: What does implementation of the GR Model do to impact academic achievement?

Finding 1. GR Model implementation impacted students’ academic achievement by developing settings and processes whereby educators’ behaviours and instructional processes became more focused and produced visible improvements in students’ academic achievement and attainment of academic goals. teachers’ attributions for academic gains, teachers’ attitudes toward purposeful instructional tasks, teachers’ instructional efficacy, and teachers’ expectations for themselves and for students’ academic achievement changed when, and to the extent that, teachers experienced—frequently, punctually, and directly—visible improvements in academic achievement associated with their “results-producing” behaviours and instructional processes.

Table 6. GR emphasis on systematic planning and instruction (Focusing on and Addressing Common Student Needs in Grade Level Meetings: The 7 Important Tasks/Steps ... in a Nutshell)

Task	Goal	Things to remember
1	Identify and clarify specific and common student needs to work on together	<ul style="list-style-type: none"> Options for identifying needs: Standards, Stanford 9 content clusters, Beginning or Midyear assessment results, Curriculum (reading or mathematics series) <ul style="list-style-type: none"> If you have BMEs, use those to identify needs Connect/clarify needs using standards Clarify needs by analysing student work Write a brief description of the need Review the components of clear objectives: <ul style="list-style-type: none"> What do you want students to do, under what circumstances, and with what, if any, kind of support or set up? <p>How many students do you expect to meet the objective by when, and what do you want to see to evidence success or mastery (desired qualities)?</p> <ul style="list-style-type: none"> Follow an example when writing an objective: <p><i>Sample clarified need:</i> For Grade 1, writing fluency is being able to think about a topic, find things to say about the topic, and effectively use letters to write what you want to say, such that someone else can read and understand it.</p> <p><i>Sample objective:</i> Given familiar topics, by the end of January, most Grade 1 students (say, 75%) will be able to consistently produce journal entries during daily journaling time of four or more related sentences that stay on topic, show reasonable thought, and are readable.</p>
2	Formulate a clear objective for each common need and identify related student work to be analysed	

Table 6. (Continued)

Task	Goal	Things to remember
3	Identify and adopt a promising instructional focus to address each common need	<ul style="list-style-type: none">• Consider first what you already have• Consider something about which someone in the group has expertise• Consider something about which you can get published information• Consider something about which you can secure outside assistance• Use these questions to gauge the promise of the instructional focus:<ul style="list-style-type: none">- Can you generally explain what it is, and how you do it in the classroom?- Can you fit it into your existing programme?- Can you generally explain why it will likely help address the objective (assuming your team works hard to implement it well)?- Is it something that a reasonable number of folks on your team will try?- Can you secure the resources necessary to implement it as a team?• Use worksheet to record tasks 1, 2, 3 (page 39 of GR Manual 2001)• Planning and preparation are ongoing and cyclical• Inventory the necessary planning and preparation and secure commitments• Get concrete—develop actual written plans and needed materials• Get concrete—do mock or real demonstrations so people can observe and discuss THE TEACHING
4	Plan and complete necessary preparation to try the instructional focus in the classroom	

Table 6. (Continued 2)

Task	Goal	Things to remember
5	Deliver Instruction: Make consistent and genuine efforts to try the team's instructional focus in the classroom	<ul style="list-style-type: none"> • The challenge: this step is not taken during a meeting but in each teacher's class • Try to get a public commitment during meetings from each team member to try the instructional focus in the classroom • Identify individuals who will bring student work back to the subsequent meeting or will be prepared to report on their classroom efforts • Remind, remind, remind (i.e., "Don't forget we all agreed this week to try ...") • With regard to getting people to bring student work: specify ahead of time what specific work to bring, from which teachers and for which students • When analysing student work: use overheads, focus on the objective, and identify strengths and weaknesses in the work; then connect the strengths to what worked in the instruction and the weaknesses to what needs to be addressed in subsequent lessons • Use a chart like the one on page 37 of GR Manual, 2001 to guide the analysis • Central question: Is the objective being met? • Also: Are team members willing to sustain efforts on this need and objective? • Guideline: If a majority of the members want to stay the course, then stay the course
6	Analyse student work to: (a) see whether the objective is being met, (b) better understand the need, and (c) evaluate instruction	
7	Reassess: continue and repeat cycle or move on to another area of need?	

Finding 2. GR implementation shifted some educators' attributions for student achievement toward specific, teacher-implemented, instructional actions and planning processes, and away from teacher and student traits, and non-instructional explanations (see Table 7). Teachers in schools and GLTs where GR implementation was strong frequently attributed student achievement, or lack thereof, to specific instructional actions or processes they had or had not instituted. Teachers in schools and GLTs where GR implementation was weak were more likely to attribute student achievement to global factors or student traits, such as experience and knowledge, socio-economic conditions, inexperience with the English language, academic inability, lack of readiness, and inadequate parental involvement.

