Literacy Professional Development Project: Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning

Report to the Ministry of Education

J. Parr, H. Timperley, P. Reddish, R. Jesson, R. Adams
School of Education, The University of Auckland
Literacy Professional Development Project:

Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning

Authors:
Judy Parr
Helen Timperley
Paul Reddish
Rebecca Jesson
Rebecca Adams

School of Education
The University of Auckland
Private Bag 92019

Prepared for:
Learning Media Ltd
Wellington
New Zealand
Acknowledgements

We have been privileged to collaborate with a highly committed team on the Literacy Professional Development Project: the Co-Directors Lyn Bareta and Carolyn English; Pam O’Connell, Lois Thompson; the Ministry of Education liaison Denise Arnerich, and the regional team leaders and facilitators.

We particularly value the manner in which the entire team embraced the notion of ongoing learning at all levels and the close productive relationship with the Ministry’s Literacy and Numeracy Strategy Division.

Sincere thanks are due to the schools that, in agreeing to become research schools, allowed us to be privy to their learning on the project. We are particularly grateful to those teachers whose classes we observed and whose students we talked with.

Finally, our thanks to the members of our hard working research team who are acknowledged in the authorship of this report.
# TABLE OF CONTENTS

Executive Summary ..............................................................................................................1

Chapter 1 Overview of research .................................................................................................5
   The Literature Guiding the Research .................................................................................5
   Overview of Method ........................................................................................................9
   References .....................................................................................................................20

PART A: MEASURING PROGRESS .............................................................................................23

Chapter 2 Summary of Trends in Student Achievement .................................................................23
   Method ...........................................................................................................................23
   Conclusions ..................................................................................................................46
   References .....................................................................................................................47

Chapter 3 Issues Associated with Tests of Student Achievement in Reading and the Use of These Data to Inform Decisions ..............................................................................49
   Use of Gain Scores to Measure Progress ........................................................................49
   The Tools Selected and Their Ability to Measure Progress .............................................51
   Table 3.1 Sample Size by Year ......................................................................................52
   References .....................................................................................................................67

PART B: TEACHER KNOWLEDGE, BELIEFS AND PRACTICES .....................................................69

Chapter 4 Examining the role of teacher pedagogical content knowledge in literacy .......................69
   Context .........................................................................................................................71
   Method .........................................................................................................................71
   Data Analysis ...............................................................................................................72
   Findings .......................................................................................................................74
   Discussion ....................................................................................................................79
   References .....................................................................................................................80

Chapter 5 What Are We Supposed To Be Learning? Students’ Understanding of Their Writing Lessons .........................................................................................................................83
   Goals and Feedback .....................................................................................................85
   Method .........................................................................................................................86
   The Lesson Observation ...............................................................................................87
   Results .........................................................................................................................88
   Conclusion ...................................................................................................................104
   References ...................................................................................................................106
Chapter 6  Teachers, Schools and Using Evidence: Considerations of Capacity ................................. 109
Method.................................................................................................................................................. 110
Findings .................................................................................................................................................. 114
Discussion............................................................................................................................................... 120
References............................................................................................................................................... 121

Chapter 7  TEACHER PERCEPTIONS OF THEIR ROLE AND LEARNING ........................................ 123
Perception of Influence on and Satisfaction with Student Achievement ......................................... 123
Self-reported Confidence re Dimensions of Pedagogical Content Knowledge .................................. 126
Teachers’ Attitudes towards Professional Development after One Year ........................................... 131
Conclusion ............................................................................................................................................... 133
References............................................................................................................................................... 133

PART C: FACILITATING TEACHER LEARNING ........................................................................... 135

Chapter 8  A Case Study: Promoting Teacher Learning Using a Needs-Analysis and Inquiry-based
Approach ............................................................................................................................................... 136
The School ............................................................................................................................................... 137
Discussion ............................................................................................................................................... 148
References............................................................................................................................................... 150

Chapter 9  Facilitation Challenges .................................................................................................... 153
Reflections Back on Learning .................................................................................................................. 153
Re-tracing some Aspects of the Learning Journey .............................................................................. 157
Developing Understandings of Facilitation .......................................................................................... 159
Conclusions............................................................................................................................................... 161
References............................................................................................................................................... 162

Chapter 10 Professional Learning Communities .............................................................................. 163
Method.................................................................................................................................................. 164
Results .................................................................................................................................................. 168
Discussion and Conclusions .................................................................................................................. 172
References............................................................................................................................................... 173

Chapter 11 An Analysis of Professional Learning Episodes .................................................................... 174
The Framework Guiding the Analysis ................................................................................................... 174
Method.................................................................................................................................................. 177
Results .................................................................................................................................................. 180
Conclusions ............................................................................................................................................... 195
References............................................................................................................................................... 197

Appendix A ............................................................................................................................................... 199
Appendix B ............................................................................................................................................... 203
Appendix C ............................................................................................................................................... 207
Appendix D ............................................................................................................................................... 209
Executive Summary

An outline of the research and of the context in terms of the broad themes in the international literature, together with an overview of the methods employed in the research, forms the introductory chapter.

Part A of the report is concerned with measuring progress. Chapter 2 concludes that students in the 13 research schools made large gains in writing and more modest gains in reading. In writing, there was a very large average effect size gain of 1.28 while, in reading, the effect sizes were more moderate with an average effect size of 0.46 in the first year of the project and 0.49 in the second year. While the average effect sizes are a little lower than those reported for the national cohort, the relativity between gain in reading and writing is similar.

With respect to differential achievement by ethnic group, in reading, there was a difference in performance of the major ethnic groups at both points in time. However, all groups made significant progress in terms of gains in total score and stanine level from Time 1 to Time 2 and there was no difference in extent of gain. In writing, the median score at Time 1 was barely different across ethnic groups. At Time 2 there was a significant difference between groups with both NZ European and Other significantly higher than Maori or Pasifika. However, this may be a function of a very small sample of schools as in a school that made excellent progress, all groups progressed equally.

There were no significant differences by gender in reading achievement scores in this sample or differences in rate of progress. In writing, however, females scored significantly better at both points in time but both girls and boys made significant progress and there was no difference in the rate of progress.

A discussion of problematic issues associated with the use of gain scores to measure progress is presented in Chapter 3. It concludes that, despite reservations in the somewhat dated literature, gain scores can be as reliable and valid as other measures. However, significant drawbacks associated with using STAR as an indicator of progress are identified. Sizeable ceiling effects are shown to occur. These appear at year 5 and are particularly marked at years 6 to 8. The reliability of difference scores for STAR appears low.

Part B of the report deals with teacher knowledge, beliefs and practices. The level and extent of growth in teacher knowledge and its relationship to student progress are examined in Chapters 4 and 6. The first of these, Chapter 4, presents a way of exploring teacher pedagogical content knowledge (PCK), that is, their knowledge of their subject (reading or writing) from the point of view of teaching it. While there are issues around measuring this knowledge (and these are discussed in the chapter), the results show growth in knowledge of writing, particularly in teachers’ ability to give feedback to students at higher levels of text. Most significantly, results demonstrate that the level of a teachers’ knowledge of writing is significantly and strongly related to the extent of their students’ progress in writing.

Chapter 5 examines the conditions under which teachers promote student self-regulation of their learning. It is argued that self-regulation depends on students being clear about their learning goals, knowing what success looks like and receiving feedback that assists them to reach their goals. Twenty writing lessons for
15 different teachers were observed and, at the end of each lesson, students were interviewed about their understanding of the above features. The observations took place at three phases of the professional development. The first observations were undertaken prior to the professional development, the second before the end of the first year, with the third at the end of the second year.

Teachers showed substantive change in the extent to which they shared learning aims, success criteria and directed their feedback to the learning aims. Students’ responses were very different in the lessons of those teachers who were explicit in these lesson features and those who were not. When teachers were explicit, interviewed students referred to the deeper features of writing when asked what they were learning about writing, what success looked like and what their teachers told them to work on in their writing. When those teachers were less explicit, the students typically referred to learning aims and success criteria in terms of the surface features of spelling, punctuation, neatness and length.

Chapter 6 presents data with respect to teachers’ knowledge of the use of evidence in decision-making. This includes knowledge of the principles of evidence-informed decision-making and knowledge of how to interpret data. Significant progress was demonstrated by teachers in terms of accurately referencing data when drawing inferences about student needs from reading or writing assessment data. However, the level of this knowledge did not relate directly to student progress as it is argued that knowledge of the principles of using evidence in decision-making need not mean placing value on its use. Nor does the ability to interpret data mean that a teacher has the pedagogical content knowledge needed to apply that interpretation.

The last chapter in this part of the report, Chapter 7, presents a series of findings from teacher self-perceptions of both their role and their learning. Teachers’ beliefs about how much influence they had on student achievement changed significantly over the course of the project. They attributed to themselves increased influence on student achievement while reducing the amount they thought was attributable to the home and the child him/herself. Teachers also reported increased satisfaction with the level of achievement of their students. Similarly, their self-reported confidence level with respect to a number of dimensions of pedagogical content knowledge increased significantly. Finally, the results of a questionnaire designed to enquire as to whether teachers thought their skills were improving and as to whether they felt supported and not overwhelmed yielded positive reports. Teachers reported they were learning new material; they anticipated learning more and the time was well spent. It is of some concern, however, that the patterns of ratings seem not to be related to improvements in student learning and this aspect is something that needs further investigation.

