

# **Transformational School Leadership for Large-Scale Reform: Effects on students, teachers, and their classroom practices**

Kenneth Leithwood\* and Doris Jantzi

*Ontario Institute for Studies in Education, University of Toronto, Ontario, Canada*

Using data from a larger 4-year evaluation of England's National Literacy and Numeracy Strategies, this study tested the effects of a school-specific model of transformational leadership on teachers (motivation, capacities, and work settings), their classroom practices, and gains in student achievement. Some 2,290 teachers from 655 primary schools responded to 2 forms of a survey (literacy and numeracy) measuring all variables in our framework. Our measure of student achievement was gains in the British government's own *Key Stage 2* tests over either 2 (numeracy) or 3 (literacy) years. Path analytic techniques were used to analyze the several different versions of the results. Results indicate significant effects of leadership on teachers' classroom practices but not on student achievement.

## **Introduction**

Policy-makers aiming to improve schools on a large scale invariably assume that the success with which their policies are implemented has much to do with the nature and quality of local leadership, especially leadership at the school level (Brown, Anfara, Hartman, Mahar, & Mills, 2002). They are likely correct. A compelling, although still modest body of empirical evidence now demonstrates the significant effects of such leadership on school conditions and students' learning (e.g., Hallinger & Heck, 1996a, 1996b, 1998; Leithwood & Jantzi, 2000; Waters, Marzano, & McNulty, 2003). This evidence has given rise to an avalanche of recent interest in how best to develop effective school leaders; governments, foundations, universities, and private sector organizations are all in the business of energetically evaluating existing programs and developing new ones (Hallinger, 2003).

---

\*Corresponding author. OISE/University of Toronto, 252 Bloor Street West, Toronto, Ontario, M5S 1V6, Canada. Email: kleithwood@oise.utoronto.ca

Without minimizing the considerable progress that has been made over the past 15 years, however, it is safe to say that the nature of effective school leadership still remains much more of a black box than we might like to think. For example, arguably the most robust efforts to assess school leader effects (Hallinger & Heck, 1996a,b, 1998) do not distinguish among forms of leadership. Furthermore, while there are several important exceptions (Hallinger & Murphy, 1985) “instructional leadership,” a concept uniquely popular in the U.S., is typically treated as a slogan rather than a precise model of leadership that would lend itself to empirical inquiry.

A substantial proportion of the contemporary literature on educational leadership, in addition, is entirely speculative or “theoretical” in nature, or offers admittedly thoughtful arguments for viewing the school organization differently (Sergiovanni, 2000) with “implications” for school leadership. And one slice of the educational literature seems mostly to be about “leadership by adjective”; a new qualifier is added to the term leadership at least annually, creating the misguided impression that something new has been discovered. Finally, and almost the reverse of much of the academic research on leadership in non-school contexts—see research reported in *The Leadership Quarterly*, for example—a substantial amount of research on school leadership is case based in nature only. Such research is a necessary but not sufficient foundation on which to build robust understandings of school leadership. These characteristics and limitations of contemporary leadership research mean that much more large-scale, sustained research about the nature and effects of precisely conceptualized and adequately measured forms of school leadership is required if it is to provide reliable support to leadership developers and others attempting significant school reform.

Our purpose in this study was to test the effects of a transformational model of school leadership, effects on teachers, their classroom practices, and student learning. We did this in a context relevant to many school leaders around the world—large-scale efforts initiated by those in government to improve local schooling. Data for the study were collected as part of a larger external evaluation of England’s National Literacy and Numeracy Strategies (NLS and NNS) commissioned by the British government and carried out over a 4-year period by our colleagues and us (Earl et al., 2002, 2003). This study is the latest in a long program of research on transformational school leadership begun in the early 1990s.

### **The National Strategies as the Context for the Research**

The National Literacy and Numeracy strategies (NLNS) were introduced as a national program beginning with Literacy in the fall of 1998, followed by Numeracy a year later. These strategies represent an ambitious and sophisticated effort at education reform, bringing together agencies and organizations across the country with the goal of raising standards of literacy and numeracy in primary schools. Key features of the reform effort include a broad range of financial and practical supports intended to ensure that the strategies are implemented and educational

standards raised. Based on the assumption that all students can achieve, the Strategies are ambitious in their scope and expectations. More than 18,500 primary schools with approximately 200,000 teachers and about 4.5 million students are expected to implement the program. Those schools opting out of the program are nonetheless expected to achieve targets set by the national Department for Education and Employment. In 1997, the newly elected government publicly announced targets for the year 2002: 80% of 11-year-olds are to achieve the standards set for their age in reading and writing and 75% are to achieve the standards in numeracy. At the time the targets were set, just over 50% of 11-year-olds were performing at the level expected for their age. New targets have been set more recently.

A large and complex set of initiatives has been undertaken to implement the strategies, for example:

- establishment of a National Literacy and Numeracy Centre, with 12 regional directors for Literacy and the same number for Numeracy, to support implementation across the country;
- establishment of over 600 Literacy and Numeracy Consultant positions at the Local Education Authority (district) level to provide both training and in-school support;
- an explicit focus on changing teaching practices linked to increases in pupil learning;
- development of high quality curriculum materials, resource documents, training programs, and videos depicting “good” practice;
- an expectation that every class will have a daily math lesson and a daily literacy hour;
- a substantial investment of new money sustained over at least 6 years and skewed towards those schools which need help most.

In terms of both breadth and depth, at the time of our study NLNS may have been the most ambitious large-scale educational reform effort attempted anywhere.

## **Framework**

Six variables and their relationships, as in Figure 1, constitute the framework for our study. This framework assumes that for large-scale reform to achieve its own goals, school staffs must be motivated to respond to the reform in some locally meaningful and productive way. As well, there must be opportunities for individual teachers to acquire the knowledge and skills needed for such a response. The school settings and wider environment in which school staffs work, have important consequences for the development of motivation and capacity.

Teachers’ motivations, capacities, and work settings have a direct effect on their school and classroom practices. These practices are clearly intended to improve student learning but may or may not do so depending on their effectiveness.

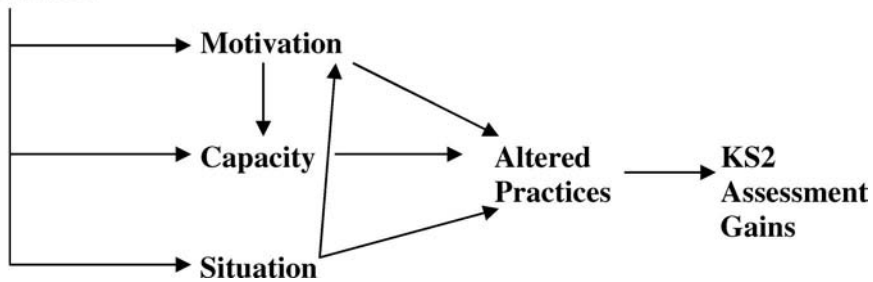
**Leader Practices**

Figure 1. Explaining leaders' effects on teachers and their practices

Transformational school leadership practices on the part of “those in positions of responsibility,” to use the language in our measures, have both direct and indirect effects on teachers' practices, the indirect effects being realized through leaders' influence on teachers' motivation, capacity, and work settings.

### *Transformational School Leadership*

Our study tested the effects of a school-specific model of transformational leadership practices. While there is much discussion in the educational literature, both supportive and critical, about transformational orientations to leadership, empirical evidence about its effects in school contexts is extremely thin. Virtually all of this evidence, however, attests to the suitability of transformational leadership practices in schools faced with significant challenges for change (e.g., Day, Harris, Hadfield, Tolley, & Beresford, 2000; Leithwood, Jantzi, & Steinbach, 1999), and to the contribution of this form of leadership, when exercised by heads or principals, to a wide array of organizational and pupil outcomes (e.g., Leithwood & Jantzi, in press; Leithwood, Tomlinson, & Genge, 1996). Comparable claims are made for this approach to leadership in non-school contexts as well (Lowe, Kroeck, & Sivasubramaniam, 1996; Yukl, 1999). To our knowledge, there is no empirical support for the factual claims or speculations contained in the small but steady stream of criticisms of this form of leadership in schools.

