

The Leadership of Teaching and Learning: Implications for Teacher Evaluation

CLAIRE E. L. SINNEMA and VIVIANE M. J. ROBINSON

The University of Auckland, New Zealand

This article reports on a series of empirical studies that investigated the extent to which teacher evaluation policies and procedures promote teachers' inquiry into the relationship between their teaching and their students' learning. Four possible explanations for the weak focus on student learning in teacher evaluation are discussed. They relate to the alignment between school evaluation practice and a national policy which gives minimal emphasis to student outcomes, to the lack of an inquiry stance, to the compliance-driven approach taken to teacher evaluation, and to the overemphasis on norms of professional collegiality. We argue for an alternative approach to teacher evaluation that promotes teachers' capacity to inquire into and strengthen the relationship between their teaching and their students' learning in order to achieve the key purpose of evaluation—to improve teaching and learning.

INTRODUCTION

While teacher evaluation policies typically espouse the improvement of teaching and learning, there is little evidence they have been successful in this regard (Davis, Ellett, & Annunziata, 2002; Ellett & Teddlie, 2003; Gunter, 2002; Kleinhenz & Ingvarson, 2004). Approaches to teacher evaluation are typically focused on either teacher “inputs” (teaching style and behaviors) or on the “outputs” of teaching (Danielson & McGreal, 2000). We argue that

We are grateful to the participating schools and teachers for allowing us to conduct these studies on teacher evaluation. In addition, we thank Graeme Aitken for his assistance with the section of effective teaching.

Address correspondence to Claire E. L. Sinnema, The Faculty of Education, The University of Auckland, Private Bag 92601, Auckland 1023, New Zealand. E-mail: c.sinnema@auckland.ac.nz

if teacher evaluation is to achieve its intended purpose of improving teaching and learning an alternative approach is needed—one that promotes teachers' capacity to inquire into and strengthen the relationship between their teaching (inputs) and their students' learning (outputs). This paper reports on a series of studies that investigated the extent to which teacher evaluation policies and procedures promote teachers' inquiry into the relationship between their teaching and their students' learning. Three linked empirical studies conducted in New Zealand elementary and middle schools addressed the following questions: (a) To what extent do the tools (e.g., policies, and indicators) that are used in teacher evaluation *focus* on the impact of teaching on student learning? (b) To what extent are those tools *used* in ways that investigate and strengthen the impact of teaching on student learning? (c) What factors within the policy environment and school and teacher culture *explain the degree of inquiry* into the impact of teaching on learning? Four possible explanations for the weak focus on student learning in teacher evaluation are discussed. They relate to the alignment between school evaluation practice and a national policy which gives minimal emphasis to student outcomes, to the lack of an inquiry stance, to the compliance-driven approach taken to teacher evaluation, and to the overemphasis on norms of professional collegiality.

There is an emerging consensus that educational leadership research and practice needs a stronger emphasis on the leadership of teaching and learning (Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004, September; Prestine & Nelson, 2005). In their introduction to the report of an AERA Taskforce on "New Directions for Research on Educational Leadership," Firestone and Riehl (2005) give several reasons why this is important. First, if research on leadership is to contribute to leadership practice, it needs to address the tasks and responsibilities that the community and politicians now expect of educational leaders. According to Firestone and Riehl, "In the past, educational leaders were judged routinely on their effectiveness in managing fiscal, organizational and political conditions in their schools and school systems. In essence they were expected simply to set the stage for student learning. Now leaders are increasingly being held accountable for the actual performance of those under their charge" (pp. 3–4). Second, the authors argue that a stronger focus on the leadership of teaching and learning is required because public expectations about what all students should learn have risen substantially. Factual recall of information and surface-level understanding of concepts and problems are no longer sufficient as the outcomes of instruction. Critical thinking and inquiry, problem solving, and deep understanding of complex ideas are the new curriculum targets. Educators need deeper understanding of content and of pedagogical strategies in order to achieve these ambitious goals with increasingly diverse groups of learners.

Given the complex and largely indirect links between educational leadership and student learning and achievement, this goal presents a considerable

challenge (Hallinger & Heck, 1996a, 1996b). While it is clear that educational leaders exercise their influence by setting and shaping some of the conditions under which teachers work, we need to know much more about the particular features of those conditions that enable and disable effective teaching. For example, even though syntheses of leadership research have shown significant correlations between leaders' reported instructional leadership and student achievement, this information is too abstract to enable leaders to design evaluation and supervision policies and procedures that will make a difference to student achievement, or to evaluate those that are already in place (Hallinger, 2005; Hallinger & Heck, 1998; Witziers, Bosker, & Krüger, 2003).

In this paper we argue that school leaders should ensure that organizational processes, such as teacher evaluation, should be designed in ways that align with, and support, the goal of instructional improvement. This is important work because, in theory, every aspect of school organization should be enabling more effective and equitable teaching and learning. Yet, there is widespread acceptance that administrative processes in schools are only loosely coupled to processes of teaching and learning (Elmore, 2004). The call for a stronger leadership focus on instructional improvement should not, in the first instance, be met with a new layer of roles, responsibilities, and specially funded initiatives. A more appropriate initial response is a critical evaluation of the alignment between existing school organization and the conditions that are required for effective teaching. Such evaluation and redesign of school organization is essential for the sustainability of instructional improvement.

One of the organizational policies and processes that is intended to improve teaching and hence student learning is teacher evaluation. The potential of teacher evaluation as a tool for instructional improvement is great. Variation in teaching quality explains more of the variation in student achievement than any other school-based factor (Cuttance, 2000; Hattie, 1999; Reynolds, Creemers, Stringfield, Teddlie, & Schafer, 2002; New Zealand Ministry of Education, 2005). Given this strong link, it makes sense to see teacher evaluation as a powerful context in which to investigate and improve the quality of teaching. Furthermore, the improvement of teaching and learning is one of the espoused purposes of teacher evaluation policies (Ellett & Teddlie, 2003). Since conscientious teacher evaluation takes considerable time and effort on the part of both teachers and administrators, it is important to maximize its benefits for both teachers and students.

There are several reasons why the potential for teacher evaluation to deliver improved instruction has not been achieved (Davis et al., 2002; Ellett & Teddlie, 2003; Kleinhenz & Ingvarson, 2004). In the United States, both summative and formative teacher evaluation is surrounded by legal complexity and risk that produces defensive behavior on the part of both teachers and administrators. One indicator of such defensiveness is the fact that

both parties acknowledge that the number of teachers receiving less than satisfactory reports is far less than the number of unsatisfactory teachers (Pajak & Arrington, 2004). A second reason for the limited impact of teacher evaluation is ongoing concern, on the part of both administrators and teachers, about the validity of the criteria of effective teaching that are incorporated in many of the tools used to evaluate teachers. In the minds of many teachers and administrators, there is little connection between scores on teacher evaluation reports and teaching quality. Such concerns need to be addressed if the time-intensive administrative processes that surround teacher evaluation are to serve instructional improvement rather than administrative compliance.