Part C, the third section of this report, comprises four chapters that focus on conditions that have been identified that are more or less effective in promoting teacher learning. Chapter 8 outlines a single school case study illustrating how one facilitator combined the challenges of meeting the four project outcomes with a needs analysis approach to promote teacher learning. Initial data were collected through the needs analysis tools used in all research schools, including observations, interviews and questionnaires. A follow-up visit from a member of the research team four months after the school began its involvement with the contract involved further interviews and observations.
The initial classroom observations showed teachers to have carefully prepared writing lessons with instruction focused primarily on motivating the students to write. Instruction about writing during the lesson related primarily to surface features. Student interviews established that all knew what it was they were supposed to be doing, but did not understand what it was they were learning about writing, or the features of effective writing beyond surface features of spelling, length, neatness and punctuation. Their understanding of feedback received also related to surface features.

When providing feedback to the teachers, the facilitator summarised the teachers’ practices, elicited the beliefs on which those practices were based, and related them to the students’ responses. When the teachers realised that their students did not understand the learning aims of the lesson or success criteria for their writing, they decided that they needed to be more explicit in their instructional approach. These student interviews provided the catalyst for change. Other information from the scenarios on the teachers’ pedagogical content knowledge and student asTTle scores helped the teachers to further define their learning needs.

During the next four months, the teachers engaged with relevant readings, classroom observations and staff meetings focused on building the knowledge and skills to make their teaching of writing more explicit. At the end of this time, further classroom observations and student interviews showed they had been successful and student achievement improved substantially during these four months.

Chapter 9 provides reflections on the national facilitators’ learning journey. By the end of the two years four themes were evident in their open-ended discussions of the most significant learning for them. These themes included, in descending order of frequency, the importance of addressing leadership and change management issues, the need to have structured conversations with teachers and leaders, key issues around the development of their content knowledge, and how to exit from schools at the end of the contract in ways that support ongoing learning.

Research data collected over the two years on facilitation showed that this journey had not been easy. At the end of the initial needs analysis phase, it became evident that some were using the data primarily for their own ongoing planning to shape the professional development activities, rather than using the information to co-construct a needs analysis with staff so that they understood their own learning needs and help to construct what needed to happen to meet them.

A few months later, facilitators were asked to write how they would respond to a hypothetical typical scenario in a school. The majority of the responses indicated the impact of the emphasis of the project on facilitating deeper professional learning by examining the beliefs on which practice is based, making decisions on the basis of the evidence in relation to student needs or the impact of practice on students, and promoting student learning through improved knowledge on which to base more effective teaching. However, when facilitators were asked to articulate exactly what they would say in these situations in the follow-up discussions, it became apparent that most restricted their input to asking questions. They were careful not to give their own opinions directly. Although the questions, as facilitators formulated them, were probing of teacher and literacy leader beliefs, at the same time they were highly leading in the sense that the facilitator’s reasons for asking them and suggestions for how to improve decision-making were embedded.
within the questions. The consequences of this style of facilitation in terms of under-utilizing their expertise was discussed with them with some showing a more co-constructed approach to their ongoing work.

Chapter 10 examines how teacher learning can be promoted within professional learning communities by examining the evolution of teacher meetings over the two years. All participant groups gave significantly higher ratings to the usefulness of meetings to assist teachers to teach those students they found challenging. Unfortunately, with the attrition of schools over the research period, it is difficult to make any general statements about change from the meeting observations and interviews. From the available evidence, it appeared that one school made substantive changes in their use of evidence of student outcomes to reflect on the effectiveness of their teaching and to plan future units. However, little change appeared in the observed meetings of the other two schools.

The final chapter in this section, Chapter 11, identifies ways to promote teacher learning through structured learning episodes. Field notes or taped transcriptions were taken of 18 different structured learning episodes involving twelve different teachers. Nine episodes involved observation and feedback, seven of meetings and two of a teacher observing others teach. The teachers’ reactions to these episodes were variable. No particular activity was more successful than any other, rather it was what happened within that activity and how well it related to the particular teachers’ learning needs that was more important.

Three interrelated principles were identified that were associated with teachers responding positively to the activities. The first was the principle of negotiation and co-construction as ways to promote engagement and learning. Whether it is the purpose of a particular activity, the form of the activity, the learning the activity was supposed to promote, or the meaning of particular data, each aspect needed to be negotiated and co-constructed with participants. Ongoing checking with teachers was an essential part of the process. Associated with this principle is the second, that of engaging teachers’ existing theories about students and how to teach them effectively. This principle is important in the activation of prior learning so that new learning can build upon prior learning. In some situations, change may depend on challenging existing theories and creating dissonance with particular positions. In the absence of theory examination, new practices are typically overlaid on previous practice, with superficial, rather than substantive, change evident. A third related principle was that of promoting self-regulation. When teachers have participated in setting personal goals, understand how to monitor their progress towards them and have the support to make appropriate changes, they are more likely to engage at a deeper level and sustain change.
Chapter 1
Overview of research

The Literacy Professional Development Project began working with its first cohort of schools in February 2004.

The professional development in literacy project was designed to achieve the Ministry of Education’s strategic goals of raising student achievement and reducing disparity (Ministry of Education, 2006). The project has four outcomes: Evidence of improved student achievement; evidence of improved teacher content knowledge; evidence of improved transfer of understanding of literacy pedagogy to practice; and evidence of professional learning communities. The first of these outcomes relates directly to the Ministry’s strategic goals, of raising achievement and reducing disparities. The other outcomes are designed to target those skills, content areas and processes that recent New Zealand and international research indicates are likely to create the conditions for promoting the strategic goals.

The team contracted to promote these outcomes comprised personnel from Learning Media Limited who were responsible for project management and delivery of the professional development, and two researchers from The University of Auckland who worked alongside the Learning Media team to investigate particular aspects of the initiative and provide ongoing feedback. “What are the key processes that promote professional learning in ways that impact positively on student outcomes?” This report focuses on the research component of the project from February 2003 – December 2005. It is intended that this report accompany and complement the report from Learning Media (English & Bareta, 2006), rather than stand independently of it.

In this introductory section, an overview of the literature that guided the research is provided together with the overarching research questions. An indication is given of the content of the chapters that address particular aspects of the literature and research questions. This introductory chapter concludes with an overview of the methodology employed in the research. In each subsequent chapter, specific studies are reported with relevant literature discussed in greater depth with the methods associated with the study described in greater detail.

The Literature Guiding the Research

A fundamental assumption of this project and the accompanying research is that quality teaching has a significant influence on a range of student outcomes (Alton-Lee, 2003; Hattie, 2003; Nye, Konstantopoulos, & Hedges, 2004). Teachers’ influence is moderated by a number of others, such as students’ prior learning and family contexts, but it is teaching that is the greatest system influence. If teachers are to maximize their influence then they, like their students, need opportunities to deepen their understandings and refine their skills. As Smylie (1995) noted “We will fail … to improve schooling for children until we acknowledge the importance of schools not only as places for teachers to work but also as places for teachers to learn” (p.92). The need for ongoing professional learning arises because teaching challenges do not remain static.
Changing student demographics and an ever-changing knowledge base mean that teachers need to be kept, and to keep, abreast of current evidence about how best to meet the learning needs of their students. Therefore, the first of the overarching questions explored in this research comprised, “What are the key processes that promote professional learning in ways that impact positively on student outcomes?” The literature on which we have drawn to inform our approach to this research project also forms the basis of the literature informing the Best Evidence Synthesis on Professional Development and Learning (Timperley, Fung, Wilson & Barrar, 2006) because both projects are designed to answer similar questions.

A major difficulty arises when attempting to answer this question because the link between what is planned and provided in a professional development project and what students learn is not direct. Considerable effort has focused on understanding the black box between the act of teaching and what students learn (Black & Wiliam, 1998) as a result of which a growing body of evidence (see, e.g., Alton-Lee, 2003) has contributed substantially to uncovering the processes and possibilities within it. This project, however, faced an even more demanding task in that a second black box is added to the teaching – learning sequence. It is situated between the professional learning experiences offered to teachers and the impact of these experiences on teaching practice. Little is known about how teachers interpret the understandings and skills targeted in any particular kind of professional development and the consequent impact on teaching practice, except that the relationship is far from simple. How teachers change their practice, of course, then impacts on student outcomes, depending on how students interpret and utilize what is taught. Figure 1.1 illustrates the parallels and processes in each situation.

**Figure 1.1**

**The black boxes of teacher and student learning**

The national and international evidence related to unpicking the first black box in the chain is relatively recent and it is hoped that this research will assist in furthering our knowledge about the conditions needed to promote teacher learning in these contexts. Much of what was undertaken previously has been based more on provider preference and assumptions about effectiveness rather than specific evidence of effectiveness, or has been of such short duration that it could not be expected to have any impact on students. For example, a national survey by the United States National Staff Development Council (National Staff Development Council, 2001) reported that in nine out of ten content areas, most teachers said that they spent one day or less on professional development during the previous year. Although no equivalent data is available for New Zealand, the popularity in this country of conferences and brief workshops indicates that professional development that is sufficiently intense and sustained to have an impact on students is often not the norm.