All transformational approaches to leadership emphasize emotions and values (Yukl, 1999) and share in common the fundamental aim of fostering capacity development and higher levels of personal commitment to organizational goals on the part of leaders' colleagues. Increased capacities and commitments are assumed to result in extra effort and greater productivity. Authority and influence associated with this form of leadership are not necessarily allocated to those occupying formal administrative positions, although much of the literature adopts their perspectives. Rather, power is attributed by organizational members to whomever is able to inspire their commitments to collective aspirations, and the desire for personal and collective mastery over the capacities needed to accomplish such aspirations. Recent evidence suggests that practices associated with transformational leadership may be widely

distributed throughout the organization (Leithwood, Jantzi, Earl, Fullan, & Levin, 2004). So there is no need, as some would claim (e.g., Lingard & Christie, 2003), to view this as a “heroic” model of leadership; it was neither conceptualized nor measured as such in this study.

Although influenced in its initial formulations by well-known, “classical,” views (Bass, 1985; Burns, 1978), our model of transformational leadership was developed from our own qualitative and quantitative research in schools, specifically. This means that some of the central features associated with transformational leadership by others either are not part of our model (e.g., charisma, see Hunt & Conger, 1999) or are awarded quite different significance (e.g., transactional practices, see Yukl, 1999); it also means that our model includes dimensions of practice not found in other models (e.g., creating productive community relationships). Differences such as these are almost never acknowledged by those commenting on our work.<sup>1</sup>

Three broad categories of leadership practices, including a total of nine more specific dimensions of practice, are encompassed in our model. Included in the category *Setting Directions* are the dimensions building school vision, developing specific goals and priorities, and holding high performance expectations. In the category *Developing People* are the dimensions providing intellectual stimulation, offering individualized support, and modeling desirable professional practices and values. The third category, *Redesigning the Organization*, includes the dimensions developing a collaborative school culture, creating structures to foster participation in school decisions, and creating productive community relationships. Each dimension is made up of multiple, more specific, practices which encourage contingent responses on the part of leaders depending on the contexts of their work. Justification for each of these categories, dimensions, and specific practices is provided elsewhere (e.g., Leithwood et al., 1999).

The specificity of practices in our model addresses one of the main conceptual weaknesses—ambiguity about behaviors—attributed by Yukl (1999, p. 288) to other transformational models. The wide range of dimensions in our model addresses a second conceptual weakness, the omission of leadership behaviors demonstrated by research to be important; some of the behaviors in our model not found in Bass’s (1996) model, for example, include “facilitating agreement about objectives and strategies,” “increasing followers self-efficacy,” and “articulating a vision and strategy for the organization” (Yukl, 1999, p. 290). Yukl also claims that most transformational leadership models overemphasize dyadic relationships—influence on individual only, rather than on both individual and group processes. The third category of practices in our model, *Redesigning the Organization*, explicitly focuses not only on group processes but on organization-wide conditions, as well.

### *Teachers and Their Workplaces*

Motivation, capacity, and work settings are key variables in a general model of employee performance developed most fully in the literature on organizational and industrial psychology (Rowan, 1996). According to this model, for example, neither

high ability and low motivation, nor high motivation and low ability produce high levels of employee performance; neither does high ability and high motivation in a dysfunctional job situation. Furthermore, a dysfunctional situation will likely depress initially high levels of ability and motivation.

The more detailed formulation of this framework developed for our larger study (Leithwood, Jantzi, & Mascal, 2002) and briefly described here, was based on contemporary theoretical and empirical accounts of the conditions required for the development of both motivation and capacity on the part of those in schools to productively engage large-scale reform initiatives. It also incorporated accounts of organizational conditions and characteristics of the infrastructure or "situation" which facilitate the successful implementation of large-scale reform, what Van Den Berg and Sleegers (1996) refer to as the organization's "innovative capacity."

*Motivation.* For this study, motivational processes were conceptualized as qualities of a person oriented toward the future and aimed at helping the person evaluate the need for change or action (Ford, 1992). These processes are a function of one's personal goals, beliefs about one's capacities, beliefs about one's work context, as well as emotional arousal processes (Bandura, 1986; Ford, 1992).

Personal goals influence the actions of individual teachers and administrators when they make them their own. And while personal goals are an important launching pad for motivation, they must be perceived to possess at least three qualities in order actually to energize action. This occurs when a person's evaluation of present circumstances indicates that it is different from the desired state, when the goals are perceived to be hard but achievable, and when they are short term but understood within the context of longer term and perhaps more important, more obviously valuable purposes.

To be motivational, people must also believe themselves to be capable of accomplishing these goals (whether or not they actually are). Perceived capacity or self-efficacy increases the intrinsic value of effort and contributes to the possibilities for a sense of collective capability or efficacy on the part of a group, as well. Increased perceptions of capacity or self-efficacy may result from teachers considering information from three sources: perceptions of success perhaps formed through feedback from others; vicarious experience often provided by role models; and verbal persuasion.

Context beliefs are beliefs about congeniality of the situation in which one finds oneself for implementing whatever the reform initiative might be. Many experienced teachers have developed negative context beliefs over their careers as a consequence of being associated with mismanaged or ill-conceived innovations.

Finally, emotional arousal processes are feelings which have motivational value when they are associated with a personal goal that is currently influencing one's actions. Positive emotions arise when an event promises to help meet a personal goal; negative emotions when chances of achieving one's goal are harmed or threatened. Such emotions may arise from frequent positive feedback about one's work and a dynamic changing job.

*Capacity.* We are referring here to the actual ability required to perform, not, as in the case of motivation, the self-belief in one's ability. Identifying the conditions for capacity-building in schools, depends on an understanding of how learning occurs within different organizational "units"—the individual, the small group, and the whole school. Learning within these units does not just happen naturally. It is stimulated by sources that also must be better understood. Furthermore, such characteristics of the organization as its goals, culture, and structure, for example, not only mediate organizational learning, they are influenced by such learning, as well. These conditions change in response to organizational learning, and these changed conditions have a direct or indirect effect, eventually, on the educational experiences of pupils.

Because it is individual capacity that is of concern at this point, the nature of individual learning is most relevant. Describing such learning as "self-regulated knowledge construction," Vermunt and Verloop (1999, p. 258) suggest that it is both cognitive and metacognitive (or regulative) in nature. Cognitive processes include, for example, relating, analyzing, applying, memorizing, critical processing, and selecting. Metacognitive or regulative processes involve planning, monitoring, adjusting, and evaluating. Since understanding is constructed through such processes, moving toward a reasonably uniform, widespread view of a reform initiative, especially one consistent with what its developers had in mind, will require repeated communication of many forms, and many opportunities to engage in discussions aimed at clarification of policy initiatives.

Theories of individual mental functioning often distinguish, as well, between declarative knowledge (the "understanding" described above) and procedural (how to) knowledge. Procedural knowledge, the basis of skilled practice, develops through repeated cycles of (a) developing a knowledge structure to guide one's mental or physical activity, (b) engaging in that activity guided by the knowledge structure, (c) obtaining feedback about the adequacy of one's "actions," and (d) refining the guiding knowledge structure. Reform initiatives requiring significantly different practices in schools need to provide opportunities for repeated iterations of this cycle.

*Work settings.* While "capacity" is typically considered a quality of individuals, its meaning and importance as a collective or organizational property increasingly is being recognized. To acknowledge the school as a unit of change, for example, implies that its capacity is more than the sum of its individual members' capacities. Research on organizational learning provides considerable support for this implication (Hutchins, 1995; Leithwood, Leonard, & Sharratt, 1998).

Basic conditions influencing collective capacity are evident in the growing understandings of learning as socially constructed and situated (Salomon & Perkins, 1997). Such understandings assume a significant role for the immediate situation in which the learner finds herself, as well as the larger cultural context in which that situation is embedded. According to theories of distributed cognition, each member of a well-functioning group must share some of the same understandings—purposes

for the collective work and constraints within which the group must function, for example. But each member also brings some unique capacities to the group's work. It is in this sense that the total capacity of the group is distributed across its members (Hutchins & Klausen, 1998).