For some educators, GR's emphasis on systematic instruction cycles (Table 6) altered their instructional habits and beliefs about what constituted good teaching. For other educators, GR implementation affirmed these habits and beliefs. In both instances, educators clarified and raised their expectations for performing essential tasks needed to promote student achievement. In educational contexts fraught with competing and shifting demands, GR implementation helped many educators, individually and collectively, to prioritise goals and focus instructional efforts accordingly. Some teachers shifted from strict adherence to timeline-driven or page-sequenced coverage of materials based on actual mandates or perceived pressures. They recognised increasingly the implications of purposeful planning and "teaching less, better" instead of "covering" material. Similarly, GR implementation crystallised, for most principals, the importance of focusing on academic goals and essential tasks that most directly impact student achievement. GR "reminded" principals not to get distracted by the "operations side" of their job, or by the gauntlet of numerous, emerging, competing demands.

Finding 3. GR implementation fostered a group ethos among some teams of grade-level teachers and ALT representatives—a collective willingness and commitment to formulate, adapt, implement, and evaluate instructional processes that targeted student achievement. This group ethos was revealed during many, but not all, GLT and ALT meetings via participants' focused academic discourse, systematic planning, purposeful use of assessment data, and agreements to implement and evaluate goal-directed teaching strategies. By providing adequate time and assistance (e.g., substitute teachers and class coverage) for teachers to analyse student work and evaluate assessment data with their peers, GR implementation altered teachers' attitudes toward such tasks. Teachers increased and improved their analysis of assessment data to better inform important instructional decisions that impacted student achievement. Viewing and interpreting information—especially student work products and test results that provided teachers with timely evidence of academic improvements—increased teachers' instructional efficacy and promoted attributions that their instructional decisions and actions improved student achievement. Data indicated that GR Model implementation: (a) increased teacher's willingness to share assessment data with colleagues; (b) altered teachers' understanding and

Table 7. Comparison of teachers' attributions for student achievement

Stronger GR Model implementation associated with teacher beliefs that instruction accounted for student achievement	Weaker/non-implementation of GR Model associated with teacher beliefs that factors other than instruction accounted for student achievement
<p>A ... most of the grade levels were doing demos [right]. Teachers would volunteer if something was working well to demo it. So, part of these grade level meetings started becoming demonstration time-doing little mini-lessons...We shared different lessons ... someone would demonstrate a lesson--whatever's working in the classroom. So, I think it was also very helpful for us ... and then, when you find out what's working in one class, then you can try it in the other classroom, and so eventually everybody's doing the same thing because it works.</p> <p>A ... it goes back to something that I said earlier. Um, yeah, with whatever pressure and whatever downside, we are showing results. (<i>Transcript, Pme, teacher focus group</i>)</p> <p>2:23. T holds up some of the bar chart results school wide. He says that he did not make copies of this for group. He mentions that (students' performance on) high frequency words actually dropped from beginning to mid-year assessment, especially in Grades 4 and 5.</p>	<p>Comment about academic achievement and we know about the students' academic achievement (i.e., how students are doing) ... statements about loss of control, "we're professionals", don't need to have people telling us what to do, we know, we have degrees, master's degrees ... (<i>Field notes, Fir, grade-level team meeting</i>)</p> <p>[Teacher] ... Says these kids aren't ready for what we're asking them to do ... And I have eight students who don't speak English, don't know anything about writing ... they don't get it. [Principal asks teacher] In general or just writing? Teacher: In general. (<i>Field notes, Fir, grade-level team meeting</i>)</p> <p>Coach distributes first one-page document with standardised test results from last year. Teachers note that 3rd Grade scores from last year were high and relate that this year's group (i.e., who are now their Grade 4 students) should do well on test this year. Male T states reason why the scores were good/his students are doing well in math is because most of the/his students are male. He also states that that's why the reading scores are low; girls do better with reading, males do better with math ... (<i>Field notes, Kam, grade-level team meeting</i>)</p>

Table 7. (Continued)

Stronger GR Model implementation associated with teacher beliefs that instruction accounted for student achievement	Weaker/non-implementation of GR Model associated with teacher beliefs that factors other than instruction accounted for student achievement
<p>R asks: Are people doing word wall everyday?</p> <p>OC: R seems to provide on-target questions/ comments that focus on analysis of the issues. (<i>Field notes, Oak, grade-level team meeting</i>)</p> <p>T says nobody is going to hand you a blueprint ... it has to come from us ... one reason why we're doing so much better than 1995. It's a huge change since 1995; it's not just personalities, it's how we view the students' work/ work. (<i>Field notes, Oak, academic leadership team meeting</i>)</p>	<p>B comments about when kids are/are not ready</p> <p>"developmentally" for writing tasks ... references to their (lack of) English reading/English language proficiency ... B makes reference to Open Court scores ... B says we have a goal/focus on vocabulary. Then discussion of comprehension scores being not too good. T says she thinks it was the story ... (<i>Field notes, Fir, grade-level team meeting</i>)</p>

expectations about the purposes of assessment data; (c) fostered an “improvement over time” versus a “one-shot” orientation for collecting, analysing, and using data; and (d) shifted grade-level teams from talking about indicators to systematic analysis and actual use of indicators to plan and implement instruction. The following example illustrates aspects of this finding. A teacher at Pine explained how GR-initiated, BME writing assessments changed the GLT’s use of assessment tools.