If professional development is to make a difference to students, then it needs to promote the kind of professional learning that is internalised by individual teachers through the development of professional
knowledge and skills (Hannay, Mahony, & MacFarlane, 2004). This may appear to be stating the obvious, yet the professional development literature typically describes the delivery of some kind of information to teachers in order to influence their practice. It may include reference to how practice was changed as a result but rarely addresses the processes of professional learning that lead to these changes in practice. In this project, the issue of professional learning has been addressed explicitly, with different chapters of this report examining the specifics of what was learned (Chapters 4, 5 & 6) and examining the circumstances, which promoted particular kinds of learning (Chapters 8, 10 & 11).

Another consideration underpinning this report is that changes in teaching practice do not necessarily lead to changes in outcomes for students. The evidence shows that sometimes apparently common-sense assumptions are not supported by outcome evidence in a range of situations. For example, Ross (1994) reported counter-intuitive outcomes for the students of teachers participating in professional development on co-operative learning that was designed to help teachers feel more efficacious in using these teaching practices. Teachers who indicated through their self-ratings they were feeling both personally and generally more efficacious over the period of the professional development, had students who rated their willingness to offer and seek help lower on average than those in classes whose teachers did not experience the same improvement in self- and general-efficacy. These student outcomes were not consistent with the principles of co-operative learning.

A second example comes from technology. Harwell, Gunter, Montgomery, Shelton, and West (2001) established, through a range of measures, including classroom observation, that professional development that resulted in increased integration of technology into science and mathematics classrooms was not associated with any positive changes in student perceptions of their classroom learning environment. Two additional studies, one in the Netherlands (Van der Sijde, 1989) and another in the United States (Gottfredson & Marciniak, 1995) further found that the students of teachers who participated in professional development had poorer outcomes than the students of the teachers who did not participate. We have, therefore, consistently referenced our findings in relation to changes in teaching practice to consequent changes in students’ acquisition of desired literacy skills and knowledge.

The measurement of the relationships between what is offered in professional development, what is learned by teachers through that process, and outcomes for students, however, is complex. Some of these measurement issues are described in Chapters 2 and 3.

**Issues Related to Promoting Teacher Learning**

Teachers in New Zealand, as in other countries, are diverse in a number of ways and have diverse learning needs. Demographic descriptions of ethnicity, age structures and teaching contexts provide some limited information about these needs. New teachers to the profession, for example, clearly have different learning needs than their more experienced colleagues. Beyond this relatively superficial level of analysis, however, demographics cannot tell us much. Even knowledge of the number of new teachers to the profession does not provide information about the prior learning experiences of these teachers in their pre-service programmes. In addition, experienced teachers who have developed competence in one situation may not have competence in another. Changing contexts, created by evolving student demographics and the development of new
research knowledge may render today’s expertise inadequate or less than optimal for teaching tomorrow’s citizens.

When considering how to construct an effective teacher learning environment for this project, it was recognised that not only are learning needs likely to be dissimilar among any group of teachers participating in a particular professional learning ‘event,’ but also the types of learning environments are likely to be responded to differently by those participating. What needs to be learned in any given situation depends on both the prior learning and skills of individuals and groups, and the demands of their current teaching context because different practice contexts create different demands. For this reason, an initial analysis of teachers’ learning needs was undertaken in this project in order to tailor what was offered to the participants in the different schools. Chapters 4 and 8 provide more detail on the effectiveness of and some of the challenges related to such an approach.

Another reason for undertaking an analysis of learning needs prior to offering teachers the opportunity to learn particular knowledge and skills, is that teachers do not implement discrete bits of practice independently of a set of integrated beliefs and practical constraints in their classrooms, (Donovan, Bransford, & Pellegrino, 1999; Kennedy, 2004; Robinson & Lai, 2006). Rather, these practices are based on coherent conceptual frameworks that include understandings of students and how they learn, the kinds of knowledge and skills that are valued, and the teaching methods believed most likely to be effective in promoting their acquisition. It is important, therefore, for any new knowledge and skills to be presented in the context of what currently underpins practice. For teachers, the most powerful context is their daily teaching practice and the social relationships with other professionals in which that practice is situated (Cobb, McClain, Lamberg, & Dean, 2003). The influence of these contexts on professional learning in this project is discussed in Chapter 10.

What Needs to be Learned
A focus on learning processes alone, however, leaves unaddressed what it is that teachers need to learn in order to make a difference for their students. Our second overarching research question, therefore, comprised; “Through these processes, what is the important content that needs to be learned in order to have this impact?”

It has long been accepted that teacher content knowledge is central to their effectiveness as a teacher. Shulman (1986, 1987) called attention to this fact when he called it a “missing paradigm” (1986) in the study of teaching, arguing further that teachers require a particular sort of content knowledge; they need to know the subject in a way that helps them to teach it to others. Shulman’s (1986) notion was that such pedagogical content knowledge “embodies the aspects of content most germane to its teachability” (p. 9). This includes not only the major topics but also representations of knowledge (including, particularly, the transformation of subject matter for teaching) and an understanding of what makes learning easy or difficult, and of student learning difficulties and strategies to deal with them.

Despite this acceptance of the importance of pedagogical content knowledge, there is surprisingly little evidence linking such knowledge directly to improved student outcomes, particularly in literacy. Part of the
problem in establishing this link is attributed to difficulties in measuring teacher knowledge (Choi & Ahn, 2003; Ingersoll, 1999; Friedman, 2000). In chapter 4, these links are explored in some detail.

Another area of teacher knowledge generally accepted as important to effective teaching is an understanding of the developmental progressions in a particular curriculum area, how to assess these progressions and craft teaching practices targeted to the next step in children’s learning. Unfortunately, assessment is often thought of in terms of summative evaluations of a students’ learning, rather than an opportunity to provide effective feedback and to plan more effective teaching experiences for students. Changes in teachers’ understandings of the more formative uses of assessment and how these understandings are evident in their teaching practices are described in Chapters 6 and 8.

**Curriculum Development and Schooling Improvement**

Although what happens in classrooms has the most impact on student learning, the teacher–student relationship is strongly influenced by the organisational and social environment in which teachers practice. Curriculum development for teachers cannot be considered independently of the school situation in which they practice. For this reason, the project has been school-based, rather than focussed on individual teachers independently of their school situation. Leadership development has focused on the roles of literacy leaders and their principals. There is a growing awareness in the research community that these positions require more than organisational capacity to be effective (Robinson & Lai, 2006). Those in leadership positions also need reasonably sophisticated pedagogical content knowledge in order to provide the conditions for teachers to learn. Although we have not specifically addressed leadership roles in this report in any detail, many of our measures have included the principals and literacy leaders’ knowledge and skills.

**Overview of Method**

In this part of the chapter, an overview of the method is provided. Details of the features of the major instruments are outlined. However, as each chapter in the report functions as an independent entity, additional details of the sample, the procedure and of the materials are provided in each chapter.

**Design**

Essentially the research reported is that which accompanies and informs the implementation of the national Professional Development in Literacy Project. The research involved examining the processes and outcomes of this project. Data were obtained from a convenience sample of schools participating in the project and these schools functioned as multiple cases. Baseline data to gain a picture of school strengths and needs (referred to as the needs analysis) were collected at the time schools joined the project and compared with like data at the end of a time period, for all schools one year and, for some schools, two years.

The variables considered included student achievement data and teacher pedagogical content knowledge in either comprehension or writing; knowledge of the principles of the use of evidence and more specific knowledge of data interpretation; elements of teacher classroom practice and of leader meeting practices; teacher beliefs about their influence; their confidence in setting learning objectives, teaching and assessing comprehension or writing, and measures of school organisation and climate with respect to innovations. Multiple methods were used to collect data. Details of what happened during the period between the initial
and post tests were not researched in detail, although post needs analysis interviews were undertaken and ongoing episodes of teacher learning sampled.

**Participants**

This project was available nationally to schools with students in Years 1-8. The project had two intake points for schools in 2004 so there were two cohorts of schools, one beginning in February and one in July. Seven schools from the February intake were invited and agreed to join the research. These were from the Auckland/Northland cluster. Five of these schools had elected a comprehension focus and two, a writing one. The six research schools in intake 2 in July were from the Otago/Southland cluster. Of these, two were focused on reading comprehension and four on writing.

Intake 1 schools were five schools in Auckland and two in Whangarei. There were three intermediates, two contributing primary schools and two full primary schools. Intake 2 schools in Southland included five contributing primary schools and one full primary.

The project was initially conceived, given funding cycles, as a one year project. When it became clear that an extension to two years was to happen, schools were invited to continue. Of the research schools in cohort one, five continued while all of the six cohort two schools continued into the second year to make an 18 month project time frame for them. Data were collected at the end of the second year of the project. However, two large intermediate schools had been lost from intake 1 and this, coupled with natural attrition meant that the sample sizes of teachers and students with a third data point at the end of 2005 were considerably reduced.