For purposes of the study, the teachers' work setting was defined by two dimensions. One dimension included the teachers' collective practices relevant to the large-scale reform, along with the physical and social infrastructure supporting such practices. The second dimension was the collective efficacy of the staff, a construct that compelling recent evidence suggests is capable of explaining a large proportion of the variation in pupil performance (Hoy, Sweetland, & Smith, 2002). Like individual self-efficacy, collective efficacy is associated with effort-based conceptions of success on the part of teachers.

### *Classroom Practices*

The orientation to classroom practices adopted for purposes of this study, reflected two distinct lines of evidence: evidence about implementation processes, and evidence about the importance of time in accounting for pupils' learning. Implementing new policies and programs seldom means to school staffs exactly what policy-makers or change advocates have in mind. Indeed, early research on program implementation argued that the ideal form of implementation was "mutual adaptation" (Berman & McLaughlin, 1977), a compromise between implementers' existing practices and those practices specified by new policies and programs. At its best, such adaptation entails customizing externally developed policies and programs in light of unique features of the context, and, in the process, making them more suitable for that context.

Based on this view of local implementation, we inquired about changes in the teaching of literacy and mathematics (without asking about the nature of those changes). Our results, however, need to be interpreted in light of evidence that many teachers respond to calls for complex changes in classroom practice by adapting their practices in response to only the relatively superficial aspects of those changes, aspects unlikely to have the intended effects on students (e.g., Cohen & Barnes, 1993). So teachers' self-reported changes may range from the superficial to the fundamental. It is also important to acknowledge that we did not directly ask teachers whether changes in their practices corresponded to those practices advocated by the Literacy and Numeracy Strategies.

Research about pupil time conceptualizes it variously as allocated time, teaching time, and academically engaged time. But increases in all these conceptions of time are associated with increases in pupil learning. The amount of evidence supporting this claim is both large and unambiguous (Glass, 2002). A second set of questions included in our research about changed practices asked whether more time was being spent on planning for, and actually delivering, teaching in literacy and numeracy in response to the strategies. Either changed teaching practices or increased time potentially could account for increased pupil learning.



### *Student Achievement*

Because our larger study was about England's efforts to improve primary-age students' literacy and numeracy skills, we used the government's own Key Stage 2 tests as our concept of student achievement. The Key Stage 2 literacy test includes measures of reading, writing, and spelling, while the Key Stage 2 numeracy test measures skills associated with, for example, mental arithmetic, number and algebra, shapes, space and measures, handling data, and using calculators (Qualifications and Curriculum Authority, 2001).

## **Methods**

### *Sample*

Data for our larger study were collected from those in many roles including a national sample of teachers and headmasters in England. Only the teacher data are used for this study. Two representative samples of 500 schools each were selected, one sample to provide evidence from teachers about NLS and one to provide evidence about NNS. Both samples were selected at random from England's National Foundation for Educational Research (NFER)<sup>2</sup> database of schools to be representative of the whole of England's primary school population in terms of school type, national curriculum test results, region, and proportion of pupils eligible for free school meals. Independent schools were included in the sample, even though the NLS and NNS were not compulsory in these schools, since some of them were choosing to implement all or part of the strategies anyway.

### *Instruments*

Two forms of a Likert-type teacher survey were developed to measure all constructs in the framework except student achievement. One form focused on NLS and one on NNS. These instruments, field tested and refined over several stages, included a 5-point response scale for most questions (1 = strongly disagree, 2 = agree, 3 = undecided, 4 = agree, 5 = strongly agree). Our measures of student achievement were gains in Key Stage 2 results.

### *Data Collection Procedures*

All surveys were sent to headteachers with an accompanying letter from the external evaluation team to introduce the surveys and to request that surveys be distributed to all staff. Numbers of staff in each school were estimated using the NFER's most recently available database of schools. Number of fulltime equivalent teachers was used to form the basis of a very generous estimate of total teaching staff to allow for the possibility of large numbers of parttime or job-share teachers. Each individual teacher survey contained a reply-paid envelope. Greater detail about these procedures is provided in Earl et al. (2003).

Pupil performance data for all schools in the sample were obtained from England's Qualifications and Curriculum Authority (QCA). QCA provided end of Key Stage 2 results in English and mathematics for the years 1997–1998, 1998–1999, and 1999–2000. In addition to performance results, data were provided for each school that gave percentages of pupils who sat the relevant test, were absent, were excused under sections 364/365 of the Education Act, or were exempted because they were working below the level of the tests.

### *Data Analysis*

Survey responses were analyzed at both the individual and school levels. Statistical Package for the Social Sciences (SPSS) was used to calculate means, standard deviations, reliability coefficients (Cronbach's alpha), correlation coefficients, and to aggregate individual responses by school, as appropriate. Independent sample *t* tests were used to compare mean ratings of literacy respondents to those of numeracy respondents to determine whether differences were statistically significant. Paired-samples *t* tests were used to compare mean ratings of specific components within a strategy, that is reading and writing for literacy or mental mathematics and mathematical concepts for numeracy. *t*-test results should be viewed with caution, however, because most item responses are skewed in a positive direction, rather than being distributed normally.

LISREL was used to assess the direct and indirect effects of leadership on motivation, capacity, and situation, as well as the effects of all these variables on altered teacher practices. This path analytic technique allows for testing the validity of causal inferences for pairs of variables while controlling for the effects of other variables. Data were analyzed using the LISREL 8 analysis of covariance structure approach to path analysis and maximum likelihood estimates (Jöreskog & Sörbom, 1993). Hierarchical linear modeling (HLM) is the analytic technique of choice for some researchers, at present, exploring data bases similar to this one. For a variety of practical reasons, however, we were unable to collect our data in a way that allowed us to link the responses of individual teachers with data from their pupils, a prerequisite for HLM. The consequences of aggregating data to the school level is loss of information about potentially significant variation across classrooms within schools. Tests of a partial model, which included the relationships up to altered classroom practice but excluded assessment results, were conducted at the individual teacher level as a means of examining the data available at both the school and individual teacher level.

## **Results**

### *Characteristics of the Achieved Sample*

School response rates for NLS and NNS respectively were 64% ( $N=340$ ) and 63% ( $N=315$ ) and features of the responding schools largely mirrored primary schools

in the country as a whole<sup>3</sup> on the basis of school type, urban (metropolitan/not metropolitan), region, and 1998 Key Stage 2 results. The sample of schools with a full set of achievement data was slightly smaller at 256 (51%) for Literacy and 258 (52%) for Numeracy. The only difference large enough to note is that, as compared with the whole school population, the achieved sample of schools was slightly overrepresented by junior schools and underrepresented by independent schools. There is no apparent reason to think this difference jeopardized the representative nature of the achieved school sample.

For teachers from all schools surveyed, the response rate was 20.4% ( $N = 1110$ ) for the literacy strategy and 20.3% ( $N = 1180$ ) for the numeracy strategy. Response rates rose to 32.4% and 33.4%, respectively, for teachers in schools from which there was at least one response. These rates likely underestimate the proportion of teachers responding, perhaps significantly, because our method of calculating the intended sample was based on the numbers of instruments sent to each school. These numbers intentionally overestimated the size of teaching staffs and made provision for parttime teachers, as well, since we did not have access to actual school staff lists.

While teacher response rates were modest, teachers in the achieved sample were very similar to the population of England's primary school teachers in terms of gender and experience. We explore the significance of the teacher survey return rate in the concluding section of the article, providing additional evidence in support of our view that it is less of a threat to the purposes of this study than it might seem.

#### *Summary of Responses to Surveys*

Because the main purpose of our study it to test the effects of transformational leadership, our initial description of teachers' responses to survey items and scales is limited to a narrative summary of results, the details of which can be found in Table 1. For both forms of the survey, Table 1 outlines the means and standard deviations of teachers' responses to each item and each scale on the surveys, as well as the reliability (Cronbach's alpha) of each scale.