Our grade-level meetings, which were supposed to be geared toward improving scores, really had no relation to what tests we were giving. But now this year, since we’re able to choose a goal to pick with the writing assessment ... Those [assessments] have become a valuable tool, whereas previously, I was just given them and that was it. They had no value to me at all. (Teacher focus group)

Finding 4. GR implementation, through collaborative goal setting, analysis of indicators, and reflection on teacher-controlled instructional variables, impacted teachers’ expectations for student achievement. Some teachers’ raised their expectations for student achievement after they viewed student work or analysed data that provided explicit evidence of students’ academic improvement. However, most teachers’ expectations did not appear to rise dramatically. Modest expectations for student achievement were reflected in teachers’ initial specification and subsequent revision of academic goals, objectives, and criteria. Document retrieval indicated that GLTs frequently set a criterion of “75% of the students” when formulating academic goals. More notably, the practice of collaborative goal setting brought to the surface teacher expectations for student achievement, and prompted some teachers to examine implications of their individual assumptions and collective expectations, as reflected in the following exchange between teachers at Oak Elementary:

- T1: I give them a piece of writing almost every night, or some kind of poem.
T2: And you’re expecting them now to be able to look at a rubric and get a three or a four?
T3: I see an improvement, you know, big time ... But I think some kids will never have the ability to be a four in my mind.
T1: At the fifth-grade level?
T3: Yeah, because, um, you know, whatever factors ... there are; their families or, I have a girl who’s only been here, you know, for a few months ... let’s say from Argentina ... you know, that’s just the way it is.
T1: But I think kids that have been here ... that have had exposure to this writing process, exposure, it’s only going to help them.
T2: ... I think the first graders now, let’s say in five years when they come to fifth grade, hopefully they’ll be, uh, able to achieve better because they would have been exposed to the process.
T1: Yeah. You know, I’m a little dismayed to hear teachers say that they can never achieve that level.

Finding 5. GR implementation required teachers to assume academic leadership roles and to chart the academic course and outcomes of their schools. As teachers

performed leadership tasks that impacted their colleagues and promoted academic achievement, this altered teachers' professional responsibility, instructional efficacy, and collegial accountability, particularly among teachers who served as ALT representatives for their GLTs. This finding is supported strongly via stark contrasts between what constituted a GLT meeting before and after GR implementation, and at GR versus comparison schools (see Table 5). Data collection identified many examples associated with the transformation of GLT meetings from "chat sessions" to settings where teachers took responsibility to meet collaboratively and focus on improving academic achievement. Germane examples appeared in ALT meetings, too. These examples reflected teachers' attitudes toward performing important tasks that impacted student achievement, including sharing expertise with colleagues, supporting and pressuring colleagues, preparing instruction, utilising meeting time, and completing professional tasks.

A heightened sense of professional responsibility was illustrated vividly in two cases in which staff confronted very challenging issues. In the first case, GR and school staff formed and implemented a strategy designed to get "reluctant" teachers to participate more actively in GLT efforts to improve students' writing. In the second case, GR and school staff formed and implemented an intervention designed to assist a "struggling" upper-grade GLT. In both cases, ongoing external assistance from GR staff was paired with internal expertise and leadership of school staff. This melding of assistance and leadership triggered a sense of urgency, agency, and responsibility among staff (Earl & Lee, 1998). It enabled staff to confront problematic issues that had inhibited necessary changes in instruction.

RQ 4: What helps and hinders schools' implementation of the GR Model?

Five factors facilitated or inhibited GR Model implementation. The first factor that affected GR Model implementation was combined leadership of school administrators and teachers who represented their GLT on school-wide ALTs. Administrators and ALT representatives at Oak and Pine focused more consistently on academic goals than their colleagues at Elm and Fir, as evidenced in their respective actions, participation, and discourse during GLT and ALT meetings. Administrators at Oak and Pine attended GLT meetings more frequently than their colleagues at Pine and Fir. Administrators at Elm were distracted frequently by the day-to-day operational demands of their jobs. Administrators at Fir and Elm more frequently used and responded to "walkie-talkie" communications during GLT and ALT meetings. These off-topic communications interrupted the flow and academic focus of the meetings. Elm's Principal reported being "overwhelmed" by the combined academic and operational demands of the job. Some ALT representatives (i.e., teachers who were GLT leaders) at Elm tended to "pick up" some of the leadership roles and responsibilities for implementing GR activities. The principal at Fir was not as distracted as the principal at Elm, but the leadership provided by ALT representatives at Fir was the least effective of all GR schools. Consequently, the overall rating

for the leadership element at Fir was 2—slightly lower than Elm’s rating of 3, and lower still than the rating of 3 to 4 for both Oak and Pine.