**Instruments**

**Student achievement**

For the Year levels where they were available, nationally normed instruments were used to measure student achievement in reading and writing to permit calculation of initial position and then extent of change, relative to both starting point and to expected movement. Thus, the research data we have for reading is for Years 3 to 8 and, for writing, for Years 4-8. In 2004, reading comprehension was measured using the Supplementary Test of Reading (STAR, New Zealand Council for Educational Research). In 2005, reading focused schools remaining in our sample, bar one, agreed to administer Assessment Tests for Teaching and Learning (asTTle): Reading in addition to STAR. Writing achievement was measured using Assessment Tests for Teaching and Learning (asTTle): Writing.

The **Supplementary Test of Reading (STAR)**: was used to measure reading ability and, more specifically, comprehension. It is a generalised test of reading ability rather than having criterion specifically referenced to the curriculum. As its title suggests, it is intended to supplement and validate the assessments of teachers; to provide a comparison point for teachers in relation to a wide range of students at a comparable level in certain reading tasks and to help them to group; to identify students in need of extra assistance; to assess new students, and to evaluate programmes. The STAR test is intended for Years 3-9. The test has two parallel forms and may be administered at any time in the school year.
There are four timed subtests for Years 3-6, word recognition, sentence comprehension, paragraph comprehension and vocabulary range and two additional subtests for Years 7 to 9, namely, the language of advertising and reading different genres or styles of writing. Word recognition tests the decoding of words familiar to students in their oral vocabulary. It involves a picture word match. The words for Years 3-6 are taken from the NZCER Noun frequency list (Elley & Croft, 1989) and for Years 7-9 from the more common words in the New Zealand Oxford Primary School Dictionary. The sentence comprehension subtest involves reading sentences and completing with the appropriate word. The paragraph comprehension is a cloze procedure task where students fill in the blanks in three paragraphs (15 blanks at Year 3 and 20 at higher levels). At Year 3, a set of words that include the correct ones plus distractors is provided for students to use to replace the missing words. Vocabulary range measures knowledge of word meanings. For Year 3 these are taken from NZCER “Spell-Write” list (Croft & Mapa, 1998) and for Years 4-9 from the NZ Oxford Primary School dictionary. The additional two subtests for Years 7-9 include one that requires the identification of emotive words that provide no information, while the second subtest includes paragraphs representing a range of genre and the student has to select the phrases which best fit the style and purpose of the writer.

For each subtest the New Zealand mean and typical range is given. Also, a critical score on each subtest is identified and students who score below this are seen to be “at risk”. A critical total score (the sum of the subtests) is provided, below which students are considered to be non readers, that is they have not demonstrated that they ‘can read’ in terms of the criteria stated in the Literacy Task Force list. The cut-off score for “being able to read connected text….takes advantage of the known performance of thousands of such pupils, and of the theoretical fact that pupils can guess correctly, without knowledge in 25% of the questions on four-choice tests” (p. 20 Teachers’ Manual). Stanine norms for total scores are provided for three points in the year (February-May, June-August and September- November), together with an indication of the percentage of students who typically score in that stanine. In addition, a guide is given in the form of a range of scores to allow teachers to see whether students are reading at the typical level for their age. Raw score means and standard deviations for each sub test are given.

A norming project established the stanine norms for each class level. This involved for Year 3 testing 3,000 students in 60 schools in 2003; for Years 4-6, 4,500 students (1,500 at each year level) in 52 schools in 1999 and for Years 7-9, a corresponding number in 2001 from 68 schools for Years 7-8 and 48 for Year 9. To select the sample of schools, a stratified procedure was used according to size, decile and location. Schools were randomly selected from within those strata. In the Teachers’ Manual, Elley states that while stanine scores are technically useful in that they can be regarded as approximately equal interval scores, they are a somewhat coarse scale and that normally two stanines difference would be needed to conclude a reliable difference, significantly different from chance (Elley, 2001).

The overall reliability of total scores is high (0.9 with a measurement error of about three points) but for subtests it varies from 0.6 to 0.8 (measurement error about 1.5 points). The subtests are noted to be quite highly correlated (0.5 – 0.7) so a substantial difference of three or four raw score points is needed to say that sub-skills are really different. Elley notes that it is easier to identify real differences between the means of groups of students than for individuals. With respect to the validity, content validity is argued to be established, in that the tasks assess skills central to the reading process as defined by the Literacy Taskforce and the English Curriculum statement. Statistical validity is established from the correlation with reading...
comprehension in the Progressive Achievement Test of Reading. A limited sample (n= 75) of Year 6 and of Year 4-5 showed correlations of 0.77 and 0.73, respectively. The correlations established in three different schools for Years 7-9 were similar, 0.70, 0.78 and 0.74. Practice effects have been noted when the STAR is taken twice within a three month period. For Years 4-6 if different forms are used, the effect is negligible, 0.42 points on the total score. However, if the same form is used the effect is 2.24. For Years 7-9 the effects were 1.5 and 2.0 points, respectively.

Assessment tools for teaching and learning (asTTle): This is a suite of assessment tools designed to help teachers assess diagnostically in order to allow more targeted teaching to enhance learning and instruction (www.asttle.org.nz). asTTle tests are referenced to the national curriculum and aim, through student responses, to highlight features related to curriculum strands, achievement objectives or functions and levels. This provides teachers with information about what to watch for or to assess further or what to attend to in order to progress learning. Normative data from a nationally representative sample are available for various groupings including year level.

Writing: The writing part of the asTTle (Version 31) literacy tools, Levels 2-4 was utilised in this research. It provides the only measure of writing achievement with associated national norms. Theoretically, in terms of this tool, writing is seen as purpose or function driven and there is a scoring rubric for each of the six functions or purposes writing is seen to serve (to persuade, to instruct, to narrate in order to inform or entertain, to describe, organise and classify in order to report; to explain and to recount. A large selection of writing prompts, that serve to have students write for one of the purposes for writing, is provided. Scoring criteria are provided for each of the seven dimensions of writing at each curriculum level. The four deep features are rhetorical or audience; content inclusion; structure/organisation of text and language resources. The surface feature dimensions are grammar, spelling and punctuation. Within a curriculum level, dimensions are scored at one of three levels, namely basic, proficient and advanced according to the weight of evidence. Examples of scored annotated texts for each purpose as well as scoring tips are provided within the asTTle CD Rom.

Standardisation of asTTle followed similar procedures for versions 1 to 3. A stratified sample of schools (based on location, size, type, and decile group characteristics, in proportion to the percentage of students in each group nationally) was drawn. Calibration and teacher feedback studies followed.

Studies have established that it is possible for teachers with minimum training to score scripts at a satisfactory level of reliability. Project asTTle reports several reliability studies concerning the use of the writing rubrics (Brown, Glasswell & Harland, 2004). Levels of reliability vary according to dimensions scored but with adjacent levels counted as agreement, coefficients are above 0.6, which is considered satisfactory in the writing assessment field. However, for the research, we scored all scripts at Time 1 and where the total of school scripts was greater than 50, we scored approximately 50 percent of samples at Time 1. In the course of the project Version 4 was released. The scoring rubrics were recalibrated to accommodate Levels 5-6. Although schools began to work with Version 4, which included Levels 5-6, for research purposes, we retained Version 3 scoring rubrics.
2. This meant that in all schools, bar one, we scored all samples at both points in time and can, thus, be confident of the reliability of the scoring for the data obtained.

**Reading:** Achievement in reading is analysed and reported by curriculum achievement objectives and processes. All questions in the reading test bank and the materials were classified by panels of teachers into the various curriculum objectives and important dimensions (through a series of “item signature studies”). The panels judged items in relation to three processes (a) exploring the English language (b) thinking critically about language and meaning and (c) processing information to identify, understand, store, organise, retrieve, combine, and communicate it.

In addition, the development of curriculum indexed assessment items (including the reporting of student achievement) are framed by the notion of surface and deeper cognitive understandings, derived from the SOLO taxonomy. Surface responses and questions require the knowledge or use of one piece of given information, fact or idea obtained directly from the problem or the knowledge or use of more than one piece, each used separately or in two or more distinct steps requiring no integration of ideas. The deep processes require a change in the quality of thinking so that it is cognitively more challenging. Questions that do this may require relational thinking where two separate pieces of knowledge, information, fact or ideas need to be integrated, that is the learner needs to impose an organising pattern on material. At the highest level, the respondent has to go beyond the information given and work out a more general rule for example.

In reading, deep processing involves a range of comprehension processes across a spectrum of difficulty, namely, **finding information, knowledge, understanding, connections and inference**. (Table 1.1 describes the type of item associated with each of these). Reading items may fit more than one objective in one or more content areas.