*Transformational leadership.* Whether in reference to the Literacy or Numeracy Strategies, teachers expressed weak levels of overall agreement that transformational practices were being used to assist with implementation ( $m = 3.71$  for literacy and 3.77 for numeracy on the 5-point Likert scale). Among such practices, most in evidence to teachers were leaders helping to clarify the reasons for implementing the Strategies ( $m = 3.87$  and 3.91), demonstrating high expectations for their work with pupils ( $m = 3.92$  and 3.94), and modeling high levels of professional practice in relation to the Strategies ( $m = 3.73$  and 3.85). From the perspective of teachers, however, leaders were providing relatively little individualized support for them as they implemented the Strategies ( $m = 3.53$  and 3.60). Nor was there much evidence of leaders creating conditions in the school to allow for wide participation in decisions about the Strategies ( $m = 3.60$  and 3.53), or helping to develop good relationships

Table 1. Means, standard deviations (*SD*) and scale reliabilities of teachers' survey responses

|  | Literacy    |            | Numeracy    |            |
|--|-------------|------------|-------------|------------|
|  | Mean        | <i>SD</i>  | Mean        | <i>SD</i>  |
| <b>Measures of Transformational Leadership</b>   | <b>3.71</b> | <b>.72</b> | <b>3.77</b> | <b>.66</b> |
| <i>Those in positions of responsibility in your school:</i>  |             |            |             |            |
| <b>Setting Directions</b>  | <b>3.83</b> | <b>.74</b> | <b>3.91</b> | <b>.69</b> |
| 1. Helped clarify the reasons for implementing the Strategy.   | 3.87        | .84        | 3.98        | .75        |
| 2. Provided useful assistance to you in setting short-term goals for l/n teaching and learning.  | 3.70        | .92        | 3.80        | .85        |
| 3. Demonstrated high expectations for your work with pupils in l/n.  | 3.92        | .85        | 3.94        | .79        |
| <b>Developing People</b>   | <b>3.66</b> | <b>.85</b> | <b>3.74</b> | <b>.77</b> |
| 4. Given you individual support to help you implement the Strategy.  | 3.53        | 1.01       | 3.60        | .96        |
| 5. Encouraged you to consider new ideas for your teaching of l/n.  | 3.68        | .91        | 3.76        | .84        |
| 6. Modeled a high level of professional practice in relation to the Strategy.  | 3.73        | .99        | 3.85        | .87        |
| <b>Redesigning the Organization</b>  | <b>3.66</b> | <b>.79</b> | <b>3.65</b> | <b>.73</b> |
| 7. Encouraged collaborative work among staff.  | 3.72        | .94        | 3.70        | .90        |
| 8. Created conditions in the school which allow for wide participation in decisions about the Strategy.  | 3.60        | .98        | 3.56        | .92        |
| 9. Helped develop good relationships with parents as part of the school's efforts to respond productively to the Strategy.   | 3.65        | .86        | 3.70        | .82        |
| Reliability of Administrative Leadership Scales: Cronbach's alphas Setting Directions = .81 NLS, .82 NNS; Developing People = .85 NLS, .84 NNS; Redesigning Organizations = .80 NLS, .78 NNS; Administrative Leadership Aggregate = .90 NLS, .88 NNS |             |            |             |            |
| <b>Measures of Teacher Capacities</b>  | <b>3.55</b> | <b>.62</b> | <b>3.98</b> | <b>.53</b> |
| 10. I have the knowledge and skill I need to implement the Strategy.   | 3.98        | .77        | 4.09        | .68        |
| 11. I am having success implementing the Strategy.   | 3.84        | .76        | 3.03        | .63        |
| 12. I have had many opportunities to think and talk about how practices recommended by the Strategy relate to my own classroom practices.  | 3.41        | 1.02       | 3.52        | .99        |

*(continued)*

Table 1. (Continued)

|   | Literacy    |            | Numeracy    |            |
|---|-------------|------------|-------------|------------|
|   | Mean        | SD         | Mean        | SD         |
| 13. I have had opportunities to practice and refine new teaching skills required for implementing the Strategy.                                       | 3.60        | .87        | 3.70        | .82        |
| 14. My teaching of reading/mental mathematics has become more effective as a result of the Strategy.  | 3.18        | 1.10       | 4.23        | .77        |
| 15. My teaching of writing/mathematical concepts has become more effective as a result of the Strategy.   | 3.35        | 1.10       | 3.93        | .90        |
| 16. My pupils have benefited from the reading/mental mathematics component of the Strategy.   | 3.61        | .94        | 4.30        | .65        |
| 17. My pupils have benefited from the writing/mathematical concepts component of the Strategy.  | 3.43        | 1.03       | 3.04        | .76        |
| Reliability of Capacity Scale: Cronbach's alpha = .81 NLS, .83 NNS  |             |            |             |            |
| <b>Measures of Teachers' Motivation</b>   | <b>3.70</b> | <b>.47</b> | <b>3.82</b> | <b>.43</b> |
| 18. The aims of the Strategy are clear to me.   | 4.15        | .67        | 4.28        | .57        |
| 19. The aims of the Strategy are consistent with my own aims about teaching l/n in my classroom   | 3.75        | .83        | 4.08        | .66        |
| 20. I have been involved in setting Key Stage 2 targets in this school.   | 3.81        | 1.14       | 3.80        | 1.16       |
| 21. I have been involved in setting curriculum targets for pupils in my class.  | 4.33        | .67        | 4.26        | .72        |
| 22. I use the school's curriculum targets in my planning.   | 4.09        | .78        | 4.01        | .83        |
| 23. If a pupil did not understand what I was teaching in a lesson, I would know what to do next to increase his/her understanding.                    | 3.88        | .71        | 3.88        | .72        |
| 24. If a pupil in my class becomes disruptive and noisy during l/n lessons, I feel confident that I know some techniques to redirect him/her quickly. | 3.97        | .80        | 4.02        | .71        |
| 25. If one of my pupils could not do the work in l/n, I would be able to assess whether the task was at the correct level of difficulty.              | 4.18        | .58        | 4.18        | .57        |

(continued)

Table 1. (Continued)

|   | Literacy    |            | Numeracy    |            |
|---|-------------|------------|-------------|------------|
|   | Mean        | SD         | Mean        | SD         |
| 26. If I really try hard, I can teach l/n to even the most difficult or unmotivated pupils.   | 3.57        | .97        | 3.72        | .85        |
| 27. The Strategy training in which I participated was very useful.  | 3.43        | 1.07       | 3.83        | .90        |
| 28. The climate in my school is consistent with efforts to implement the Strategy.  | 4.12        | .73        | 4.24        | .59        |
| 29. I have the flexibility that I need to implement the Strategy in a manner that I believe is effective for my pupils                      | 3.80        | .94        | 3.93        | .76        |
| 30. I have all the information that I need about the expectations associated with the Strategy.   | 3.54        | .84        | 3.71        | .75        |
| 31. I have access to the resources (e.g., people, materials) that I need to implement the Strategy.   | 3.40        | 1.07       | 3.55        | .99        |
| 32. I am able to find the teaching time needed to implement both the Literacy and Numeracy Strategies.                                      | 3.66        | .99        | 3.66        | 1.00       |
| 33. I am able to find the planning time needed to implement both the Literacy and Numeracy Strategies.                                      | 2.79        | 1.26       | 2.94        | 1.21       |
| 34. I receive useful feedback about my use of teaching practices related to the Strategy (e.g., from the head teacher, co-ordinator, etc.). | 3.44        | 1.01       | 3.46        | 1.00       |
| 35. The Strategy has made my job more satisfying and engaging.  | 2.76        | 1.08       | 3.27        | .99        |
| Reliability of Motivation Scale:<br>Cronbach's alpha = .82 NLS, .82 NNS   |             |            |             |            |
| <b>Measures of Teachers' Work Settings</b>  | <b>3.64</b> | <b>.51</b> | <b>3.68</b> | <b>.45</b> |
| 36. My colleagues and I function as a team in implementing the Strategy.  | 3.95        | .87        | 3.96        | .79        |
| 37. My colleagues and I build on one another's strengths in implementing the Strategy.  | 3.75        | .92        | 3.71        | .88        |
| 38. My colleagues and I assist one another, as needed, in implementing any new classroom practices required by the Strategy.                | 3.89        | .86        | 3.84        | .83        |

(continued)