The second factor that impacted GR implementation was the relative frequency and continuity of expert assistance provided by GR staff. Direct observations of GR staff’s assistance confirmed school staffs’ repeated assertions that GR staff members provided essential expertise and leadership. GR assistance was strong over time at Elm, Pine, and Oak. At Fir, a GR staff member provided frequent assistance to the principal and staff during Year 1. However, that assistance was provided much less frequently during subsequent years. The principal and some teachers at Fir reported that they “missed” this external assistance. Ratings and direct observations of GLT and ALT meetings at Fir confirmed this finding (see Table 4).

A third factor that affected GR implementation was school staffs’ readiness for change, which varied by individual teachers and GLTs. Most nascent and veteran teachers were optimistic about school change initiatives and programmes. They invested energy to implement GR Model activities and other initiatives. However, a few veteran teachers were pessimistic. They limited their investment and participation based on past experiences with short-lived initiatives and programmes that “come and go”. In addition, most staff at GR schools were relatively unencumbered by “baggage” or past experiences that negatively impacted current relationships and job performance. However, lingering issues, differing expectations, and idiosyncratic habits among some GLTs and individuals diverted valuable energy and time from school change efforts.

A fourth factor that impacted GR implementation was concurrent job demands. As stated previously, K-3 GLTs implemented GR procedures more effectively than their colleagues in Grades 4 and 5. Unlike their K-3 colleagues who had implemented a new structured reading programme the preceding year, teachers in the upper grades, at three of four GR schools, were implementing those same programmes for the first time. Upper-grade teachers reported consistently that they expended extraordinary time implementing this new reading programme. First-year implementation demands, combined with concurrent first-year implementation of a new mathematics programme, proliferation of assessments, and difficulties resolving pacing plans, accounted, in part, for greater stress and less effective GR implementation among upper versus lower GLTs.

A fifth factor that impacted GR implementation was the type of operating schedule used at each school. Year-round multi-track scheduling—a response to overcrowding and shortages of school buildings in this dense urban district—clearly complicated communication, planning, and scheduling at Elm, Fir, and Oak (and comparison schools), but not at Pine, which used a traditional, 9-month single-track schedule. Staff at Elm, Oak, and Fir expended considerable time during GLT meetings, ALT meetings, faculty meetings, and professional development sessions trying to determine how teachers who were “off-track” could be informed, accommodated, or trained on GR and related activities. Teachers and administrators consistently expressed concerns about the “hand-off” and continuity of initiatives between incoming and outgoing tracks. Multi-track scheduling made “teacher buy-in” more

challenging because one-third of the staff and their students were not present; they were on their off-track vacation period. Consequently, only two-thirds of the staff were present, at any given time, when important decisions were made, such as when GLTs formulated academic goals and agreed to use corresponding instructional techniques.

RQ 5: What has changed as a result of GR Model implementation from the perspectives of participants and the external evaluator?

According to participants and the external evaluator, GR implementation increased educators' focus on academic learning and on getting academic results in four important ways (see Table 8). First, GR transformed GLT meetings, from settings where non-academic topics and activities consumed time into settings where systematic academic planning, instructional modeling, analysis of student work, and other purposeful, goal-directed tasks informed teachers' instruction. Second, GR assisted and challenged teachers to provide leadership in charting the academic course of the school through participation in newly created ALTs. Third, GR staff assisted and challenged administrators and staff to prioritise and emphasise purposeful, goal-directed efforts aimed at improving teaching and learning. Fourth, GR fostered systematic collection and use of assessment data, particularly BME writing assessments, in ways that more punctually and more effectively informed teachers' instructional decisions.

Discussion

In this final section, we discuss limitations of this external evaluation study, as well as implications of this study's findings for school reform in general, and more specifically for understanding what goes on inside the black box of reform.

Limitations of the Evaluation Study

Important limitations of this study relate to the case-study approach used in this process-focused evaluation. Case studies, inherently, must be bounded (Merriam, 1998). We purposefully selected seven cases—four GR schools and three demographically similar comparison schools. Thus, our findings, discussion, and implications must be placed into context. One obvious limitation is that our findings emanate strictly from primary schools. We did not and have not investigated how GR processes might operate in middle schools or high schools. So, we are left to wonder if and how GR processes might apply to such schools. Some researchers have reported that educational reform processes, and even evaluating such processes, are more challenging in middle schools and high schools compared with primary schools (Greene & Lee, 2006).

Another major limitation of this study is that the external evaluator conducted relatively few observations of teachers instructing students in their classrooms.