In constructing a reading comprehension test, teachers are able to select the difficulty level of the test in terms of curriculum levels and to decide whether they want most, many, some or few items to come from the curriculum levels specified. The programme, however, still selects a few items from outside the levels specified. When creating a reading test, the teacher also selects up to three foci/curriculum functions from the range of comprehension processes shown in Table 1 and nominates the relative amount (few, some, many, most) of each type of item. Again, there are a few items from outside the specified focus areas that are also included. There are many thousands of different permutations of tests that can be created. The scores derived from any of these tests are comparable because an item response model was used to locate both items and students on a common scale. So, students can be compared on a common scale no matter which items they actually answered or a student can be compared at different points in time with different tests. The main reason for this is that the difficulty of each and every item has been carefully estimated using item response theory (IRT) (IRT is a mathematical model that allows the difficulty of items and the ability of students to be located on one common logistic scale).
Table 1.1
Curriculum area objectives levels 2-4

<table>
<thead>
<tr>
<th>Language Deeper Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding Information</td>
<td></td>
</tr>
<tr>
<td>Find, select &amp; retrieve information</td>
<td></td>
</tr>
<tr>
<td>Skim/scan for information</td>
<td></td>
</tr>
<tr>
<td>Note take in a variety of ways</td>
<td></td>
</tr>
<tr>
<td>Use dictionary, thesaurus, atlas</td>
<td></td>
</tr>
<tr>
<td>Identify fiction &amp; non-fiction texts</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
</tr>
<tr>
<td>Knowledge of vocabulary</td>
<td></td>
</tr>
<tr>
<td>Knowledge of poetic &amp; figurative language</td>
<td></td>
</tr>
<tr>
<td>Knowledge of semantic, syntactic &amp; visual grapho-phonetic cues</td>
<td></td>
</tr>
<tr>
<td>Knowledge of strategies to solve unknown words &amp; gain meaning</td>
<td></td>
</tr>
<tr>
<td>Knowledge of publishing conventions</td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td></td>
</tr>
<tr>
<td>Read for meaning</td>
<td></td>
</tr>
<tr>
<td>Understanding / identification of main ideas</td>
<td></td>
</tr>
<tr>
<td>Understanding of detail to support main ideas</td>
<td></td>
</tr>
<tr>
<td>Use understandings &amp; information</td>
<td></td>
</tr>
<tr>
<td>Question to clarify meaning</td>
<td></td>
</tr>
<tr>
<td>Discuss texts &amp; identify aspects</td>
<td></td>
</tr>
<tr>
<td>Connections</td>
<td></td>
</tr>
<tr>
<td>Compare similarities &amp; differences within &amp; between texts</td>
<td></td>
</tr>
<tr>
<td>Make links between aspects of text</td>
<td></td>
</tr>
<tr>
<td>Make use of prior knowledge</td>
<td></td>
</tr>
<tr>
<td>Understand &amp; organise or sequence material</td>
<td></td>
</tr>
<tr>
<td>Empathise with characters &amp; situations</td>
<td></td>
</tr>
<tr>
<td>Make links between verbal &amp; visual information</td>
<td></td>
</tr>
<tr>
<td>Inference</td>
<td></td>
</tr>
<tr>
<td>Explore author's purpose &amp; question intent</td>
<td></td>
</tr>
<tr>
<td>Make inferences</td>
<td></td>
</tr>
<tr>
<td>Read critically for: bias, stereotyping &amp; propaganda</td>
<td></td>
</tr>
<tr>
<td>Predict possible outcomes</td>
<td></td>
</tr>
<tr>
<td>Identify &amp; discuss purposes of text</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language Surface Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar</td>
<td></td>
</tr>
<tr>
<td>Identify word classes</td>
<td></td>
</tr>
<tr>
<td>Use grammatically correct structures</td>
<td></td>
</tr>
<tr>
<td>Identify features or characteristics of text</td>
<td></td>
</tr>
<tr>
<td>Punctuation</td>
<td></td>
</tr>
<tr>
<td>Spell correctly</td>
<td></td>
</tr>
</tbody>
</table>

Source: asTTle Psychometric Manual Version 3.0 p. 13
The research team constructed asTTle reading tests to use at three different levels of the school. There were two forms constructed at each level to use as pre and post tests (these are parallel by definition).

The level 2 reading test: For Year 4, we opted for most of the items at level 2 and a few at level 3. The resulting test had 6 items at 2B; 7 at 2P, 10 at 3A, 3 at 3B, 1 at 3P, 0 at 3A, 1 at 4B and 1 at 4P. The foci we chose for this test were finding information (some items), knowledge (some) and understanding (most). As a result there were 11 finding information items, 10 knowledge items and 29 understanding items. There were also 3 connections items and 5 items dealing with surface features. There were 17 items regarded as tapping surface features and 18 deep feature items.

The level 3 reading test: For Years 5 and 6 we selected some items at level 2 and many at level 3 (few at Level 4). The result was 2 items at 2B, 6 at 2P, 5 at 2A, 8 at 3B, 5 at 3P, 3 at 3A, 5 at 4B and 1 at 4P. For this level 3 test we selected knowledge (few - there were 13), understanding (some - there were 29) and inference (many there were 30). However, although these were nominated, there were also items the programme selected dealing with finding information (9) and connections (4). There were 15 surface feature and 22 deep feature items.

The intermediate school reading test: For Years 7 and 8 we chose few items at level 2 and most at level 3 with a few at Level 4. The result was 4 items at 2B, 1 at 2P, 0 at 2A, 11 at 3B, 8 at 3P, 5 at 3A, and 5 at 4B. For this test we selected understanding (some - there were 16), connections (some - there were 11) and inference (many there were 23). However, although these were nominated, there were also items the programme selected dealing with finding information (14), knowledge (5) and surface features (1). There were 12 surface feature and 23 deep feature items.

Leader interview questionnaire
The principal and literacy leader were interviewed on the initial visit to gain contextual details about the school’s composition; their practices, particularly in relation to innovations and to assessment, and regarding their preparedness for the project. They were asked about their envisaged and, subsequently, their actual role in the project.

Teacher (and Principal and Literacy Leader) questionnaire
This instrument had several reporting formats. Teachers rated and provided reasons for ratings for a range of statements. They answered questions that required them to identify issues from scenarios. They self reported, by rating, confidence in various aspects of practice.

The instrument contained measures of several variables. Specially constructed scenarios aimed to tap teachers’ knowledge in three areas, namely pedagogical content knowledge, knowledge of the logic of evidence-informed decision-making and knowledge of data interpretation and use. The first scenario described teacher classroom practice in either comprehension or writing. Responses to the scenario provided an indication of the teacher’s level of pedagogical content knowledge. Some aspects of the practice described in the scenario could be considered effective in terms of features identified in the best practice in literacy literature, while others were less effective. Teachers responded by identifying, by rating and by giving reasons for their choices. Similarly, the second scenario was designed to examine teachers’ knowledge of
the processes that characterise the use of evidence of learning in the classroom to make decisions about
effectiveness of teaching and about future teaching moves. Again, ratings and reasons for them were the
response format. A third scenario required teachers to examine a set of data and draw interpretations from it
by responding, largely, to open-ended questions.

There were also a series of questions which asked teachers to rate their confidence in carrying out various
pedagogical practices in either reading comprehension or writing like, for example, setting learning
objectives, sharing them with children and teaching or assessing specific aspects of either comprehension or
writing.

Another series of questions examined variables thought to impact on teacher practice including their beliefs
about their influence. Teachers were asked to name the proportion of influence they thought that each of
several sources, including their teaching, had on student achievement. They were also asked to rate the extent
to which they felt they might be held accountable for student achievement levels. Other questions asked
about aspects of the school climate, for example with regard to carrying through new programmes.

Teacher classroom practice observations
In each school, three classroom teachers were observed. The management in the school was asked to
nominate the three teachers by selecting one teacher who was “well on the way” as far as literacy teaching
was concerned, one who was “making reasonable progress” and one teacher whom they felt “had a way to
go”.

These teachers were observed for between 30 and 45 minutes in a reading or writing lesson, which we asked
to be one that was a teacher-active lesson. We particularly stressed that we wanted to see the initial stages of
a lesson where the teacher would be setting up the learning. Variables under observation included whether
learning intentions were specified; the nature of these; whether they were shared with children and whether
success criteria or their equivalent were evident. Also of interest were the extent of alignment between aims
and activities and the nature of explicit instruction.

The teachers in the February cohort were interviewed briefly prior to the lesson to ascertain the lesson focus;
where the lesson sat in the ongoing programme; whether it was a whole class or group work lesson and, if
the latter, what is planned for groups at times they are not working with the teacher. For the July cohort and
at the end of year for the February intake, teachers completed a brief pre-observation questionnaire dealing
with these issues. Also, for these latter groups, the teacher was taped while wearing a radio microphone and
there were, where possible, two observers present.

In each classroom, teachers were asked to indicate between three and nine students of a range of ability with
whom we could have an informal conversation. The number of students depended on the time available
during independent seat work to carry out the interviews. We talked with the students about what they were
currently doing and learning; what their teacher had told them to work on in their reading or writing and
what they thought made for a successful reader or writer (in relation to what they were currently focusing
on). We also noted (and obtained photocopies of, where possible) any teacher comments or evaluative
feedback in draft books or language books in use during the lesson.
For the teachers still present at the end of 2005, a further round of observations was conducted.