Table 1. (Continued)

|   | Literacy    |            | Numeracy    |            |
|---|-------------|------------|-------------|------------|
|   | Mean        | SD         | Mean        | SD         |
| 39. Structures (e.g., timetables, planning arrangements) in this school allow for opportunities to collaborate in teaching with colleagues.   | 3.28        | 1.17       | 3.26        | 1.15       |
| 40. The physical layout of the school is conducive to discussion with colleagues about teaching and learning.   | 3.12        | 1.09       | 3.14        | 1.00       |
| 41. Good teachers can teach l/n to even the most difficult pupils.  | 3.46        | .96        | 3.56        | .94        |
| 42. Teachers are very limited in what they can achieve in teaching l/n because a pupil's home environment is a large influence on learning. (scale reversed)                                | 3.14        | .95        | 3.53        | .88        |
| 43. When it comes right down to it, a teacher really can't influence l/n much because most of a pupil's motivation and performance depends on his or her home environment. (scale reversed) | 3.67        | .88        | 3.89        | .81        |
| 44. Colleagues in the school feel a sense of responsibility for the quality of teaching in their classrooms.  | 4.54        | .59        | 4.49        | .58        |
| 45. Colleagues in the school feel a sense of responsibility for engaging in school-wide decisions that influence their teaching.  | 4.09        | .81        | 4.05        | .80        |
| 46. Parents are supportive of this school's efforts in l/n.   | 3.56        | .87        | 3.56        | .79        |
| 47. Non-parent members of the community are supportive of the school's efforts in l/n.  | 3.15        | .88        | 3.04        | .85        |
| Reliability of Situation Scale:<br>Cronbach's alpha = .79 NLS,<br>.76 NNS   |             |            |             |            |
| <b>Measures of Teachers' Classroom Practices</b>  | <b>3.37</b> | <b>.59</b> | <b>3.52</b> | <b>.62</b> |
| 48. I have changed my teaching of reading/mental mathematics as a result of the Strategy.   | 3.43        | 1.04       | 4.01        | .90        |
| 49. I have changed my teaching of writing/mathematical concepts as a result of the Strategy.  | 3.56        | .99        | 3.69        | .98        |

(continued)

Table 1. (Continued)

|  | Literacy |      | Numeracy |      |
|--|----------|------|----------|------|
|  | Mean     | SD   | Mean     | SD   |
| 50. I spend significantly more of my daily classroom time teaching reading/mental mathematics than I did before the Strategy.    | 2.93     | 1.14 | 3.98     | .99  |
| 51. I spend significantly more of my daily classroom time teaching writing/mathematical concepts than I did before the Strategy. | 2.96     | 1.12 | 3.26     | 1.11 |
| 52. I spend significantly more time planning for l/n teaching than I did before the Strategy.                                    | 4.33     | .95  | 3.85     | 1.12 |
| 53. I spend significantly more time assessing pupils' work in l/n than I did before the Strategy.                                | 3.45     | 1.14 | 3.37     | 1.10 |
| 54. My pupils spend significantly more of their time in school focusing on l/n than they did before the Strategy.                | 3.46     | 1.15 | 3.13     | 1.12 |
| 55. My pupils spend significantly more of their homework time on l/n than they did before the Strategy.                          | 3.07     | 1.08 | 3.20     | 1.07 |
| 56. Parents spend significantly more time helping their children with l/n now than they did before the Strategy.                 | 2.62     | .86  | 2.89     | .88  |
| 57. The Literacy and Numeracy Strategies require me to use complementary forms of teaching.                                      | 3.81     | .75  | 3.74     | .76  |
| Reliability of Classroom Practice Scale: Cronbach's alpha = .76 NLS, .82 NNS   |          |      |          |      |

with parents ( $m = 3.65$  and  $3.70$ ), as part of the school's effort to respond productively to the Strategies.

Although levels of transformational leadership provided in schools were perceived to be low, as we demonstrate later in these results, variation in levels of such leadership was significantly related to changes in school conditions and teacher practices.

*Teacher motivation.* A total of 18 items on the teachers' surveys measured the four sources of motivation identified by our framework (teachers' professional goals, beliefs about their capacities to implement NLNS and the supportiveness of their



work setting, and emotions invoked by the effort. Considered together, results suggest modest levels of motivation for implementing either of the Strategies. There was, however, significantly greater motivation to implement the Numeracy Strategy (overall mean = 3.82) than the Literacy Strategy (overall mean = 3.70), these differences reaching statistical significance for 12 of the 18 survey items. In particular, as compared with the Literacy Strategy, teachers perceived the Numeracy Strategy to have goals more consistent with their own goals (Item 19:  $m = 3.75$  vs. 4.08). Teachers also perceived the Numeracy Strategy to offer more useful training (Item 27:  $m = 3.43$  vs. 3.83), and to foster greater, though still not high, levels of job satisfaction and engagement (Item 35:  $m = 2.76$  vs. 3.27).

For teachers, the strongest sources of motivation were participation in setting curriculum targets (Item 21:  $m = 4.33$  and 4.26), using the results of such target setting for their own planning (Item 22:  $m = 4.09$  and 4.01), confidence in their abilities to diagnose the level of difficulty of pupils' tasks (Item 25:  $m = 4.18$  and 4.18) and to manage classroom disruptions (Item 24: 3.97 and 4.02), and beliefs about the supportive nature of the overall climate in their school (Item 28;  $m = 4.12$  and 4.24). Most likely to erode teachers' motivation to implement both Strategies were lack of planning time needed to implement the Strategies (Item 33:  $m = 2.75$  and 2.74) and low levels of job satisfaction and engagement resulting from implementing the Strategies, especially in Literacy (Item 35:  $m = 2.76$  and 3.27).

*Teacher capacities.* Eight items on the survey assessed the extent to which teachers perceived opportunities to acquire relevant knowledge and skill, the possession of such knowledge and skill, estimates of success with the Strategies, and the effects of implementing the Strategies on improving teaching. Results indicated modest levels of overall agreement with all eight survey items measuring capacities. Agreement was substantially greater for the Numeracy than the Literacy Strategy (overall mean = 3.55 and 3.98).

In relation to the Literacy Strategy, teachers were most confident that they had the knowledge and skill needed for implementation (Item 10:  $m = 3.98$ ), and that they were having success with the Strategy (Item 11:  $m = 3.84$ ). Greatest improvement in Literacy teaching was reported to be for writing, but pupils were considered to have benefited most from changes in the teaching of reading (Item 15:  $m = 3.61$ ) as compared with writing (Item 16:  $m = 3.43$ ). For the Numeracy Strategy, teachers were most confident that their teaching of mental math (Item 16:  $m = 4.30$ ) had improved as a consequence of the Strategy resulting in substantial benefits to pupils in that area.

The common view that teachers know less about teaching math than language was not supported by these findings. Teachers reported that they were slightly, to quite a bit better, prepared to teach numeracy and that their pupils were benefiting more, as well.

*Teachers' work settings.* Responses to the 12 items measuring work setting indicated that teachers were most positive about the teamwork used to implement the Strategies

in their schools (Item 36:  $m = 3.95$  and  $3.96$ ) and about the assistance staff provided to one another (Item 38:  $m = 3.89$  and  $3.84$ ). Weakest agreement was with the contribution of the schools' physical layout (Item 40:  $m = 3.12$  and  $3.14$ ) and structures to allow staff opportunities to collaborate with one another (Item 39:  $m = 3.28$  and  $3.26$ ).

*Teachers' classroom practices.* The overall response to the 10 items measuring the extent to which teachers and headteachers perceived changes in their classroom practices in response to the Strategies was uncertainty (Literacy: overall mean =  $3.37$ ) to weak agreement (Numeracy: overall mean =  $3.52$ ). Teachers rated both changed practices (see Items 48 and 49 in Table 1) and increased teaching time (see Items 50 and 51 in Table 1) significantly higher for mathematics than for literacy. In response to NNS, results suggest that teachers were more likely to have changed their teaching of mental mathematics than math concepts (Items 48 and 49:  $m = 4.01$  vs.  $3.69$ ), as well as increased the time they spent on mental math in class (Items 50 and 51:  $m = 3.98$  vs.  $3.26$ ). Least change was reported to have occurred in the overall amount of time devoted to mathematics in school (Item 54:  $m = 3.13$ ), time spent on homework (Item 55:  $m = 3.20$ ), and time spent by parents helping their children at home (Item 56:  $m = 2.89$ ).