Table 8. Participants' perspectives on changes resulting from GR implementation

<p>Grade-level team meetings transformed</p> <p>"We're using our grade-level meetings to focus on instruction now...the grade-level meetings are very well planned...have agendas...grade-level meetings are no longer [used for] planning treats and assembly performances, unfortunately [some-one chuckles]. But it's good...for the students because it's not a lot of down time at these meetings...focus on [students' academic] needs ... most of the grade levels were doing demos. Teachers would volunteer if some- something was working well to demo it. So part of these grade-level meetings started becoming demonstration time—doing little lessons." (<i>Pine focus group</i>)</p> <p>... having meetings that focus on student achievement and improving it, versus faculty meetings that talk about copiers and things like that all of the time." (<i>Elm focus group</i>)</p>	<p>Administrators prioritised and focused</p> <p>[GR] "helped me focus" given all of the many things that draw principals away from instruction. She says, "I didn't know how to lead" in the past when she came on as a principal, but that Bruce/[GR] showed her how, the specifics of leading people and focusing them on instruction. She showed the ALT (how to lead and focus), then they lead and did likewise with their colleagues ... when she first changed from grade-level chairs format to ALT, that teachers still used the meetings as complaint sessions and to deal with the many non-instructional items ... managed to get them focused on instructional goals ... via pressure. (<i>Pine principal interview</i>)</p> <p>"And, Bruce has been, really, a godsend. He challenges me and he pushes me, and he the monthly meetings that we have with him here at school—the administrative team [ALT] here have been wonderful. Because he pushes you, and he challenges you, and he just supports you all the way. So, we—I don't think we would be as far as we are if it weren't for, for the Getting Results team." (<i>Oak principal interview</i>)</p>
<p>...when she took over as Principal GLT meeting time was something that teachers might or might not have attended, served as complaint time... getting teachers to attend was a first step & accomplishment. Getting them to start talking about instruction was/is another step. (<i>Fir Principal interview</i>)</p>	<p>Assessments used to inform instruction</p> <p>"No. That's never happened before. At the beginning, that was also a tough nut to crack because people didn't want to bring student—"I'm not going to bring my student work because then people will think I'm not a good teacher". ... it took a while to get away from that ... to change the mindset ...</p>
<p>Teachers provided leadership via ALT</p> <p>"I think that AALT, the team ... they had a lot of leadership this year. And I know they felt the pressure. And I think they were the leaders because they were the ones who were kept pushing us ... leading the meetings ... Yes and keeping us focused ..." (<i>Pine focus group</i>)</p>	

Table 8. (Continued)

<p>We started the Academic Achievement Team. And that's become our curricular leadership, uh, tool ... It's been a challenge to, um, bring teachers actively into a leadership role because, um, they're very uncomfortable telling their colleagues something they don't want to hear ... when it comes to something that they've come to a conclusion needs to be done that's not going to go down well, they still want the boss to tell them" [laughs]. (<i>Fir principal interview</i>)</p>	<p>Because teachers were very uncomfortable with sharing and bringing work in. Now it's not an issue anymore ... We've never sat down with colleagues to analyse student work ... And we never got together to talk. And we never got together to, to plan together. I mean, it's just these last two to three years has really changed things ... (<i>Oak principal interview</i>)</p>
	<p>"I remember my sub days were just to grade the assessments ... actually got me to grade the assessment ... really, really helped ... and discussed the results ... made notes of student needs on the spot, so then when we had our next meeting we ... set goals ... And not only assessing but looking at those ... helping guide our instruction where our weaknesses are ... where we can focus ... and how we can improve the weaknesses and continue with the strengths ... most useful is the writing assessment that we do at the beginning of the year, middle, like she said, and the end of the year. It helps us find out what needs, uh, we need to address with the students. And I think it really helps. [Right away.] [Yeah.] Whereas not, "Oh, school's ending and where are, what are we lacking". (<i>Pine focus group</i>)</p>

Frequent classroom observations of teacher–student interactions would have provided greater insights about linkages between GR Model elements, GR-related processes, and teaching and learning at GR schools. Likewise, observing teaching and learning, *in vivo*, in classrooms at comparison schools would have further illuminated our findings. Additionally, findings reported here are subject to the benefits and limitations of using a sole external evaluator, as opposed to an evaluation team.

Implications for School Reform

Findings indicated that although implementation levels varied somewhat across GR sites, the GR Model was successfully implemented at GR sites. Moreover, GR implementation established apparently unique processes that contributed to improved academic achievement at low-achieving schools. GR schools also were rated higher than comparison schools on key elements of school improvement. These findings have important implications for school reform.

First, findings from this external evaluation contribute to the body of literature that supports the efficacy of change elements in the GR Model—elements that are not unique to the GR Model. These elements are, indeed, evident in other reform designs (Fullan, 2000; Marzano & Kendall, 1996; Schmoker, 1996), although we would note the GR model stands in contrast to much more prescriptive reforms that specify curricula and instruction (Borman et al., 2003; Desimone, 2002) and that have shown the strongest evidence of impact. Nonetheless, the evaluation findings reported here and in Saunders (2003) support the premise that goals (or standards), indicators (or assessments), assistance (or collaboration and professional development), leadership, and settings are influential levers for change. We have not yet established empirically the individual contribution of each element to improved achievement. In one sense, the individual contribution of each element is irrelevant. The premise of the model is that no one element is sufficient; all are necessary. We do know that, in combination, and given reasonable development (i.e., Level 3 on the GR Model Rubric), these elements were associated with more rapid gains in academic achievement at GR schools, compared with the gains at demographically similar comparison schools, and the overall school district, over a 5-year period.