Leader meeting practice
In the Auckland and Northland research schools, we asked to attend a meeting. For the first meeting, early in the research, we asked to observe a meeting where matters concerning teaching and learning or the examination of student achievement information were to be a major component of the agenda. In the latter data collection phase, we specifically asked for a meeting where teachers were considering assessment data (usually that from the standardized assessments associated with participation on the project). This was most often a syndicate level meeting and was targeted to be one that included the teachers who had been observed. The meetings were audiotaped where participants agreed. The person conducting the meeting (usually the literacy leader or an associate or deputy principal) was subsequently interviewed. After the first meeting, the teachers whom we had observed were interviewed individually. After the late in the year meeting, teachers were customarily interviewed as a focus group. These interviews were also audiotaped with permission. For two schools, there were additional follow-up meetings that the researchers attended although interviews were not conducted after these.

The variables under consideration include the focus of the meeting and whether evidence, particularly of student achievement, was used in decision making. Other aspects examined include the accuracy of the interpretation of evidence; the level or levels of aggregation of data examined; whether collective responsibility was taken for the interpretation, and whether and what decisions were made. Specifically, we examined whether links were made to teaching practice.

Teacher interviews post needs analysis
The three teachers observed in each school were also interviewed after the needs analysis and feedback phase of the project was completed. They were asked about what feedback they had received from the analysis and what they had learnt as a result of it. The principal and literacy leaders were similarly interviewed. Again, all interviews were audiotaped.

Facilitator interviews
At a mid point in the first year, each of the Auckland regional facilitators involved in the research schools was interviewed. They were asked about leadership within the school; the use of the needs analysis materials; the work they had done within schools using the tools and about issues that had arisen.

End of Year 1 teacher survey
After teachers had completed a year on the project they were asked to respond to a brief 10 item questionnaire that asked them to rate the project’s effect on aspects of their learning.

Teacher learning-related interviews
In 2005, an effort was made to document episodes of teacher learning. This was seen as likely to happen through several means. One might be the receipt of feedback after an observation from a facilitator or school literacy leader. Another might be a special meeting taken by a literacy leader or facilitator. To try to establish what helps teachers to learn about their practice, a sequence of data collection was undertaken. This sequence was seen to comprise an episode. First, one of the research team attended the possible learning site
Literacy Professional Development Project: Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning

Documents

The major documentary source was in the form of facilitator evaluation reports on each research school at midway through the year (July) and at the end of the year. In addition, details of the number and purpose of visits by facilitators to the research schools was supplied, with permission, from the Learning Media database.

Procedure

Ethics approval was obtained for the research from the University of Auckland Human Participants Ethics Committee. Schools were invited to be research schools and ethical guidelines followed in terms of informed consent and confidentiality.

The instruments were piloted initially in a relatively limited way in schools not part of the project, given time constraints imposed by factors surrounding the project’s start. Consequently, some changes in instruments and procedures were made for intake 2 schools and the follow-up data collection point for intake 1 schools.

The first visit to each school, for the first intake took place in late February and March and for the second intake in July and August. In conjunction with the facilitator for the school, an explanation of the project and, more specifically, what being a research school would entail, was given. The principal and literacy leader or leaders were interviewed together for a portion of the interview, then separately and these were audiotaped. Either at this time or shortly thereafter, the teacher questionnaire was administered in a staff meeting situation and observations were made in three classrooms of a reading comprehension or a writing lesson. The first intake classroom observations were recorded using an observation schedule and field notes. In subsequent observations, with specific consent, the teacher wore a radio microphone to record teacher talk.

Schools undertook to administer, with facilitator guidance, the student achievement measures. For writing, schools selected the function or purpose of writing they wanted to test and selected or devised with assistance an appropriate prompt.

The meeting observation was scheduled to fit in with the regular school timetable of meetings and normally took place before or after school and lasted about an hour. Interviews also were scheduled to fit school requirements but were as close as possible after the meeting time.

The second round of data collection for the first intake took place largely in November and comparable tools were used to collect data on the same variables as at the beginning of the project.

Data analysis

With respect to scoring the student achievement data in reading and writing, procedures were employed to ensure reliability. To check reliability of scoring of the STAR, a ten percent random sample including at least three scripts from every teacher was scored. The decision rule was that if less than 10 percent of the
scripts contained an error, we would then accept the marking as accurate. All teachers marked with at least 90 percent accuracy. We assumed for the second administration that this was unlikely to be any different.

The writing samples from the two schools in the first, February intake and the four schools in the second, July intake, were scored by the researchers as it was reasoned that teachers had had virtually no opportunity to become familiar with the asTTle scoring rubrics. In many cases the teacher scored their classes as well and compared their scoring with that of the researchers. At the second data collection point, either 50 randomly selected scripts or a 20 percent sample were scored, whichever was the larger. This meant that we again scored all of the scripts in one Auckland school and the one Southland school that submitted November data. An analysis of the data obtained for the other Auckland region school showed considerable difference between the teacher scoring of their own class scripts and the research team expert scorer who scored a sample of around a third of the school. The reliability coefficients were much lower than any reported in reliability tests for asTTle. As a consequence, any data used in this report is data from research scored scripts.

**Analysis of questionnaire data**

In order to provide research schools with feedback particularly for the needs analysis phase, for questions with ratings or similar, means and ranges were calculated. The responses to open-ended questions were grouped and direct quotes listed. From these verbatim or the recorded gist of these responses, coding categories were subsequently developed through an iterative process that both considered the actual responses but also considered the research question being addressed and how theoretical views might frame decisions about categories. For example, our research question about what teachers knew about data interpretation suggested that we should code responses broadly in terms of whether they were generalised or made reference to data and whether they were accurate or not. Similarly, theory and empirical work would suggest that in providing feedback to students (in this case about their writing) there needs to be an evaluative element that tells the student how well they have achieved what the learning aim was and also that there needs to be some element of diagnosis and information that comprises feedforward, helping the student to take learning forward and improve the piece of writing. Details of these categories appear in chapters where the instrument is used to provide data on a particular variable. For example, Chapter 4 on Teacher Pedagogical Content Knowledge details how the reading and writing scenarios were coded and reports reliability levels obtained.

**Analysis of teacher interview data**

The first interviews with teachers were post needs analysis to ascertain the impact of the needs analysis phase on their learning. Interviews were also held with teachers in association with meetings observed and in relation to learning opportunities provided. Interviews, where recorded, have been transcribed. All participants had the opportunity to review transcripts but few availed themselves of this. Again the coding categories and coding procedures, including reliability, are discussed in the relevant chapters of the report.

**Analysis of observations**

Transcriptions were made of the audiotapes from those lessons observed at Times 2 and 3 for the Auckland regional cluster and at Time 1 and 2 for the South Island schools. These were considered in terms of
significant features like clarity of the lesson aim; alignment of non-teacher led activities with the learning aims expounded and in terms of the extent and type of explicit teaching exhibited.

References


PART A: MEASURING PROGRESS

Chapter 2

Summary of Trends in Student Achievement

This chapter considers evidence of the success of the professional learning project in terms of the extent and nature of change in student achievement in reading or writing. Detailed analyses of the progress of individual schools were presented in the report of February 2005 and it is not proposed to repeat that level of detail here but rather, to summarize the trends. Similarly, detailed analyses by ethnicity were presented in the July, 2005 report so an abbreviated version is presented here. This chapter, therefore, considers the overall trends in terms of progress, then summarises progress by ethnicity and gender. The emphasis is on the period of the first year of the project, Time 1 to Time 2. The sample size with respect to the research schools reduced considerably from Time 2, the end of the first year, to Time 4, the end of the second year of the project. This is largely due to the fact that two large schools completed their contract at the end of a year but also there was natural attrition in terms of obtaining data for individual students at four points in time. However, for those students for whom we have all data points, the achievement trends are presented. Further, it is possible to view our results in light of the trends for all schools in the wider project.

Descriptions of the tools for assessing student achievement were presented in the previous chapter as part of the outline of the method but additional details, relevant to interpretation, are noted here. Technical issues associated with the use of these tools are touched on in this current chapter. However, the substantive discussion of the use of difference scores and of significant measurement issues follows in Chapter 3. At this point we need to foreground our warrant that the results from all analyses have to be viewed in conjunction with the demonstrated shortcomings of any measure employed. Similarly, sample size needs to be considered when viewing analyses particularly those by ethnicity.

Method

Procedures and Tools

Student achievement, and progress in terms of this achievement, in writing and in reading, was obtained from two cohorts of project schools. There were seven research schools that were part of the first (February 2004) cohort of schools and there are data at two points in time for all of these. In the second year of the project two schools withdrew although we have data at two further time points from one of these schools. The first data collection point was early in the 2004 year, but the exact time varied by school from late February to early April. The second data collection point was towards the end of the year in October or during November. The continuing schools collected data early in the second year. A final data collection point occurred at the end of 2005. For the second cohort, of six schools, there are two data points. The first data collection point varied from August to October 2004. The second data point was towards the end of 2005 and, again, varied from school to school.
Writing
In the research schools whose focus was writing, student achievement in terms of a particular purpose for writing was examined using an appropriate writing prompt from Assessment Tools for Teaching and Learning (asTTle). Schools both selected and administered the writing task. They mostly selected tasks involving the same purposes for writing at the two points in time. asTTle, with the use of item response theory, allows students to be compared on a common scale no matter what test they sat or the time of the year (as the difficulty of items has been estimated).