Results for NLS indicate that teachers had changed their teaching of writing significantly more than their teaching of reading (Items 48 and 49:  $m = 3.56$  vs.  $3.43$ ), and that they had increased the amount of time spent planning for literacy teaching (Item 52:  $m = 4.33$ ). There was little evidence that this extra planning time actually resulted in any more class time devoted to the teaching of reading or writing in the classroom (Items 50 and 51:  $m = 2.93$  and  $2.96$ ), or any more of parents' time helping their children at home (Item 56:  $m = 2.62$ ).

*Student achievement.* The 1998 English results provided baseline data for Literacy sample and the 1999 mathematics results for numeracy. Performance targets were set by the strategies for the percentage of students who performed at Level 4 and test results were reported as percentages of students at each of the levels within a school, educational authority (district) and nationally. This study used the school mean percentage at Level 4 as the achievement measure. Results for our sample were virtually identical to the national average. The national means for percentage of students performing at Level 4 in English for the 3 years were 64%, 70%, and 75%, respectively, and results for our sample were 66%, 72%, and 76%. Mathematics results for the same 3 years were 58%, 68%, and 72% nationally and 59%, 70%, and 73% for our sample.

#### *Transformational School Leadership Effects*

Structural equation modeling (LISREL), with the school as the unit of analysis, was used to test the direct and indirect effects of transformational leadership practices provided by "those in positions of responsibility" (teachers' own interpretations of who that phrase included) on teachers, their classroom practices, and changes in

student achievement. Leadership was treated as an exogenous variable, teacher motivation, capacity, and work settings as mediating variables, teacher practices as proximate dependent variables, and student achievement as the distal, dependent variables. Depending on their effectiveness, altered teacher practices may contribute directly to changes in student performance.

Table 2 reports correlation coefficients for variables included in the model testing effects on classroom practice and gains in KS2 pupil assessments. Correlation coefficients are based on teacher responses aggregated to the school level. For both NLS and NNS results, altered classroom practices has the strongest relationship with capacity (both  $r = .50$ ) followed by motivation ( $r = .30$  and  $.29$ ). Leadership also has a significant, but weak, relationship with classroom practices in both cases. While work setting is not related to altered classroom practices in the literacy schools, the relationship is weakly significant in the numeracy schools.

Gains in student achievement results are not significantly related to any of the variables in the literacy schools but are related weakly to leadership and work setting in the numeracy schools. These correlation coefficients reflect only the relationship between two variables at a time without considering how other variables might influence that relationship. The results of path analyses reported below show relationships between variables while taking into account the effect of other variables on that relationship.

Several measures are available in the LISREL program to assess the extent to which the model being tested fits the data. We used three indices to measure adequacy of

Table 2. Relationships between variables in two models tested

|                        | Leader | Motivation | Capacity | Work setting | Classroom practice | Achievement gains |
|------------------------|--------|------------|----------|--------------|--------------------|-------------------|
| <b>Literacy:</b>       |        |            |          |              |                    |                   |
| <b>N = 256 schools</b> |        |            |          |              |                    |                   |
| Leader                 | 1.00   | .56**      | .41**    | .71**        | .17**              | .06               |
| Motivation             | .56**  | 1.00       | .73**    | .62**        | .30**              | .01               |
| Capacity               | .41**  | .73**      | 1.00     | .43**        | .50**              | .02               |
| Work setting           | .71**  | .62**      | .43**    | 1.00         | .04                | .03               |
| Classroom practice     | .17**  | .30**      | .50**    | .04          | 1.00               | .03               |
| Achievement gains      | .06    | .01        | .01      | .03          | .03                | 1.00              |
| <b>Numeracy:</b>       |        |            |          |              |                    |                   |
| <b>N = 258 schools</b> |        |            |          |              |                    |                   |
| Leader                 | 1.00   | .65**      | .47**    | .73**        | .12*               | .14*              |
| Motivation             | .65**  | 1.00       | .77**    | .64**        | .29**              | .12               |
| Capacity               | .47**  | .77**      | 1.00     | .39**        | .50**              | .07               |
| Work setting           | .73**  | .64**      | .39**    | 1.00         | .14*               | .12*              |
| Classroom practice     | .12*   | .29**      | .50**    | .14*         | 1.00               | -.03              |
| Achievement gains      | .14*   | .12        | .07      | .12*         | -.03               | 1.00              |

\*\* $p < .01$ ; \* $p < .05$ .

absolute fit. The root mean square error of approximation (RMSEA) should be less than .10 for a good fit, the root mean squared residual (RMR) should be less than .05 (Kelloway, 1998), and the adjusted goodness of fit index (AGF) should be above .90. In addition, a measure of comparative fit was used to assess whether the tested models were better than the null model for which there would be no paths connecting the variables. The normed fit index (NFI) should be over .90 (Kelloway, 1998). As indicated in Table 3, all but Model 6 met all these criteria and although it met only one criterion, the pattern of results for that model was similar to results for the other five.

The results of model testing (Figure 1) are reported in Tables 4 and 5. These tables summarize results of testing the same model with six slightly different sets of data reported in columns labeled 1 through 6 on each table. These tests differ as follows:

- Columns 1 (for NLS) and 2 (for NNS) report use of teacher-level data to test the model using classroom practices as the dependent variable. Because all measures of pupil performance available to us were aggregated to the school level, pupil performance could not be included in these two tests. So, while our data do not

Table 3. LISREL analyses of six versions of the model: indices of model fit

| Versions of the model                    | RMSEA | RMR | AGF | NFI  |
|--|-------|-----|-----|------|
| 1 = 1110 NLS teachers                    | .07   | .01 | .97 | 1.00 |
| 2 = 1180 NNS teachers                    | .09   | .02 | .94 | 1.00 |
| 3 = 256 NLS schools                      | .00   | .02 | .99 | 1.00 |
| 4 = 258 NNS schools                      | .09   | .06 | .91 | .98  |
| 5 = 160 NLS schools with > 2 respondents | .00   | .02 | .99 | .99  |
| 6 = 157 NNS schools with > 2 respondents | .17   | .10 | .77 | .94  |

Table 4. Effects of leadership on teachers and their classroom practices significant standardized total effects,  $t > 1.96$

|                              | Versions of the model |                |                 |                 |                 |                 |
|------------------------------|-----------------------|----------------|-----------------|-----------------|-----------------|-----------------|
|                              | 1                     | 2              | 3               | 4               | 5               | 6               |
|                              | NLS<br>1110 T.        | NNS<br>1180 T. | NLS<br>256 Sch. | NNS<br>258 Sch. | NLS<br>160 Sch. | NNS<br>157 Sch. |
| <b>Leadership effects on</b> |                       |                |                 |                 |                 |                 |
| Mediating variables:         |                       |                |                 |                 |                 |                 |
| Motivation                   | .62                   | .61            | .56             | .65             | .67             | .61             |
| Capacity                     | .52                   | .51            | .41             | .47             | .58             | .53             |
| Work setting                 | .68                   | .65            | .71             | .71             | .79             | .73             |
| Classroom practices          | .21                   | .18            | .17             | .12             | .23             | .19             |

T. = number of teachers; Sch. = number of schools.

Table 5. Effects on classroom practices significant standardized total effects,  $t > 1.96$ 

|                                 | Versions of the model |                     |                      |                      |                      |                      |
|---------------------------------|-----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|
|                                 | 1<br>NLS<br>1110 T.   | 2<br>NNS<br>1180 T. | 3<br>NLS<br>256 Sch. | 4<br>NNS<br>258 Sch. | 5<br>NLS<br>160 Sch. | 6<br>NNS<br>157 Sch. |
| <b>Motivation effects on:</b>   |                       |                     |                      |                      |                      |                      |
| Classroom practices             | .28                   | .23                 | .40                  | .32                  | .53                  | .34                  |
| <b>Capacity effects on:</b>     |                       |                     |                      |                      |                      |                      |
| Classroom practices             | .60                   | .54                 | .59                  | .70                  | .76                  | .83                  |
| <b>Work setting effects on:</b> |                       |                     |                      |                      |                      |                      |
| Classroom practices             | ns                    | ns                  | ns                   | .22                  | ns                   | ns                   |

allow hierarchical analysis, the tests summarized in these versions of the model do estimate effects at the lowest (teacher) level.