Second, data collected in this study indicated that GR change elements were, at least, somewhat evident at comparison schools. We believe that the challenge is not installing these elements, but developing and utilising them effectively and at a high level (e.g., a 3 or 4 on the GR Model Rubric). Recall that most GR schools were rated at rubric levels 3 and 3 to 4 for each element, and most comparison schools were rated at 2 and 2 to 3. If this school district reflects reform efforts in the U.S., then many schools, based solely on local, state, and national emphases (e.g., content standards, annual assessment, professional development, school governance), and without local assistance of an entity like the Getting Results Network, might function at Level 2 or slightly higher (see GR Model Rubric, Table 2). In other words, the following are likely to occur: (a) achievement goals are probably identified in the form of state or district content standards; (b) school-wide indicators, in the form of standardised

tests, are probably used, at a minimum, to examine annual achievement trends; (c) assistance or professional development is likely to be available, at least intermittently; (d) leadership is most probably focused on improving academic achievement, at least in general terms; and (e) settings, such as leadership teams, grade-level meetings, and faculty meetings, are available and used, to some degree, to foster school improvement efforts. The GR Model provides one example of how schools might move these change elements toward a higher level of development and utilisation. We would then expect this to result in achievement benefits beyond those currently observed in schools where standards, mandated tests, professional development, and other reform features have been mandated from the top down and with less systematic attention to change elements.

Third, findings from this study, like countless other school reform studies, point to the critical role of leadership. In particular, this study's findings seem to demonstrate the promise of increasingly distributed leadership that involves both administrators and teachers working together. Indeed, some prominent scholars have called for more creative forms of distributed leadership that not only improve student achievement, but also do so in socially just ways (Hargreaves & Fink, 2006), and in ways that transform the culture of learning and teaching in schools (Fullan, 2002). GR implementation required teachers to assume academic leadership roles (RQ 3, Finding 5), helped establish tighter linkages between administrators and teachers (RQ 2, Finding 2), and brought about change by challenging and assisting teachers to actually provide leadership to staff and GLTs (RQ 5, Finding 2). Not unexpectedly, some teachers in GR and comparison schools expressed concerns about their roles as change agents (e.g., worrying about the extent to which the focus on academic achievement in reading and mathematics impacted other aspects of schooling, including physical education and health, music, and art). Most prominent among those factors that helped schools successfully implement the GR Model was the combined leadership of school administrators and teachers via the ALT (RQ 4, Finding 1). We address, here, three implications of our findings for research on, and practice of, leadership as a lever for school change.

First, we conceptualise and focus on leadership as it relates to other elements of reform and school effectiveness; that is, not leadership per se, but leadership as it bears on setting and sharing goals or standards, utilising indicators or assessment information, seeking out and cultivating assistance and collaboration, and establishing and maintaining productive settings such as leadership teams, grade-level meetings, and faculty meetings. We could have written a very different version of this article, one that might have isolated our findings on leadership. But it seems more relevant and theoretically more important to examine and discuss leadership in the context of the full GR Model, wherein leadership functions as one of five highly interdependent change elements. At least as it is conceptualised in the GR Model and enacted at GR schools, leadership operates as a necessary although not sufficient condition for productive change.

Second, related to our first point, the findings of this study provide one example of leadership as it functions within the context of a specific, and fairly concrete,

approach to school change. Although the GR Model/Network is curriculum-free (unlike, e.g., Success for All) (Slavin & Madden, 2006) and process-oriented, it has advanced over the years to become increasingly “nuts-and-bolts” oriented in its approach to the process of change and to the administrator and teacher leadership that successful execution requires. For example, GR assistance and trainings for principals and teacher-leaders, for the most part, do not focus on leadership *per se*. We certainly highlight the concept of “leadership that supports and pressures”—one way to distill the large body of empirical and theoretical work on leadership. And we work with teachers to help them understand principles of keeping meetings focused and people on-task (e.g., by creating and adhering to agendas, and dealing with problematic personalities). But, more commonly, GR assistance and training for principals and teacher-leaders focus very directly on leading and facilitating specific and concrete instructional tasks. These tasks include, for example, using a specific protocol to review and interpret standards; applying detailed procedures to administer, score, tally, and analyse results of specific assessments at the BME of the school year; and using a specific protocol to identify common student needs, formulate objectives, and analyse student work in the context of grade-level meetings.