We share these concerns about the validity of teacher evaluation and conclude from our data that there is a significant misalignment between the content of teacher evaluation policies and procedures and the nature of effective teaching. There may also be, as many authors have argued, a misalignment between some of the processes of teacher evaluation and the processes that teachers need to be reflective about to improve their practice. For example, Down, Chadbourne, and Hogan (2000) and Gunter (2002) argue that teacher evaluation that is linked to high-stakes employment decisions is not conducive to teacher learning. It is the content of teacher evaluation, however, rather than the process by which it occurs, that we are concerned with in this paper.

Educational leaders who wish to ensure more instructional payoff from teacher evaluation systems need to formulate an account of effective teaching that they can use as a yardstick against which to judge the adequacy of teacher evaluation systems. Even if educational leaders delegate this task to evaluation specialists, they need sufficient knowledge of the field to be able to critically evaluate and contribute to their work. We begin by arguing for a particular account of effective teaching and showing how it provides a framework for the evaluation of teacher evaluation policies and procedures. We then use this framework to investigate the extent to which teacher evaluation policies and procedures are aligned to the goal of instructional improvement. The investigation comprises three linked studies of teacher evaluation in a sample of urban elementary schools.

ALTERNATIVE APPROACHES TO EVALUATING TEACHING

In addressing the question of what is effective teaching, we briefly review three approaches that could be used to evaluate teaching. The first approach, which uses student outcomes as a basis of teacher evaluation, is flawed because of the difficulty of attributing student achievement and learning to the efforts of a particular teacher or teachers. The second approach, which evaluates teachers on the basis of their conformity to a

preferred style of teaching, is also flawed because, even if the preferred aspects of teacher style are based on research evidence, such evidence cannot be uncritically applied to any given teaching situation. The third approach, and the one we favor, evaluates teachers on the basis of their ability to inquire into and strengthen the relationship between their own teaching and the learning and achievement of their students.

Using Student Outcomes to Evaluate Teaching

The view that teaching effectiveness should be determined by what students achieve has popular appeal and a simple logic. The assumption that effective teachers get better results than less effective ones is reflected in the current push in the United Kingdom (Farrell & Morris, 2004) and United States (Pajak & Arrington, 2004) to hold teachers and school leaders accountable for results through performance pay or other contractual mechanisms.

The first problem with this approach to teacher evaluation is that differences in students' achievement cannot be directly attributed to differences between their teachers. The best predictor of what a student learns in any teacher's class is what they already know when they start the year. As David Ausubel concluded nearly thirty years ago:

If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly. (Ausubel, 1968, p. 18)

In a twenty-year research program based on intensive monitoring of the links between teaching and what students learned and remembered, Nuthall and Alton-Lee (1993) drew a similar conclusion about the impact of prior knowledge on what pupils learn from their classroom experience. The cumulative and path-dependent nature of student achievement means that evaluations of teaching quality based on single-point comparisons of students' achievement are invalid.

The measurement of "added value" provides a much more sophisticated version of the outcomes approach to teacher evaluation. The outcome of interest in this case is not student achievement but student learning—i.e., the difference between their achievement at two points in time. The difficulty here is that the measurement of learning is a complex business. Large data bases, a sophisticated hardware and software infrastructure, and skilled teams of educational practitioners, scientists, and mathematicians are required to collect, analyze, and report the data in ways that are useful for teachers and administrators. The Curriculum Education and Management Centre at the University of Durham in the United Kingdom provides such a

service to about 4000 English primary schools and one in three English secondary schools. The center puts considerable effort into ensuring that the data are used to build a research culture in schools and not an “aggressive management culture, including target setting and performance related pay” (Fitz-Gibbon & Tymms, 2002). Despite the high quality of the value-added evidence they produce for subscribing schools, they reject its use for high-stakes teacher evaluation purposes because it does not address the crucial question of the cause of the added value. If the students’ learning cannot be unequivocally attributed to their teacher, then it cannot be used as a basis for judgments about effective teaching.

Using Aspects of Teaching Style to Evaluate Teaching

The logic that underlies the style-based approach to teaching effectiveness rests on the claim that we have reliable knowledge of the characteristics of teaching and teachers that have positive impacts on students. The source of such knowledge may be the experience and practical wisdom of teachers themselves, or process-product research identifying how particular qualities of teachers or teaching impacts students. These qualities can include teaching approach (e.g., teachers are evaluated on their use of direct instruction or cooperative learning), personal attributes (e.g., the use of humor, or displaying warmth) or particular instructional strategies (e.g., the use of advance organizers or praise).

Despite the fact that rigorous research may show the benefits of such qualities, their use as indicators of teaching effectiveness in teacher evaluation is flawed. The central problem is the assumption that a research generalization, no matter how reliable, is applicable in any given setting (Scriven, 1989). The style-based approach to teacher evaluation implies one right way to teach that ignores the impact of context on teaching. Evaluators who use observational checklists, no matter how well researched, presume to know in advance of any context-specific inquiry how a particular teaching episode will affect students. As David Berliner points out, the fundamental flaw in this approach is that teachers can be judged to be good if they model the desired practices, irrespective of whether the students learn (Berliner, 1987). As long as teacher evaluation tools focus on what the teacher is doing rather than on how particular students are experiencing and benefiting from the teaching, teacher evaluation will be treated by many as an administrative compliance task which has little to offer teachers about how they might improve their impact on the learning and achievement of their students.

Our rejection of style-based approaches to teacher evaluation is not a rejection of the research activity that is now delivering robust, consistent, and useable generalizations about effective teaching (Cuttance, 2000; Darling-Hammond & Bransford, 2005; Richardson & Placier, 2001). Indeed, the convergence of findings from recent syntheses of evidence suggests that

research on teaching is now at a point where the educational community can claim that the “art of teaching is rapidly becoming the science of teaching” (Marzano, Pickering, & Pollock, 2001, p. 265).

Educational leaders and teachers should be familiar with such research because it provides fruitful hypotheses about how to teach in ways that are likely to be effective. Our point is that, in any given context, these generalizations must be treated as hypotheses, which may or may not turn out to be true in any given context. The increasing diversity of the student body and the complexity of the teaching-learning interface mean that effective teaching cannot be sustained by enforcing prescribed styles and approaches. The widespread problem with teachers’ cynical compliance with evaluation is partly due to their correct perception that the assessment rubrics represent their leaders’ preferred style of teaching, rather than what is necessarily effective with their particular students.