- Columns 3 (NLS) and 4 (NNS) report use of school-level data to test the model including both classroom practices and pupil performance. Survey data used in the analyses summarized in these figures include all teacher survey respondents.
- Columns 5 (NLS) and 6 (NNS) replicate analyses reported in columns 3 and 4 of the model but using survey responses from only those schools with three or more teacher respondents. The literacy sample had a mean of 5.7 teacher respondents and the numeracy sample had 5.9 respondents.

The models as a whole explain a moderate and significant proportion of the variation in altered teacher practices across schools; in the six versions of the model tested, this was 26%, 20%, 29%, 29%, 39%, and 35%, respectively. Table 4 summarizes the total effects of leadership on the three mediating variables in the model. In all six analyses, leadership has the strongest effect on work setting (.65 to .79), followed by motivation (.56 to .67) and capacity (.41 to .58). Although Version 3 of the model shows significant direct effects on changed classroom practices, leadership effects on such practices are primarily through the mediating variables and range from .12 to .23. So while overall levels of transformational leadership in the school were perceived by teachers to be low, variations in levels of such leadership explained a significant proportion of the variation in teacher's changed classroom practices.

The effects of mediating variables alone on changed classroom practices are reported in Table 5. Total effects on teacher practices are greatest for capacity (.54 to .83), with motivation having weaker effects (.23 to .53). Situation or work setting has weak total effects (.22) in one analysis and no significant effect in the other cases.

## Conclusion

Using data from a larger 4-year evaluation of England's National Literacy and Numeracy Strategies (Earl et al., 2003), this study aimed to test the effects of a

potentially distributed, school-specific, model of transformational leadership on teachers (motivation, capacities, and work settings), their classroom practices, and gains in student achievement. Some 2,290 teachers from 655 primary schools responded to two forms of a survey (literacy and numeracy) measuring all variables in our framework. Our measure of student achievement was gains in the British government's own Key Stage 2 tests over either 2 (numeracy) or 3 (literacy) years. Path analytic techniques (LISREL) were used to analyze the several different versions of the results.

Two methodological limitations of the research warrant caution in the interpretation of results—teacher response rates and the measure of student achievement. While response rates for both literacy and numeracy sample *schools* seem reasonable by most standards (51% and 52%, respectively), *teacher* response rates hovered in the 21% to 30% range, depending on whether all schools were considered, or only schools in which two or more teachers responded. Even though these rates are likely significant underestimations, much higher rates would give us more confidence that the sample was nationally representative. Nevertheless, we believe the teacher sample does not fatally undermine the purposes of our study for four reasons.

First, while the response rate was low, the size of the sample is significantly larger than samples used in most original educational leadership research. Second, it is unusual for educational leadership research to be based on national samples, but other comparators we were able to locate did not achieve better response rates and most were lower.<sup>4</sup> Third, a comparison of our achieved sample with the total population of primary teachers in England indicated differences of only 2% or 3% in proportion of male and female teachers. With respect to years of teaching experience, when the sample and total population are split between 0–10, 11–20, and 20+ years experience, they are very similar with 40%, 49%, and 46% in the younger group and 54% in the older group for the total population, the NLS sample, and the NNS sample, respectively; none of the groups is skewed by age. Correlational analysis indicated that years experience was weakly related to two variables: Less experienced literacy and numeracy teachers felt more capable ( $r = .08$  and  $.09$ ) and more emotionally engaged ( $r = -.08$  and  $-.10$ ); less experienced literacy teachers claimed to have changed their practices more than their more experienced peers.

Finally, the purposes for this study, unlike our larger study, do not depend on representing the views of teachers across the whole country. It is sufficient to know that respondents shared a number of relevant characteristics with a great many other English primary teachers.

A second methodological limitation of the study was our measure of student achievement gain scores over either 1 (numeracy) or 2 years (literacy). Linn (2003) has argued that gain scores over only a 2-year period may be unstable for a variety of reasons and recommends tracking gains over 3 or more years. We agree with Linn on this point and urge caution in interpreting evidence in our study related to the numeracy performance of students. In light of this caution, we re-ran all of our

analyses using, instead of gain scores, mean annual achievement scores, but found essentially no differences in results.

Three key sets of results were evident in our multiple tests of the effects of transformational leadership practices. First, transformational leadership had very strong direct effects on teachers' work settings and motivation with weaker but still significant effects on teachers' capacities. The size of these effects was approximately the same for either NNS or NLS. Second, transformational leadership had a moderate and significant effect on teachers' classroom practices. This effect was not as strong as either teacher capacity (the strongest effect) or teacher motivation but it was substantially stronger than teachers' work settings. Leadership, along with teacher motivation, capacity, and work setting explained approximately 25% to 35% of the variation in teachers' classroom practices. But our model, as a whole, failed to explain any of the variation in student achievement gains. These results have important implications for research and policy.

The evidence provided in this study (acknowledging the limitations outlined above) leads us to three straightforward yet important conclusions. First, school leadership has an important influence on the likelihood that teachers will change their classroom practices. Second, transformational approaches to school leadership seem to hold considerable promise for this purpose. Third, there is a significant gulf between classroom practices that are "changed" and practices that actually lead to greater pupil learning; the potency of leadership for increasing student learning hinges on the specific classroom practices which leaders stimulate, encourage, and promote.

In addition to these conclusions, we flag two implications from our study for future research. First, such research would do well to inquire, more precisely than this study was able to, about sources of transformational leadership practices. Our leadership model suggests that such practices may be widely distributed throughout the school and beyond, but our measures provided little information about the extent and sources of such distribution. Evidence from one of our recent studies in the same policy context suggests that the distribution of transformational practices is common but raises questions about which sources are best suited to the delivery of which practices (Leithwood et al., 2004). Research of this sort has important contributions to make to practice since "distributed leadership" (e.g., Gronn, 2002) has become such a popular idea in school systems, in spite of the extremely thin body of empirical evidence about its effects.

A second implication for future research concerns alternative designs for large-scale empirical studies of leader effects. They should be designed differently in order to better estimate the size of such effects. This study, quite similar in design to other large-scale empirical studies of the past, systematically underestimates leader effects because it invariably tests only relatively weak forms of leadership. This is the case because the data are collected from samples of leaders who range from highly successful to mediocre, at best. On average, ratings of leadership effects, as in our study, are usually relatively low. So designs such as this one do not test the effects of strong leadership. We need large-scale studies of strong leadership or empirical

studies of “outliers” (exceptionally strong as compared with relatively weak leadership).<sup>5</sup>

Results of our study, finally, also have several implications for educational reform policy. We stray a bit from our evidence here but were prompted to do so by the significant leader effects on classroom practices but lack of effects on student achievement. When governments prescribe preferred local action in some detail, as in the case of NLNS, and provide serious incentives and sanctions for compliance, the responsibility and accountability for student achievement ought to look very different than is typically the case. In such highly prescribed policy environments, governments typically hold school leaders accountable not only for compliance (in the case of this study, implementing the literacy and numeracy hours) but also for improving student achievement by, for example, the publication of league tables and the labeling of schools as failing.

This is clearly an instance of “one-way” accountability for governments who have gotten away with it for far too long. Indeed, adopting Wagner’s (1989) reasoning, it is highly unethical, since local leaders are being held accountable for things over which they have, at best, only partial control. In the context of our study, local leaders had some control over the implementation of prescribed classroom practices but not over whether the practices actually improved student achievement. If policy-makers insist on detailed prescription of local practices, then local leaders should be held accountable for implementing those practices with fidelity and policy-makers should be held accountable for the effects of those practices on students. Policy-makers unwilling to be held publicly accountable for the consequences of their policies should get out of the business of prescribing local practices and insisting on compliance.