The scoring rubric associated with the writing purpose was used to assess the student samples. Reliability, in terms of consistent and accurate scoring of the writing samples, was a paramount consideration. For cohort 1, at Time 1, the beginning of the year, all scripts from the two writing research schools were scored by an experienced primary teacher who had had previous training in asTTle scoring and had considerable practice in scoring using the asTTle rubrics. At the end of the year, Time 2, the same person scored all scripts from one writing school and a large sample of scripts (n = 204) from the second writing school. Again, at Time 3 all scripts from the remaining cohort 1 writing school were scored by the research team. For cohort 2, all scripts were scored by the research team at each point in time; for one writing school we scored samples at several time points.

Using criteria, the rater allocates a curriculum sublevel to each of the seven curriculum functions or dimensions of writing (audience, content, structure, language resources, grammar, spelling and punctuation). From this asTTle generates two types of scores - a writing scale score and curriculum level scores. The first type of score is a numeric standardised score that represents performance against that of the norming population used by asTTle to calibrate the assessment tasks. The mean of all students in all curriculum areas in the asTTle sample was 500 with a standard deviation of 100. A score of 500 is approximately what a level 3 basic, student across New Zealand can achieve. Scaled scores typically range from 150 to 1250 (levels 2-6, Years 4-12). No score is below 100 and it is possible to score above 1250.

The standard error of measurement of asTTle instruments has been calculated as 15 scale points. This creates a 65% confidence interval around a reported score. This means that any score is accurate two out of three times to within plus or minus 15 points of the reported score. Scores that are within 15 points of each other are probably not different by a statistically significant margin. So, in practice, it means that scores within 15 points are not regarded as different and for improvement to be seen as “real” and not chance, it has to be more than 15 points.

The asTTle Version 4 Manual provides details to allow a consideration of growth rate using the standardised total score. This score is considered to be an all-year score. As it is a curriculum referenced test, the score represents the average achievement on curriculum objectives for that year. A table is provided in the Version 4 manual of annual growth obtained by calculating the difference between the mean scores of adjacent years. The mean scores in writing for Years 4 to 8 from the norming population are shown in Table 2.1. The average increases (difference) in scale score by year are also indicated in Table 2.1. The average growth (across all curriculum areas) per year based on the multi-year cross-sectional sampling is generally between 25 and 30 scale points at primary school (average effect size .28). Growth in writing tends to be somewhat
less. A table of percentile scores is also available in the Version 4 Manual and allows schools to consider the relativity of their writing scores in this format.

### Table 2.1
**Mean Scale Scores and Difference in Scale Scores from One Year to the Next for Writing**

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>↔</th>
<th>5</th>
<th>↔</th>
<th>6</th>
<th>↔</th>
<th>7</th>
<th>↔</th>
<th>8</th>
<th>↔</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>434</td>
<td>↔</td>
<td>454</td>
<td>↔</td>
<td>490</td>
<td>↔</td>
<td>504</td>
<td>↔</td>
<td>517</td>
<td>↔</td>
<td>581</td>
</tr>
<tr>
<td>Difference</td>
<td>20</td>
<td>→</td>
<td>36</td>
<td>→</td>
<td>14</td>
<td>→</td>
<td>13</td>
<td>→</td>
<td>64</td>
<td>→</td>
<td></td>
</tr>
</tbody>
</table>

The other type of score that asTTle generates is a score that expresses performance calibrated to the achievement objectives linked to the levels within each subject’s curriculum statement. In writing, where the marker allocates a curriculum sub-level for each of the seven curriculum functions or dimensions of writing, the score given is the midpoint of a range for that sublevel. The Version 4 Manual lists the range of scores associated with each curriculum sub-level.

#### Reading

A measure of reading comprehension was obtained from the Supplementary Test of Reading: STAR (Years 3-9). The five research schools with a reading comprehension focus administered one of the two parallel forms of the test (A or B) in February or March and the other one at the end of the first year. All but one of the research schools administered STAR at the beginning of the second year on the project. For this one school, STAR data for the whole cohort of students was not available at Time 3, the beginning of Year 2, as they only administered STAR to the new intake.

In Year 2 of the project, three of the four continuing reading research schools agreed to administer asTTle: Reading at the beginning and end of the school year. The schools administered the test we devised while the research team scored and entered these data (Note: The discussion of the relationship of asTTle to STAR results for individual students is in the following chapter).

The *STAR Teachers’ Manuals* contain most of the technical information needed for interpretation. There is a standard error of measurement for the total reading score of plus or minus three points. Tables in the Manual provide stanine norms for four points in the year for Year 3 and three points in the year for the other year groups. A table shows the expected percentages of students in each stanine and the percentile rank associated with a raw score. Tables also provide the typical scores (mean score and range of scores) and critical scores.

---

2 Note that prior to asTTle Version 4, there were criterion statements for each of the three sublevels within a curriculum level, namely, basic, proficient and advanced. Version 4 has criterion statements only for the middle category, proficient, and the rater makes a judgement of lower or higher according to these. In addition, the criterion statements in Version 4 for level 4 have been adjusted to reflect the feedback from secondary teachers and the necessity to accommodate development and the specification of this at higher levels.
for each sub-test for Year 3, Years 4-6 and Years 7-9. The Year 3 scores are given for three points in the year (Feb-March, mid year and Nov-Dec). The other years have only one set of scores and these are beginning of the year scores, according to the test designer, Prof. Warwick Elley (personal communication).

Issues
There are issues that relate to technical details of the instruments used (discussion of measurement issues is expanded considerably in the following chapter) and issues that relate to the accuracy of the data reported.

Instruments: For a consideration of student achievement in writing, there are no significant issues regarding the technical data available from asTTle. Reports by year are not possible unless the year group is re-entered as a class. For other analyses, including analyses of progress, the data have to be exported; it is not straightforward to export details like ethnicity.

In the reading data, there is an issue concerning the mean scores for Year 4. In the STAR Year 3 Teachers’ Manual, Table 4 gives raw score means for each subtest and total means and standard deviations for Years 3 and 4 in March. The scores for Year 4 in Table 5 of the Teachers’ Manual Years 4-9 are different to these. Although the time of the year the means apply to in the latter manual is not given, when asked, Elley said the tables in the Year 4-9 Teachers’ Manual showed beginning of the year data. The differences may be due to the fact that an additional norming group was used for year 4 at the time the Year 3 norms were established. For all Years there were reportedly two norming groups but these do not appear separately in the Tables. Presently, the norms from the Years 4-9 Manual are those being used for Year 4.

To calculate gain scores for STAR is problematic given that the test does not utilise item response theory that allows precise measurement of the degree of difficulty of an item. So, we cannot be sure of how the items for the years equate. Another reason that it is difficult to calculate gain is that, given the mean raw scores provided are for the beginning of the year (except for Year 3 where additional means are provided), a decision has to be made whether to use, as indicative of the mean score for a particular Year group in November, the February mean for the next Year group. This would involve, for example, using the Year 5 normative mean in February as the normative mean score for Year 4 in November.

While there are standard deviations for the total score in the Manuals, there are none for subtests. Professor Elley, the test designer was approached and provided standard deviations for the sub scales as these indicate variability. These SDs are reported for each norming samples and for all years except Year 5 (There is insufficient data to calculate Year 5 readily but Elley thinks they will be similar to 4 and 6). In general, these indicate variability greater than 20 percent of the subscale score.

Accuracy of scoring and entering data: The other area where there are issues concerns the accuracy of data recording on the part of the schools. We are confident about the asTTle writing and reading data as we scored and entered all data used in analyses. However, in the case of one writing school, at Time 2 we marked and entered only a sample, albeit a large sample of 204 students. The school were to mark and enter all students and send on these data. They subsequently provided data for around 500 students that included those we had marked. We are unable to ascertain whether the missing 200-odd students were not tested or whether the scripts were not marked or the data not entered. The degree of reliability in scoring between the
scoring of the research team and teachers for the 206 scripts in common was unacceptably low so only the research data for those students were included in the analyses.

At Time 1, the scoring of STAR scripts was checked for accuracy. We asked schools to provide a random sample of scripts from each class. The decision rule was that if any class sample (of 20% of scripts) contained ten percent or more errors, the scripts would be rescored. The error rate across all classes was negligible (less than 10%).

In addition, at Time 1, we checked the accuracy of recording data as, in most cases, schools provided the data from STAR. Some data were provided in hardcopy and we entered the data but, mostly, data were sent electronically on a spreadsheet. Initially, there was some level of inaccuracy or imprecision. This ranged from an inability to add the four subscales correctly, to reading off and recording the wrong stanine for the score and year, to failing to provide exact details of students at Time 1 and Time 2 to enable matching, to providing only one or two subtest scores and neglecting to indicate whether others were administered and zero scored or whether they were not administered. As an extreme example, incorrect addition and entering the wrong scores, including stanine, accounted for 16 errors within one class in a school. In some schools, there were entire classes with missing data. If we did not attempt to check and rectify these errors and omissions we stood to lose around one third of our sample at Time 1 alone.