Using Inquiry into the Teaching-Achievement Relationship to Evaluate Effective Teaching

The problem with both the outcome and the style-based approach to teacher evaluation is that they assume, rather than investigate, the relationship between teaching and student learning. Under the outcome approach, it is assumed that the teacher being evaluated is the cause of student learning. Under the style approach, it is assumed that the preferred teaching qualities cause learning in this particular class with these particular students. The alternative approach we advocate is one that evaluates and promotes teachers’ capacity to inquire into and strengthen the relationship between their teaching and what their students learn. There is now growing empirical support for the view that effective teachers inquire into, rather than take for granted, the relationship between what they do (style) and its impact on students (outcomes). In addition, they take action as a result of their investigation to improve student outcomes and continue to inquire into the result of their interventions (Earl & Katz, 2002; Stoll, Fink, & Earl, 2003; Timperley, 2005).

It is desirable, of course, that teachers’ actions are informed by a knowledge of research generalizations about effective teaching, because such knowledge makes it more *likely* that they will be effective. But as indicated earlier the applicability of such knowledge must be tested in each particular context. Thus, effective teaching is more than style and more than outcomes—it is the continual interrogation of the relationship between these two dimensions with the aim of enhancing the achievement of all students. Such a model implies particular attitudes or dispositions (open mindedness, flexibility) and particular actions (questioning students about what they are understanding) but it does not prescribe or checklist such qualities. It simply prescribes inquiry, action, and improvement. In this sense it is

consistent with Scriven's observation that "Teachers can teach however they like, as long as it is ethical and effective in imparting valuable learning, within applicable curriculum and resource constraints" (Scriven, 1994, p. 157).

Under this third approach, teacher evaluation is concerned with teachers' ability to inquire into the impact of their own teaching and to invent or adapt strategies that will help all their students to succeed. Robust inquiry into the teaching-learning relationship requires evidence about what the teacher's students have learned or achieved. Such evidence could comprise anything from disaggregated standardized tests or exam results, to students' assignments, to worksheets assessing a unit of work, or how students' talk about their understanding of what they are supposed to be learning. Whether teacher evaluation has a summative or formative purpose, this framework judges teachers' capacity and willingness to understand and use evidence from their own students to change their teaching in ways that help them be more successful. This inquiry approach to teacher evaluation provided the normative framework for the empirical studies reported in the remainder of this paper. The questions these studies were designed to answer were:

1. To what extent do the tools (e.g., policies, and indicators) that are used in teacher evaluation focus on the impact of teaching on student learning?
2. To what extent are those tools used (e.g., in the development of teachers' goals and in evaluation discussions) in ways that investigate and strengthen the impact of teaching on student learning?
3. What factors explain the degree of inquiry into the impact of teaching on learning?

Research Context

The above questions were addressed through three linked empirical studies of teacher evaluation practices in 17 New Zealand elementary schools. Under New Zealand's radical self-management policy, teacher evaluation is the responsibility of every self-managing school, constrained only by a broadly specified national policy framework. The overall purpose of the national policy is to "improve the quality of teaching (and therefore learning)" (New Zealand Ministry of Education, 1997), a purpose that is shared with many United States evaluation systems (Danielson & McGreal, 2000; Davis et al., 2002; Ellett & Teddlie, 2003; Miami-Dade County Public Schools, 2002). The national policy provides the framework under which every school develops its own teacher evaluation policy. That framework guides both the processes to be used in teacher evaluation and the criteria against which teaching and teacher effectiveness is to be assessed.

The national policy for performance management in New Zealand specifies that every teacher must be evaluated against a set of performance

expectations, which are aligned to the national professional standards and other agreed teaching, management, and schoolwide responsibilities. It further specifies that the evaluation must be conducted by an evaluator selected in consultation with the teacher, and that it must incorporate observation of teaching, teacher self-appraisal, discussion of the evidence, and collaborative setting of development goals. The evaluation report is confidential to the teacher, their evaluator, and the school principal. In short, New Zealand's evaluation policy and processes encourage a low-stakes formative approach.

Since 1999 New Zealand fully registered elementary teachers have been evaluated against seven professional standards which are incorporated in the national collective employment agreement. Linked to the seven standards are 24 sample performance indicators which schools may choose to incorporate into the tools they develop as part of their school's evaluation policy. None of the professional standards explicitly refers to student outcomes and only one of the accompanying 24 sample indicators, namely, "demonstrate a range of effective teaching techniques," has the potential to require data-based inquiry into student learning. The sample indicators are, rather, almost entirely style based.

Research Approach

Figure 1 outlines the mixed methods approach taken in a sequence of three empirical studies which addressed the research questions outlined earlier.

The approach is described as "mixed methods" because it involved "both qualitative and quantitative data collection and analyses techniques" (Tashakkori & Teddlie, 2003, p. 11). Mixed methods were particularly appropriate for this set of studies because it included both quantitative and qualitative questions. The quantitative questions (questions 1 and 2) required description of the extent of focus on teaching and learning. The qualitative question (question 3) required an explanation for the degree of focus. Data from three studies were used with at least two sources relevant to each question.

The first study employed document analysis to assess the extent to which each school's evaluation policy and its associated performance indicators were focused on inquiry into student learning. The second study used interviews to discover the kinds of topics that teachers reported discussing in their most recent evaluation, and the third study used a questionnaire to find out the type of goals that were agreed upon by teacher's and their evaluators. Additional methodological detail is included in the subsequent reports on each study.

All three data sources were submitted to both qualitative and quantitative analysis. While both the documents and semistructured interviews provided qualitative textual information, that information was quantified by