An ongoing debate in the reform literature weighs the comparative benefits and drawbacks of more conceptual versus more “nuts and bolts” approaches to reform (Bodilly, 1998). As the GR Network has evolved, we have consistently found it valuable to try to work on both planes with school staffs—on the concepts or principles underlying concrete procedures and the successful execution of those procedures. Without question, over time we have produced higher levels of implementation and effectiveness, and, we would argue, stronger and more effective leadership from both principals and teachers as the nuts and bolts of the work have been made increasingly clear. It is possible that our efforts have been myopic—that we have cultivated highly contextualised, task-specific leadership skills with limited generalisability. We have completed no studies to investigate this possibility. Perhaps this is one of many issues that merit study. Nonetheless, while noting this caution, GR results to date suggest that benefits do accrue when we approach leadership programs and interventions from a more contextualised, task-specific orientation. In other words, when we help educators in the schools get things done. Moreover, findings from recent multi-method studies of primary school reform are consistent with our experiences. For example, Lithwood, Jantzi, and McElheron-Hopkins (2006) found that school leadership and school improvement processes accounted for the largest proportion of variance in explaining modest yet significant differences in student achievement across primary schools.

Third, it seems important to note what the GR Model and its implementation does and does not focus on in terms of leadership. At the centre of the GR radar is teaching and learning—not school governance *per se*. Inherent in our initial theoretical orientation of school change, this heart-of-the-matter focus has been further reinforced as our work in schools evolved. Originally, we non-descriptly named our model the “School Change Model” because we saw it as a vehicle for making fundamental changes in how teachers approach the work of teaching and learning.

As the focus on student outcomes became increasingly pronounced, and the educational system as a whole bore down on tangible evidence of student achievement as an impetus of improvement (e.g., Schmoker, 1996), we started to use the name “Getting Results”. Throughout the years, we have tried consistently to stay as close as possible to the major source of improved achievement over which schools have enormous direct control—teaching and learning in the classroom (Wang, Haertel, & Walberg, 1993). Some GR schools, as well as non-GR comparison schools, have had governance training (e.g., LEARN, a local school-district reform effort that was popular in the 1980s and 1990s, and School-Based Management), and some have not. But we have not found focusing on governance issues independent of the issues that most directly impact teaching and learning in classrooms to be an effective approach to accomplishing our fundamental goal: improving student achievement.

We raise this issue of governance in light of the strong findings this study uncovered about distributed leadership. Perhaps teaching and learning, in general—and model elements such as academically focused goals, indicators, and assistance, more specifically—provide a fruitful common ground upon which shared or distributed leadership among administrators and teachers can be cultivated efficiently and successfully. This is not to negate the potential import of shared or distributed leadership that focuses on governance. However, governance has for many years received substantial attention and emphasis in the reform literature. GR results suggest substantial promise in shared leadership that focuses specifically on helping teachers get results in their classrooms.

In the GR project, strong external assistance provided by GR staff helped administrators and teachers become more effective leaders and experts—professionals who focused their efforts and who pressured, supported, and assisted colleagues toward achieving specific, measurable goals in ways that impacted teaching and learning on a daily basis. GR staff provided the expertise initially, to help implement key aspects of the operational model—ALTs, grade-level teams, collection and scoring of student papers—and subsequently, to help sustain these settings and foster expertise and leadership among key educators at GR schools (e.g., teachers who served as GLT leaders and ALT representatives, principals, academic coaches who worked across grade levels). In time, a critical mass of focused educators emerged at GR schools and teaching became more coherent across each school. One very consistent related theme was that GR staff provided assistance in ways that earned educators’ respect and impacted their attitudes and teaching practices. This assistance, which originally was perceived by educators to be from an outsider, morphed into a more internal form of assistance as educators began to interact with GR staff as “one of our own”.

Paths that lead to getting results are not likely to be smooth. Indeed, GR schools and comparison schools experienced many bumps in their respective paths, including a gauntlet of emerging demands that often impeded their way. External expertise, combined with leadership within the school, helped schools stay on track. Elements of change could wane—focus in the form of goals that are set and shared, indicators that measure progress toward achieving goals, and settings where the nuts

and bolts of change must occur in order to get results. Indeed, these elements seemed to be in danger of continual derailment since so much effort had to be expended to maintain them. The external assister played a crucial role in helping school staffs keep focused on their ultimate goals and maintain settings and activities that were essential for achieving them.