More reciprocal accountability than is common is still called for on the part of policy-makers even if they do not prescribe local practices since, as a minimum, governments are responsible for providing the resources and conditions the best available evidence would suggest is necessary to meet the targets set for student achievement in many current policy contexts. Although not the case in England, many governments have increased standards for student achievement while cutting back on the resources to those in schools. And they have done this in the face of strong evidence that “less is definitely not more” (Molnar, 2002). Practices such as these, on the part of policy-makers, should not be tolerated by communities that value the education of their children; it is policy “carpet-bagging” at its worst.

The importance attributed to leadership in explaining school effects is without parallel among most policy-makers and many educators, at present. Enormous resources are being spent on improving leadership as one of the most direct and manageable ways for policy-makers to realize their aspirations for school reform on a large scale. But empirical evidence to help guide these efforts lags badly, potentially putting reform efforts in jeopardy. We desperately need more effort put into empirically unpacking the nature of successful leadership and describing the size and nature of its impacts on school organizations and students.



## Notes

1. This lack of attention to model specification is an important case of a larger problem plaguing educational leadership studies. Such failure to take account of significant conceptual differences threatens to undermine the special features of this form of leadership, reducing it, as is now the case for “instructional leadership,” to a slogan with a wide range of assumed and imputed meanings rather than a series of empirical claims about specific leadership practices and their independent and combined effects.
2. NFER has conducted research in England for over 55 years. Its membership includes local educational authorities (districts) in England and Wales, the main teachers’ associations in the UK, and many other agencies as well. At any given time it may be engaged in as many as 100 projects.
3. Heads’ response rates were 49% (literacy) and 50% (numeracy).
4. See, for example: Farkas, Johnson, and Duffett (2003) with a U.S. national response rate of 34% of superintendents and 21% of principals achieved samples of 1,006 and 925, respectively.
5. A preliminary analysis of evidence from this study, comparing data from schools rating transformational leadership highest and lowest (top and bottom quartiles), supports the promise of such a design. Measures of almost all variables were significantly different between these two sets of schools and there was some evidence of significant leader effects on student achievement.

## References

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: The Free Press.
- Bass, B. M. (1996). *A new paradigm of leadership: An inquiry into transformational leadership*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Berman, P., & McLaughlin, M. (1977). *Federal programs supporting educational change. Vol. VIII. Implementing and sustaining innovations*. Santa Monica, CA: Rand Corporation.
- Brown, K. M., Anfara, V. A., Hartman, K. J., Mahar, R. J., & Mills, R. (2002). Professional development of middle level principals: Pushing the reform forward. *Leadership and Policy in Schools, 1*(2), 107–143.
- Burns, J. M. (1978). *Leadership*. New York: Harper & Row.
- Cohen, D. K., & Barnes, C. A. (1993). Pedagogy and policy. In D. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding* (pp. 207–239). San Francisco: Jossey-Bass.
- Day, C., Harris, A., Hadfield, M., Tolley, H., & Beresford, J. (2000). *Leading schools in times of change*. Buckingham, UK: Open University Press.
- Earl, L., Watson, N., Levin, B., Leithwood, K., Fullan, M., Torrance, N., Jantzi, D., Mascall, B., & Volante, L. (2002). *Watching & learning 2: Report of the external evaluation of England’s National Literacy and Numeracy Strategies*. Toronto: Ontario Institute for Studies in Education, University of Toronto.
- Earl, L., Watson, N., Levin, B., Leithwood, K., Fullan, M., Torrance, N., Jantzi, D., Mascall, B., & Volante, L. (2003). *Watching & learning 3: Final report of the external evaluation of England’s National Literacy and Numeracy Strategies*. Toronto: Ontario Institute for Studies in Education, University of Toronto.
- Farkas, S., Johnson, J., & Duffett, A. (2003). *Rolling up their sleeves: Superintendents and principals talk about what’s needed to fix public schools*. New York: Public Agenda.
- Ford, M. A. (1992). *Motivating humans: Goals, emotions and personal agency beliefs*. Newbury Park, CA: Sage.
- Glass, G. V. (2002). Time for school: Its duration and allocation. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 79–93). Greenwich, CT: Information Age.

- Gronn, P. (2002). Distributed leadership. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration: Part one* (pp. 653–696). Dordrecht, The Netherlands: Kluwer.
- Hallinger, P. (2003). The emergence of school leadership development in an era of globalization: 1980–2002. In P. Hallinger (Ed.), *Reshaping the landscape of school leadership development: A global perspective* (pp. 3–22). Lisse, The Netherlands: Swets & Zeitlinger.
- Hallinger, P., & Heck, R. H. (1996a). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 723–783). Dordrecht, The Netherlands: Kluwer.
- Hallinger, P., & Heck, R. H. (1996b). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, 32(1), 5–44.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, 9, 157–191.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal*, 86(2), 217–247.
- Hoy, W. K., Sweetland, S. R., & Smith, P. A. (2002). Toward an organizational model of achievement in high schools. *Educational Administration Quarterly*, 38(1), 77–93.
- Hunt, J. G., & Conger, J. A. (1999). From where we sit: An assessment of transformational and charismatic leadership research. *The Leadership Quarterly*, 10(3), 335–343.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge: Massachusetts Institute for Technology.
- Hutchins, E., & Klausen, T. (1998). Distributed cognition in an airline cockpit. In Y. Engerstrom & D. Middleton (Eds.), *Cognition and communication at work* (pp. 15–34). Cambridge, UK: Cambridge University Press.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL: Structural Equation Modeling with SIMPLIS Command Language*. Chicago: Scientific Software, Inc.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage.
- Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112–129.
- Leithwood, K., & Jantzi, D. (in press). A review of transformational school leadership research: 1996–2005. *Leadership and Policy in Schools*.
- Leithwood, K., Jantzi, D., Earl, L., Fullan, M., Levin, B. (2004). Leadership for large-scale reform. *School Leadership and Management*, 24(1), 57–80.
- Leithwood, K., Jantzi, D., & Mascall, B. (2002). Large-scale reform: What works. *Journal of Educational Change*, 3(1), 7–33.
- Leithwood, K., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times*. Buckingham, UK: Open University Press.
- Leithwood, K., Leonard, L., & Sharratt, L. (1998). Conditions fostering organizational learning in schools. *Educational Administration Quarterly*, 34(2), 243–276.
- Leithwood, K., Tomlinson, D., & Genge, M. (1996). Transformational school leadership. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 785–840). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Lingard, B., & Christie, P. (2003). Leading theory: Bourdieu and the field of educational leadership: An introduction and overview to this special issue. *Leadership in Education*, 6(4), 317–334.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3–13.

- Lowe, K., Kroeck, K., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *Leadership Quarterly*, 7(3), 385–425.
- Molnar, A. (Ed). (2002). *School reform proposals: The research evidence*. Greenwich, CT: Information Age.
- Qualifications and Curriculum Authority. (2001). *Standards at key stage 2, English, mathematics and science: A report for headteachers, class teachers and assessment coordinators on the 2000 national curriculum assessments for 11-year-olds*. London: Author.
- Rowan, B. (1996). Standards as incentives for instructional reform. In S. H. Fuhrman & J. O'Day (Eds.), *Rewards and reform: Creating educational incentives that work* (pp. 195–225). San Francisco: Jossey-Bass.
- Salomon, G., & Perkins, D. (1997). Individual and social aspects of learning. *Review of Research in Education*, 23, 1–24.
- Sergiovanni, T. J. (2000). *The lifeworld of leadership: Creating culture, community, and personal meaning in our schools*. San Francisco: Jossey-Bass.
- Van den Berg, R., & Slegers, P. (1996). Building innovative capacity and leadership. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 653–696). Dordrecht, The Netherlands: Kluwer.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9(3), 257–280.
- Wagner, R. B. (1989). *Accountability in education: A philosophical inquiry*. New York: Routledge.
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement, a working paper*. Retrieved November 15, 2003 from McREL (Mid-continent Research for Education and Learning) Web site: <http://www.hub.mspnet.org/index.cfm/11589>
- Yukl, G. (1999). An evaluation of conceptual weakness in transformational and charismatic leadership theories. *The Leadership Quarterly*, 10(2), 285–305.

Copyright of School Effectiveness & School Improvement is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.