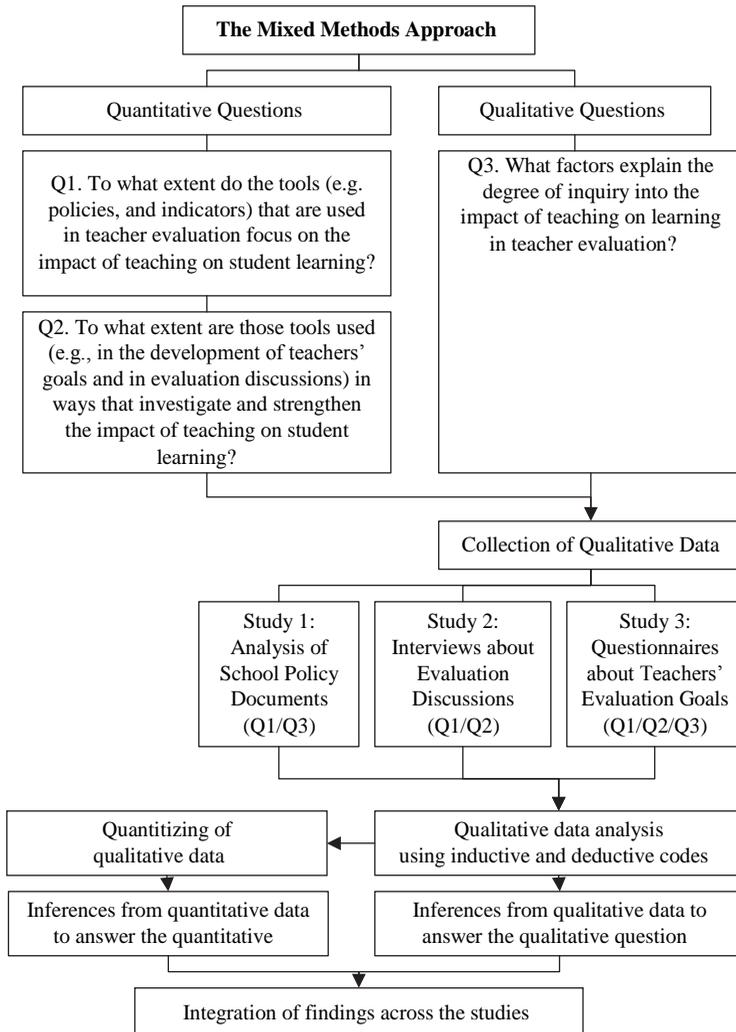


FIGURE 1 The Mixed Methods Approach Used in Three Studies of Teacher Evaluation.

counting the number of units that exemplified particular meaning codes. The codes, outlined with examples in the Appendix, were developed both deductively and inductively.

The deductive codes (about student learning, teaching, inquiry, and data) reflected the normative framework of effective teaching, which was explained earlier. By coding against these codes we were able to answer the first two research questions about the extent to which school-teacher evaluation practices and processes fostered inquiry into the teaching-learning relationship. In order to answer the third research question, however, we needed to step outside our own normative framework to discover what

other theory of effectiveness was informing a school's evaluation policies and practices. Inductively developed codes enabled us to identify and explain the focus of teacher evaluation in these schools, specifically whether or not it was focused on student learning. For example, in Study One important distinctions were apparent between indicators initially coded in the "student learning" category. Close analysis revealed a distinction between indicators that were *directly* focused on student learning, and those that were *indirectly* focused on student learning by describing a practice that was assumed to impact student learning. Hence the deductive code "student learning" was elaborated into the two inductive codes "direct inquiry" and "indirect inquiry". Similarly, in Study Two (about evaluation discussions) the deductive code "student learning" was elaborated into two inductive codes—"student learning specific" (to represent discussion that was specifically about the learning of a particular student, group, or class in relation to a specific learning outcome), and "student learning general."

STUDY ONE: DO EVALUATION TOOLS FOSTER INQUIRY INTO THE IMPACT OF TEACHING ON LEARNING?

The first study investigated the focus given to teaching and student learning in the tools of teacher evaluation. All 46 elementary and middle schools who participated in a university-schools consortium for the purpose of initial teacher education were invited to participate in this study by providing their school's teacher evaluation documents. Seventeen urban elementary and middle schools of varying size and socioeconomic status submitted their school's evaluation policy and performance indicators for analysis. The statements of purpose in each policy were identified and coded into three categories: teaching, student learning, and other. Definitions and coding rules were developed and a second coder independently coded all of the documents. Brief definitions and examples of these categories are provided in the Appendix. Intercoder agreement was 100% for the student learning category, and 89% for the teaching category. Of 119 intention statements included in these policy documents, 19 statements (15%) made some reference to student learning. In contrast, 76 statement (70%) made reference to teaching, with the remaining 42 statement referring only to other purposes.

The indicators that schools had included in their policies (to reflect national professional standards) were coded into three categories: direct inquiry into student learning, indirect inquiry into student learning, and other (see the Appendix for examples).

The distinction between "direct inquiry" and "indirect inquiry" into student learning is important since direct indicators are those deemed to direct inquiry into the actual impact that teaching practice has on student achievement. Indirect indicators, in contrast, are expressed in a manner which

assumes that a particular practice will lead to improved student achievement. An example is the indirect indicator “demonstrates an attractive, busy, and challenging physical environment that promotes student achievement and further learning across the curriculum.” This indicator primarily encourages inquiry into the teacher’s classroom environment and suggests certainty about the impact a particular environment will have on student achievement. A classroom environment that is attractive, busy, or challenging, however, will not necessarily promote student learning. Those things could, in fact, impact negatively on student learning for those children who prefer a plain, calm, and quiet, versus attractive and busy environment. The certainty implied by use of the term “that” in the indicator is misguided, since the particulars of any given teaching/learning context will impact on how such environments influence students. Such indirect indicators reinforce, rather than challenge, assumptions about the impact of teaching on particular students.

The “other” category encompassed those indicators that focused purely on teacher activities and behaviors with no explicit connection to student learning. Examples of such indicators include whether the “cloakroom area is tidy” or whether the teacher is demonstrating “professional dress standards.” Coding of the performance indicators by the second coder revealed more than 99 percent agreement for the judgments of the 503 indicators.

Analysis of the indicators revealed that although the average school included 46 indicators in its policies, only 3% were directly focused on student learning, with a further 3% providing an indirect focus on learning (see Table 1).

These findings point to two key problems. The first problem is that the schools studied use indirect indicators as often as they use direct indicators of student learning. The second problem is that the vast majority of indicators (nearly 94%) focused on matters other than student learning. Two-thirds of the schools studied only had indicators in the “other” category.

In summary, Study One showed that while improved student learning was an espoused purpose of most schools’ teacher evaluation policies, the tools that they developed were not aligned to this purpose. Instead, the tools supported evaluation of specific teacher behaviors. The underlying assumption appeared to be that performance of the specified teaching behaviors would guarantee improvements for students.

TABLE 1 Focus of Schools’ Performance Indicators.

| | Indicators (<i>N</i> = 503) | |
|--|------------------------------|-----|
| | Total | % |
| Direct inquiry into student learning | 15 | 3% |
| Indirect inquiry into student learning | 16 | 3% |
| Inquiry into other | 472 | 93% |

STUDY TWO: DO EVALUATION DISCUSSIONS FOSTER INQUIRY INTO THE IMPACT OF TEACHING ON LEARNING?