References

- American Institutes for Research. (1999). *An educators' guide to schoolwide reform*. Washington, DC: Author.
- August, D., & Hakuta, K. (Eds.). (1997). *Improving school for language minority children: A research agenda*. Washington, DC: National Academy Press.
- Berends, M., Kirby, S., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American schools: Three years into scale-up (MR-1145-EDU)*. Santa Monica, CA: Rand.
- Blase, J. (1987). Dimensions of effective school leadership: The teacher's perspective. *American Educational Research Journal*, 24, 589–610.
- Bliss, J., Firestone, W., & Richards, C. (Eds.). (1991). *Rethinking effective schools: Research and practice*. Englewood Cliffs, NJ: Prentice Hall.
- Bodilly, S. J. (1998). *Lessons from the New American Schools' scale-up phase: Prospects for bringing design to multiple schools*. Santa Monica, CA: RAND.
- Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73, 125–230.
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.
- California State Department of Education (n.d.). *Content standards for California public schools*. Retrieved March 4, 2002, from <http://www.cde.ca.gov/standards>.
- Carter, T., & Chatfield, M. (1986). Effective bilingual schools: Implications for policy and practice. *American Journal of Education*, 95, 200–232.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.
- Denzin, N. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72, 433–479.
- Earl, L. M., & Lee, L. E. (1998). *Evaluation of the Manitoba School Improvement Program*. International Centre for Educational Change at OISE/UT.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15–27.
- Fashola, O., Slavin, R., Calderón, M., & Durán, R. (1996). *Effective programs for Latino students in elementary and middle schools*. Unpublished manuscript. Baltimore, MA: The Johns Hopkins University.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London: Falmer.
- Fullan, M. (2000). The three stories of educational reform. *Phi Delta Kappan*, 81, 581–584.
- Fullan, M. (2002). The change leader. *Educational Leadership*, 59(8), 16–20.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NJ: Doubleday.
- Goldenberg, C. (2004). *Successful school change: Creating settings to improve teaching and learning*. New York: Teachers College Press.

- Goldenberg, C., & Gallimore, R. (1991). Changing teaching takes more than a one-shot workshop. *Educational Leadership*, 49(3), 69–72.
- Goldenberg, C., Saunders, W., & Gallimore, R. (1996). *Settings for change: A practical model linking rhetoric and action to improve achievement of diverse students*. Proposal submitted to the Spencer Foundation.
- Goldenberg, C., & Sullivan, J. (1994). *Making change happen in a language minority school: A search for coherence* (Educational Practice Report No. 13). Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.
- Greene, J. C., & Lee, J. H. (2006). Quieting educational reform ... with educational reform. *American Journal of Evaluation*, 27, 337–352.
- Harcourt Assessment. (1996). *Harcourt educational measurement*. Retrieved from <http://harcourt-assessment.com>.
- Hargreaves, A., & Fink, D. (2006). The ripple effect. *Educational Leadership*, 63(8), 16–20.
- Joyce, B., & Showers, B. (1983). *Power in staff development through research on training*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.
- Lieberman, A. (1988a). Expanding the leadership team. *Educational Leadership*, 45(5), 4–8.
- Lieberman, A. (1988b). Teachers and principals: Turf, tension, and new tasks. *Phi Delta Kappan*, 69, 648–653.
- Lincoln, E., & Guba, Y. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lithwood, K., Jantzi, D., & McElheron-Hopkins, C. (2006). The development and testing of a school improvement model. *School Effectiveness and School Improvement*, 17, 441–464.
- Little, J. (1982). Norms for collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal*, 19, 325–340.
- Manning, K. (1997). Authenticity in constructivist inquiry: Methodological considerations without prescription. *Qualitative Inquiry*, 3(1), 93–115.
- Marzano, R., & Kendall, J. (1996). *Designing standards-based districts, schools, and classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McDougall, D. (2002). *Getting Results project external evaluation*. Long Beach, CA: Getting Results Network.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Miles, M. (1983). Unraveling the mystery of institutionalization. *Educational Leadership*, 41(3), 14–19.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
- Peterson, K., & Lezotte, L. (1991). New directions in the effective schools movement. In J. Bliss, W. Firestone, & C. Richards (Eds.), *Rethinking effective schools* (pp. 128–137). Englewood Cliffs, NJ: Prentice Hall.
- Rosenholtz, S. (1991). *Teachers' workplace: The social organization of schools*. New York: Teachers College Press.
- Sarason, S. (1972). *The creation of settings and the future societies*. San Francisco, CA: Jossey-Bass.
- Saunders, W. (2003). *"Getting Results": Final report of achievement impact*. Long Beach, CA: California State University.
- Saunders, W., O'Brien, G., Marcelletti, D., Hassenstab, K., Saldivar, E., & Goldenberg, C. (2001). Getting the most out of school-based professional development. In P. Schmidt & P. Mosenthal (Eds.), *Reconceptualizing literacy in the new age of multiculturalism and pluralism* (pp. 289–320). Greenwich, CT: IAP.
- Schmoker, M. (1996). *Results: The key to continuous school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Slavin, R., & Fashola, O. (1998). *Show me the evidence! Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin.

- Slavin, R. E., & Madden, N. A. (2006). *Success for all/roots and wings: 2006 summary of research on achievement outcomes*. Baltimore, MD: John Hopkins University.
- Taylor, S. J., & Bogden, R. (1998). *Introduction to qualitative research methods*. New York: John Wiley.
- Tharp, R., & Gallimore, R. (1988). *Rousing minds to life*. Cambridge, England: Cambridge University Press.
- Wang, M., Haertel, G., & Walberg, H. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249–294.