Leadership of teacher evaluation is not only evident in the development of the associated tools, it is also apparent in the way those tools are used in interaction with teachers (Halverson, Kelley, & Kimball, 2004). While the analysis of the tools themselves is suggestive of how they might be used, further data were required to find out what was emphasized in the actual evaluation discussions. Eleven teachers in three of the schools involved in Study One were interviewed to find out what was emphasized in their evaluation discussions. The principals of these schools had indicated interest in a follow-up study and had sought teacher volunteers. The teachers were interviewed about their most recent evaluation (held within the previous six months), and whether their role had been that of evaluator (four teachers) or the teacher being evaluated (seven teachers). The hour-long interviews began with open-ended questions about the topics that had been discussed and then probed directly for the degree of emphasis on student learning. The final interview questions probed teachers' attitudes towards inclusion of dialogue about the impact of their practice on student learning in their evaluation discussion.

The topics teachers reported to have been addressed in their evaluation discussion were coded into the four categories shown in Figure 2.

While all 11 teachers reported a discussion of their teaching, only one teacher (an evaluator) reported discussing the relationship between a particular aspect of teaching and student learning. Since the teacher that evaluator was discussing was also a participant (they were the only evaluator-evaluated pair included in the sample), we were able to compare their reports. The evaluator reported that their discussion inquired into the connection between student learning outcomes and the teacher's practice:

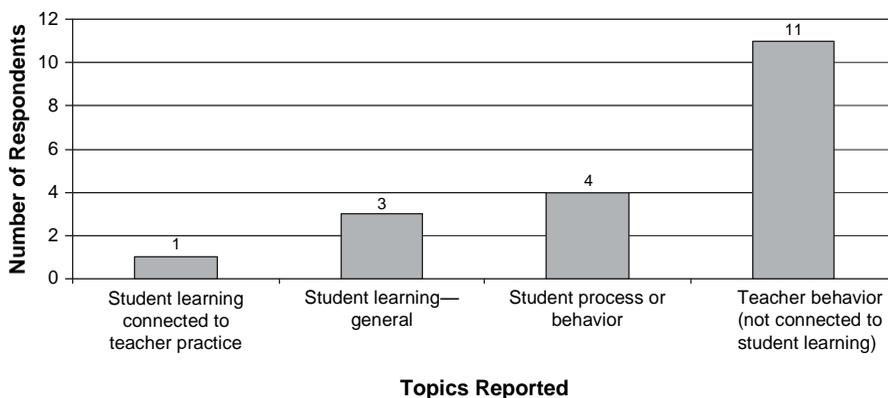


FIGURE 2 Teachers' Reports of Evaluation Discussion Topics.

It was very clear that from some of the strategies she used, more learners were accessing the thinking and the learning, and in other instances, not so many were. So, as we talked about that, which instances did it work well, which instances wasn't it working. We very clearly identified some of the kids who were very clearly like, getting it, and the kids who were very clearly not . . . and why that may have been. (Sinnema, 2005, p. 112)

The teacher who the evaluator was referring to, though, did not recall this same discussion. Rather, she recalled the positive character of the discussion and particularly the praise she was given:

There was a lot of praising so it was very, very positive. She looked at me in a very positive way, a very pleased way, and I felt really relaxed about that. (Sinnema, 2005, p. 113).

While three other teachers did report that their discussion included a general reference to student learning, these discussions were disconnected from teacher practice. Furthermore, they did not refer to the learning of a particular student or group in relation to a particular learning outcome. For instance, a teacher mentioned in her discussion that a lesson was "really pleasing . . . the art work came out how we wanted it to look." Another recalled mention of "improved numeracy skills" and that "children are making progress." The generality of these references to student learning limit their usefulness as a basis for inquiring into and strengthening teaching effects.

The predominant focus of appraisal discussions was, as reported by all 11 of the teachers, on teacher practices, such as activities they had carried out with their students and strategies or approaches they had adopted:

It [the evaluation interview] was basically surrounding a one-off lesson, the organization, the preparation, the different ways that we managed it. We just talked about the differences in our styles . . . a bit general. (Sinnema, 2005, p. 109)

Comments from teachers indicated three key explanations for the focus on teaching style in evaluation interviews. The first explanation is that particular types of teaching are assumed to connect to positive shifts in student outcomes. That assumption is so strong that it is not deemed necessary, in the context of teacher evaluation, to examine the actual connection between a teacher's practice and their students' learning. For example, one teacher reported that the evaluation discussion was about involvement in a professional development program and completion of planning and school-wide assessment requirements. When she was asked if there was any discussion about those practices in relation to student learning she replied: "No. It's interesting that you, you ask about children. I think they rather get

lost in our appraisal system.” She then responded to a question about the absence of discussion about student learning in this way:

If I am involved in the arts development program, and if my assessment practices are up to scratch and my planning is good, then does that mean that the student outcomes will be all right? I assume that’s the thinking behind it.

Another teacher explained that there was no inquiry into student learning because the development goal that her discussion focused on was not about student learning. Her goal was about planning for values exploration and social decision making in social studies. The discussion centered on the way in which her team of teachers investigated and implemented learning centers, graphic organizers, and cooperative learning. Her comments reveal the strength of taken-for-granted assumptions about the impact of particular teaching approaches and tasks:

. . . ‘cause this one [goal] is all about planning, and planning affects students’ learning. ‘Cause planning obviously, you know, it must, it must connect with the students’ learning ‘cause it’s part of that planning, learning, assessment, teaching cycle. (Sinnema, 2005, p. 11)

A second explanation for the focus on teaching style in evaluation interviews is that they are perceived as a summative, rather than as a formative, teacher learning opportunity. This was the case for five of the 11 participants in this study who suggested, for example:

I guess I’d never really thought of it as needing to be terribly helpful in terms of changing what you do in your classroom process.

It’s almost like a summative thing, it’s towards the end [of the year], so it wasn’t like an opportunity for him to go “Well have you tried this?” It wasn’t a bounce off session. It was more of a “let’s evaluate what you’ve done this year” type thing. (Sinnema, 2005, p. 117)

Perceptions of evaluation as summative can lead to a compliance-driven approach. One teacher saw the purpose of her interview as being to “. . . fill in the form [and not to be] helpful in terms of changing what you do in your classroom process” (Sinnema, 2005, p. 129). In some cases, the timing of the discussion was the reason why evaluation was perceived as unproductive:

Appraisal discussion happens too late to productively use data . . . in our appraisal cycle in term one it’s setting goals, and term two and three you do observation, then term four in the discussion you’re kind of wrapping it up and looking ahead. (Sinnema, 2005)

For these teachers, learning from evaluation is limited by perceiving it as summative, and as occurring too late in the year to be able to alter one's teaching.

The limiting influence of evaluation goals is the third explanation for an absence of focus on the teaching and learning relationship in evaluation discussions. Nine out of 11 teachers reported that evaluation goals determined the focus for the evaluation discussion. Since their goals did not focus on teaching and learning, the interview (based on those goals) did not have such a focus.

In the final phase of each interview, participants were asked to respond to the proposition that their evaluations might focus on inquiring into the teaching-learning relationship through the examination of data about student learning. The proposition received an overwhelmingly positive response from interviewees. Ten participants had a positive response to the suggestion, with seven of these having only positive things to say in relation to the idea. A typical reaction was "I think that's great . . . so you're using hard data of learning to then get the teacher to focus in on their teaching practice." Others said "I actually really enjoy getting that kind of feedback" and "I think it's a fair question to ask." The positive reactions were typically expressed with a sense of the idea being obvious, reflecting common sense, and with a tone of bewilderment that such a simple, yet important, idea is not common practice.

That'd be really good. I'd really like that. It would be useful 'cause that's what we're here for, 'cause that's what really matters. I think that's actually what it's all about. That's so fundamental. That's actually the guts of what I do. And also it's what I do really well. (Sinnema, 2005, p. 8)

Several of the other participants who had a positive response also showed genuine enthusiasm toward the notion of an evaluation interview in which the learning of their students was a focus. As they pondered the idea, their body language, expression, and questions revealed genuine enthusiasm:

. . . Well, that'd be good. It'd be really good. Like it'd give you really good feedback, wouldn't it?

. . . It's good because it'll mean that my teaching will become more effective for the children. And that's the whole idea, isn't it, really?

The suggestion also highlighted for one teacher a concern about the invisibility of children in their current approach to evaluation. She laughed at the irony of seeing a focus on student learning as novel: "it's interesting that you ask about children. I think they rather get lost in our appraisal system . . . To ask about students' learning, I mean it's so alien, it is so alien."

Only one of the 11 participant showed an exclusively negative response to the suggestion.

STUDY THREE: DO EVALUATION GOALS FOCUS ON TEACHERS' IMPACT ON STUDENT LEARNING?

Since Study Two revealed that the content of the evaluation discussion was largely determined by the evaluation goals that had been set at the start of the evaluation cycle, the third study gathered more systematic information about the nature of teachers' development goals. In this study, 68 teachers from eight schools responded to a questionnaire to establish the extent to which their goals focused on student learning. The content of each goal was coded into five categories to establish whether the goal reflected a focus on student outcomes, student behavior, teacher processes, inquiry, and data (see the Appendix for examples). Goals that incorporated more than one focus were multiply coded. Twenty-four of the 244 goals were coded by a second rater to establish the reliability of data analysis. The 24 goals were selected by highlighting every tenth goal on a spreadsheet. The second rater then coded those goals (by ticking the categories on the spreadsheet) based on coding rules (Sinnema, 2005, pp. 148–151) for the relevant categories. The intercoder agreement was calculated as 90%.

There is a strong body of evidence that suggests that “goals that are specific and difficult lead to a higher level of performance than vague, non-quantitative goals such as ‘do your best’” (Locke & Latham, 1990). Goals coded to the “student outcome” category (11) were analyzed further against coding rules for specificity and challenge (details of the coding of goal specificity and goal challenge are available in Sinnema, 2005, p. 149).

The Focus of Evaluation Goals

As shown in Table 2, less than five percent of goals (11 out of 244) were found to refer to student outcomes. The vast majority (90%) focused, instead, on elements of teachers' practice such as an aspect of their style, or a particular teaching approach. In this latter category were such goals as “implement new arts curriculum,” “support teacher aides with autistic student,” and “maintain a student-focused physical classroom

TABLE 2 Characteristics of Teachers' Evaluation Goal Statements.

| Goal Characteristics | Number of Goals (<i>N</i> = 244) | | Percentage Yes |
|--------------------------------------|-----------------------------------|-----|-------------------|
| | Yes | No | |
| Statements mention: student outcomes | 11 | 233 | 4.5% |
| student behaviors | 2 | 242 | 1.0% |
| teaching processes | 220 | 24 | 90.2% |
| Mention some form of inquiry | 9 | 235 | 3.7% |
| Mention the use of data | 0 | 244 | 0.0 |

environment.” Less than four percent of goals were coded as having an inquiry focus, and none specified the use of data.

Goal Specificity

Eleven of the goals, those that were focused on student outcomes, were also analyzed to establish how specific they were in stating which learners and aspect of learning were involved and the level of achievement desired. Approximately half of the goals were specific and half were general in stating which learners the goal related to. Specific reference to learners can be seen in the references, for example, to “gifted and talented children in my classroom,” and “lower-level students.” The way in which goals referred to the aspect of learning and the extent of improvement desired, though, was overwhelmingly general, with only one making specific reference to an aspect of learning. Of the other ten goals, two referred to learning generally, four referred to literacy, two to maths/numeracy, one to ICT, and one to questioning. While nine of the 11 goals analyzed here were about aspirations to “develop,” “up skill” and “improve” and “extend” children and their learning, not one made clear the desired *extent* of improvement. No specific reference was made to how much students might improve or how regularly they might, for example, demonstrate a skill. It was merely stated that they would “improve,” suggesting that any improvement, however small, was the goal. The extent of improvement teachers were aiming for was either not expressed at all or was ambiguous.

Goal Challenge

The eleven outcomes-focused goals were also analyzed to establish the degree of challenge they involved. Only one of the eleven goals about student outcomes also expressed *explicit* challenge for teaching and learning. The majority were limited because the outcomes they referred to were only *potentially* challenging, since even negligible improvement could represent achievement of the goal. For example, the goal to “develop literacy in grade 5” is undoubtedly a goal with potential to be challenging, if, for instance, it refers to a goal to increase the reading ages of all boys in the class (currently 80% of whom are reading at an age 1–2 years below their chronological age) to at least their chronological age, by the end of the year. Achieving this would certainly indicate that literacy for grade 5 students had been developed. However, the vagueness of the challenge in the “develop literacy in grade 5” goal means that even the slightest improvement for just one of the students could imply that the goal has been met. Unless appraisal goals clearly state a challenge, evaluation of the extent to which goals are achieved will not be possible.

DISCUSSION

These three linked studies, using a mixed methods approach to data collection and analysis, have confirmed that in these elementary and middle schools there is little focus on the teaching-learning relationship. Neither the evaluation tools themselves, nor the way they are used in teacher evaluation discussions, encourages inquiry into the impact of teaching on student learning. This is puzzling, given that the espoused intention of teacher evaluation in New Zealand, as in the United States and the United Kingdom, is to improve teaching and learning. Four possible explanations for the weak focus on student learning are discussed.

The first explanation points to the high degree of alignment between these schools' policies and evaluation tools and the national policy guidelines. While strongly espousing such purposes as improvement of teaching and student learning, the tools incorporated within national policy give very low priority to using student learning or achievement as a criterion of teaching effectiveness. Of the 24 sample indicators provided by the central Ministry of Education to illustrate how teachers would be evaluated against seven professional standards, none relate specifically to student learning (New Zealand Ministry of Education, 1998b). In addition, the Ministry issued a list with a further 78 sample performance indicators that schools could choose to incorporate into their own school policies (New Zealand Ministry of Education, 1998a). Only three of the 78 sample indicators refer to student achievement. The message given by the national policy itself is that direct inquiry into student learning is not needed because certain types of teaching have guaranteed positive impacts on students' learning and achievement. Other recent New Zealand policy initiatives relating, for instance, to teacher professional learning and school improvement emphasize the need for practitioners to focus on student learning and engage with achievement data. However, there is little coherence between those policies and teacher evaluation policy.

The second explanation for these findings is the assumption by both teachers and policymakers that we know what effective teaching looks like. Given this certainty, evaluators and teachers act as if what needs to be monitored is not what is happening for students but what teachers are doing. Many of our teachers were convinced that if they planned, arranged the classroom environment in certain ways, encouraged participation, or employed other particular strategies, then their students would inevitably be learning. Teacher evaluation has yet to make the shift from a behavioral performance orientation to a more cognitive and inquiry stance in which teachers are encouraged, indeed required, to examine their taken-for-granted beliefs about how their teaching is impacting their students (Stein & Spillane, 2005).

The third explanation for these findings is the perfunctory and compliance-oriented nature of teacher evaluation in these schools. A compliance orientation was evident in the predominance of vague rather than specific and challenging goals, and from interview comments that indicated a tendency to

focus on satisfying requirements. Most teachers are deeply committed to their students and wish to see them succeed, yet that motivation is not tapped by teacher evaluation processes that check teachers' performance of preferred practices rather than their impact on students. Our interview data showed that teachers would be much more motivated by a system that had a strong developmental focus on assessing and growing teachers' capacity to continuously evaluate and strengthen the impact of their instructional strategies.

The fourth explanation for the lack of focus on student learning is the strong belief among these teachers, and in the profession more widely, that evaluation is an opportunity to celebrate success and be supportive of colleagues (Timperley & Robinson, 2000). In theory, this belief does not preclude a focus on student learning. In a policy and school climate where the positive impacts of certain teaching styles are taken for granted, however, it seems threatening to confront such assumptions by asking for evidence that preferred instructional strategies are having the intended impact on students.

In conclusion, the leadership embedded in teacher evaluation policies and artifacts conveys the strong message that we know what teaching practices are effective and that the purpose of teacher evaluation is to monitor and develop teachers' capacity to produce them. This is not the type of leadership that will grow teacher and school cultures that meet the espoused goals of teacher evaluation systems to improve teaching and learning. Leadership practices at state, district and school levels are needed that confront false epistemological assumptions about the generalizability of teaching-learning relationships and that promote teachers' situated inquiry into the impact of their teaching on their own students. As Ball and Cohen put it, attention to the "particular" is especially important for learning in and from practice since "teaching occurs in particulars—particular students interacting with particular teachers over particular ideas in particular circumstances" (Ball & Cohen, 1999, p. 10). Teacher evaluation is an ideal opportunity for educators to give attention to, and learn about, their own "particular" context, so that their teaching, and ultimately their own students' learning, can be improved.

REFERENCES

- Ausubel, D. (1968). *Educational psychology: A cognitive view*. New York: Holt Rinehart and Winston.
- Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning professional: Handbook of policy and practice* (pp. 3–32). San Francisco, CA: Jossey Bass.
- Berliner, D. C. (1987). Simple views of effective teaching and a simple theory of classroom instruction. In N. L. Gage, D. C. Berliner, & B. Rosenshine (Eds.), *Talks to teachers: A festschrift for N. L. Gage* (1st ed., pp. xiv, 411). New York: Random House.
- Cuttance, P. (2000). *The impact of teaching on student learning. Australian College of Education yearbook 2000*. Canberra: Australian College of Education
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision & Curriculum Development.
- Darling-Hammond, L., & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.
- Davis, D. R., Ellett, C. D., & Annunziata, J. (2002). Teacher evaluation, leadership and learning organizations. *Journal of Personnel Evaluation in Education*, 16(4), 287–301.
- Down, B., Chadbourne, R., & Hogan, C. (2000). How are teachers managing performance management? *Asia-Pacific Journal of Teacher Education*, 28(3), 213–223.
- Earl, L., & Katz, S. (2002). Leading schools in a data rich world. In K. Leithwood, P. Hallinger, K. S. Louis, G. Furman-Brown, P. Gronn, B. Mulford, & K. Riley (Eds.), *Second international handbook of educational leadership and administration* (pp. 1003–1022). Dordrecht: Kluwer.
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101–128.
- Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Education Press.
- Farrell, C., & Morris, J. (2004). Resigned compliance. *Educational Management & Administration*, 32(1), 81–104.
- Firestone, W. A., & Riehl, C. (Eds.). (2005). *A new agenda: Directions for research on educational leadership*. New York: Teachers College Press.
- Fitz-Gibbon, C. T., & Tymms, P. (2002). Technical and ethical issues in indicator systems: Doing things right and doing wrong things. *Education Policy Analysis Archives*, 10(6). Retrieved from <http://epaa.asu.edu/epaa/v10n16/>
- Gunter, H. (2002). Teacher appraisal 1988–1998: A case study. *School Leadership & Management*, 22(1), 61–72.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, 4(3), 221–239.
- Hallinger, P., & Heck, R. H. (1996a). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In K. A. Leithwood, J. Chapman, P. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 723–783). Dordrecht: Kluwer Academic.
- 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, 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.
- Halverson, R., Kelley, C., & Kimball, S. (2004). *Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practices*. In C. Miskel & W. Hoy (Eds.), *Theory and research in educational administration* (Vol. 3, pp. 66–90). Greenwich, CT: Information Age Press.
- Hattie, J. (1999). *Influences on student learning*. Inaugural Lecture, University of Auckland, New Zealand.
- Kleinhenz, E., & Ingvarson, L. (2004). Teacher accountability in Australia: Current policies and practices and their relation to the improvement of teaching and learning. *Research Papers in Education*, 19(1), 31–49.
- Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004, September). How leadership influences student learning. Retrieved June, 2005, from <http://www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf>
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Marzano, R., Pickering, D., & Pollock, J. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Miami-Dade County Public Schools. (2002). Procedures for observation, assessment and instructional and school support/resource personnel (draft). Retrieved March, 2005, from <http://paces.dade.k12/fl.us>
- New Zealand Ministry of Education. (2005). *Impact of teaching and schools on variance in outcomes*. Wellington: New Zealand Ministry of Education.
- New Zealand Ministry of Education. (1998b). *Teacher performance management: Primary school teachers, primary school deputy/assistant principals. A resource for boards of trustees, principals and teachers*. Wellington: Government Printer.
- New Zealand Ministry of Education. (1998a). *Interim professional standards: Primary school deputy/assistant principals, primary school teachers*. Wellington: Ministry of Education.
- New Zealand Ministry of Education. (1997, 10 February supplement). Performance management systems: PMS1: Performance appraisal. *Education Gazette*.
- Nuthall, G., & Alton-Lee, A. (1993). Predicting learning from student experience of teaching: A theory of student knowledge construction in classrooms. *American Educational Research Journal*, 30(4), 799–840.
- Pajak, E., & Arrington, A. (2004). Empowering a profession: Rethinking the roles of administrative evaluation and instructional supervision in improving teacher quality. In M. A. Smylie & D. Miretzky (Eds.), *Developing the teacher workforce* (103rd yearbook of the National Society for the Study of Education), (pp. 228–253). Chicago: National Society for the Study of Education.
- Prestine, N. A., & Nelson, B. S. (2005). How can educational leaders support and promote teaching and learning? In W. A. Firestone & C. Riehl (Eds.), *A new agenda: Directions for research on educational leadership* (pp. 44–60). New York: Teachers College Press.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schafer, G. (2002). *World Class schools: International perspectives on school effectiveness*. London: Routledge.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 905–947). Washington, DC: American Educational Research Association.
- Scriven, M. (1989). The state of the art in teacher evaluation. In J. Lokan & P. McKenzie (Eds.), *Teacher appraisal: Issues and approaches* (vol. 28, pp. 92–136). Victoria: Australian Council for Education Research.

- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education* 8(2), 151–184.
- Sinnema, C. (2005). *Teacher appraisal: Missed opportunities for learning*. Doctor of Education Thesis, The University of Auckland.
- Stein, M. K., & Spillane, J. P. (2005). What can researchers on educational leadership learn from research on teaching? Building a bridge. In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research on educational leadership* (pp. 28–45). New York: Teachers College Press.
- Stoll, L., Fink, D., & Earl, L. (2003). *It's about learning (and it's about time): What's in it for schools?* London: Routledge Falmer.
- Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671–701). Thousand Oaks, CA: Sage Publications.
- Timperley, H. (2005). Instructional leadership challenges: The case of using student achievement information for instructional improvement. *Leadership and Policy in Schools*, 4, 3–22.
- Timperley, H. S., & Robinson, V. (2000). Workload and the professional culture of teachers. *Educational Management & Administration*, 28(1), 47–62.
- Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425.

APPENDIX

Coding Category Examples

| Coding Categories | References to | Examples |
|--|------------------------|---|
| Study One—Policy Intention Statements | | |
| Student Learning | Student learning | [To] attest to the achievement of effective teaching and learning |
| | Student success | [To] achieve improved outcomes for student |
| | Student achievement | [To] promote ongoing and improved teaching and learning |
| | Student outcomes | [To] affirm and improve the quality of teaching and learning outcomes |
| | Learning | [To] improve and extend professional skills |
| | Success | [To] promote the development of effective performance |
| Teaching | Achievement | [To] support the establishment of effective teaching programs |
| | Teaching | [To] affirm and improve the quality of teaching and learning outcomes |
| | Teaching effectiveness | |
| | Professionalism | |
| | Work | |
| | Effectiveness | |
| | Standards | |
| | Excellence | |
| | Achievement | |
| | Skills/abilities | |
| Performance | | |
| Work performance | | |
| Teaching performance | | |

(Continued)

Appendix (Continued)

| Coding Categories | References to | Examples |
|--|---|---|
| Other | Other than above | [To] clarify job-related expectations [To] create a climate for two-way communication |
| Study One—Indicators | | |
| Direct Inquiry into Student Learning | about actual: Learning Success Progress Outcomes Achievement Results | Demonstrates appropriate emphasis and successful learning in the areas of reading, writing, and mathematics Children are achieving success Achieves results |
| Indirect Inquiry into Student Learning | That express an assumption that a particular practice will: lead to student learning cause student learning be related to student learning promote student learning | Adapts own teaching approaches and techniques to maximize students' learning opportunities and achievements Encourages pupil progress by providing feedback Demonstrates an attractive, busy, and challenging physical environment that promotes student achievement and further learning across the curriculum |
| Other | Neither of the first two categories | Work dated in exercise books Cloakroom area is tidy Professional dress standards |
| Study Two—Appraisal Discussions | | |
| Student Learning—Specific | Learning Outcomes Achievement Results Success Improvement Impact (of a particular student, group, class, or year level in relation to a specific learning outcome) | "I was saying [to the appraisee] do you remember when you used the abacus here, and [student A] was sitting across the table from you and he couldn't do it, could he, whereas [student B] was right there with you, didn't need the material and said straight away 24 divided by 6" |
| Student Learning—General | Learning Outcomes Achievement Results Success Improvement generally | "I said I can definitely see the children are making progress" |
| Teacher Behavior | The activities a teacher does Strategies, approaches, techniques used by the teacher Teacher qualities | "It [the appraisal discussion] was basically surrounding a one-off lesson, the organization, the preparation, the different ways that we managed it. We just talked about the differences in our styles . . . a bit general" |

(Continued)

Appendix (Continued)

| Coding Categories | References to | Examples |
|---|---|---|
| Student Behavior | The engagement and attitude of students | “She [the appraiser] said that the children were all totally engaged in the maths activities . . . they were totally involved in their maths, and really enthusiastic about it” |
| Study Three—Content of Evaluation Goal Statements | | |
| Student (a particular individual or group) learning outcome | Improvement; increased achievement; learning gains; student learning; student success; student achievement; student outcomes; learning gains; improvement; achievement (in a specific aspect of learning) | To upskill literacy at year 1 and 2 level (reading/writing) |
| Student—behavioral | Student behavioral processes | Ensure ready student response to instructions |
| Teacher process | Teachers’ actions, activities, tasks, | make and use numeracy project resources throughout the year |
| Inquiry | See if; find out; check; ask; query; question; look at; examine; explore; analyse | [none] |
| Data | Data; evidence; assessment; data; results; scores; reading ages; percentages; assessment grades; achievement ratings; test results | [none] |