This lack of facility in dealing with data on the part of most schools is possibly indicative of the level of experience within schools in dealing with data and the state of knowledge of data use amongst teachers, a point taken up in a later chapter of this report. By Time 2 data collection, these issues were largely resolved.

**Overall Trends in Student Achievement for Reading and Writing**

The most efficient (and parsimonious) way to report the progress schools make with regard to student achievement is to use the measure of effect size. The use of effect size provides a common metric (expressed in standard deviation units) to measure and compare the effect of any or many interventions that may use different outcome measures. The effect indicates the amount of gain. A rigorous way to gauge the effect of any large-scale intervention would be to consider the amount of gain over and above what might be expected at the 50th percentile, which would allow one to say with confidence whether or not the intervention had any effect on student achievement over and above progress under normal conditions. Calculation of this effect size involves taking the difference between the Time 2 score and the population mean for that time, and dividing by the population standard deviation. However, given that there are only 13 schools with at least two data points in this sample and that these schools may not be a representative sample of New Zealand schools, it is not considered appropriate to emphasize this method of calculation of effect size.

Instead, we utilise an effect size that relates to the extent to which student achievement in the school has moved, relative to where it started. This involves taking the difference between mean scores at two points in time (between, for example, Time 1 and Time 2 for those students who completed both tests) and dividing by the average of the standard deviations at those times. This gives a measure of the gain relative to starting point and allows one to view both the extent of gain and the range of gain across schools in the sample. For STAR this effect has been calculated using raw scores. Note that, when considering progress of schools with students taking STAR tests with different maximum total scores or when considering the effect size gain for
individual classes, we used growth in raw score as a proportion of possible total score (as STAR tests for year groups are out of different totals).

The effect size gains for all 13 research schools, both reading and writing, are shown in Table 2.2. The table shows gains for Time 1 to Time 2 and then Time 3 to Time 4. It also indicates the sample size on which the effect size was calculated as, in some cases, these are very small. Further, where reading schools administered both STAR and asTTle in the same time period (Time 3 to 4 for cohort 1 and Time 1 to 2 for cohort 2) the effect size gain is shown for both tests.

These analyses show mixed results, with some schools making significant gains overall while others make small gains; the range in effect size was from 0.105 to 2.756. According to Cohen (1977) (Cohen’s $d$, called after the statistician who devised it, is the formula used to calculate an effect size), an effect size of 0.8 and above is a large effect (0.2 – 0.49 is small; 0.5 – 0.79 is medium).

For the research schools in reading (schools 1, 2, 3, 4, 5, 8, 9) the mean effect size gain for Time 1 to Time 2 using STAR was 0.463 and ranged from 0.291 to 0.691. For the period Time 3 to Time 4, the effect size gains averaged 0.488. In three of the four schools on the project for two full years, the effect size was greater between Times 3 and 4 than between Times 1 and 2, suggesting that the professional learning project may have taken some time to gain traction in these schools. In the national data, the range in mean effect size for year levels (there is no comparable analysis by school) for reading over two years was 0.54 to 1.12 with a mean effect size of 0.87.

Where the research schools that focused on reading administered asTTle, over a comparable time period to STAR, the effect sizes, in general, tended to be smaller. There were only two reading schools who administered both tests at Time 1 and 2 and, for these schools, the average effect size for asTTle Time 1 to Time 2 was 0.463 and for STAR, 0.523. For the Time 3 to Time 4 period, again only two schools have data for both tests at comparable times. The average effect size for asTTle was 0.455 and for STAR 0.536. For the one school that had an effect size gain calculated from STAR for year 1 on the project and asTTle for year 2, the effect size for STAR was 0.402 but, for asTTle, it was 0.849. There are some possible explanations for the different effect size gains. First, the ceiling effects for STAR (see Chapter 3) suggest that the test is easy and so, although high scoring students are constricted, lower scoring students are able to make quite large gains. The asTTle test was much more difficult for students (see graphs in Chapter 3 showing the gap between the top scores at Time 1 and Time 2 and the maximum possible score). Also, arguably, despite asTTle being curriculum referenced, it is much more difficult with asTTle to ‘teach to the test’ than with STAR where we had reported instances of students practicing cloze pieces. While practice effects of STAR are claimed to be low, two parallel forms appear to yield a more similar test than that generated randomly from the extensive item bank of asTTle.
Table 2.2
Effect Size Gains in Reading and Writing Schools for Year 1 (T1-T2) and Year 2 (T3-T4) of the Project.

<table>
<thead>
<tr>
<th>School</th>
<th>Effect size</th>
<th>Test</th>
<th>N</th>
<th>Effect size</th>
<th>Test</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.576</td>
<td>STAR</td>
<td>213</td>
<td>0.68</td>
<td>STAR</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.483</td>
<td>asTTle</td>
<td>152</td>
</tr>
<tr>
<td>2</td>
<td>0.291</td>
<td>STAR</td>
<td>268</td>
<td>0.476</td>
<td>STAR</td>
<td>257</td>
</tr>
<tr>
<td>3</td>
<td>0.621</td>
<td>STAR</td>
<td>90</td>
<td>0.392</td>
<td>STAR</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.427</td>
<td>asTTle</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>0.402</td>
<td>STAR</td>
<td>269</td>
<td>0.849</td>
<td>asTTle</td>
<td>266</td>
</tr>
<tr>
<td>5</td>
<td>0.305</td>
<td>STAR</td>
<td>399</td>
<td>0.406</td>
<td>STAR</td>
<td>316</td>
</tr>
<tr>
<td>6</td>
<td>1.124</td>
<td>asTTle</td>
<td>82</td>
<td>0.265</td>
<td>asTTle</td>
<td>106</td>
</tr>
<tr>
<td>7</td>
<td>0.105</td>
<td>asTTle</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.306</td>
<td>STAR</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.451</td>
<td>asTTle</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.355</td>
<td>STAR</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.476</td>
<td>asTTle</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.652</td>
<td>asTTle</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.703</td>
<td>asTTle</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2.756</td>
<td>asTTle</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.802</td>
<td>asTTle</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: School (4) STAR T3-T4 not included as only T3 data for new intake.

To illustrate these effect sizes in a more familiar way, the mean stanine for reading at each of the time points is shown in Table 2.3 (the N is the students with all data points). Figure 2.1 illustrates growth for the reading students over time although school 1 shows a discernible dip between the end of year in 2004 and the beginning of 2005.

Table 2.3
STAR Mean Stanine at Beginning and End of Years 1 and 2 by School

<table>
<thead>
<tr>
<th>School</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.25</td>
<td>5.99</td>
<td>5.20</td>
<td>6.06</td>
<td>119</td>
</tr>
<tr>
<td>2</td>
<td>5.42</td>
<td>5.33</td>
<td>5.32</td>
<td>5.79</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>3.66</td>
<td>4.11</td>
<td>4.13</td>
<td>4.71</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>5.24</td>
<td>5.75</td>
<td>5.83</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.44</td>
<td>5.69</td>
<td>5.94</td>
<td>5.86</td>
<td>154</td>
</tr>
<tr>
<td>8</td>
<td>4.52</td>
<td>4.75</td>
<td></td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.19</td>
<td>5.24</td>
<td></td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

Note: School 5 was the school not continuing on the project.
Figure 2.1
Mean Stanine at Beginning and End of Years 1 and 2 by School

Note: School 5 was the school not continuing on the project

For the six writing schools (schools 6, 7, 10, 11, 12, 13), as shown in Table 2.2, the effect size gains averaged 1.19 for the first year of the project (range 0.105 to 2.756). In the national data, the effect size for the cohort over two years was 1.28. The much larger effect size for writing compared to reading is probably partly a function of the relatively low level of writing achievement as reflected in the norms. The mean total score for each time point for each school is shown in Table 2.4. We have data over two years for only one school where the gains were large in the first year, showed the often observed decline over the summer break (not to the level of Time 1), then rose again at Time 4 but not to the level of Time 2. This is shown in Figure 2.2.

Table 2.4
Mean Total Writing Score at Beginning and End of Years 1 and 2 by School

<table>
<thead>
<tr>
<th>School</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>451</td>
<td>519</td>
<td>468</td>
<td>486</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>469</td>
<td>477</td>
<td></td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>10</td>
<td>381</td>
<td>434</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>397</td>
<td>456</td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>12</td>
<td>400</td>
<td>502</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>420</td>
<td>478</td>
<td></td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>
Student Achievement by Ethnic Group and by Gender

This section presents data in terms of ethnic group affiliation and gender. The data are only the Time 1 to Time 2 data as, by Time 3 and 4, the numbers in the various ethnic groups have become too small for meaningful analyses.

By ethnic group

The reason for such analyses is that previous analyses of literacy achievement (e.g. PIRLS, Ministry of Education, 2003) have shown that Maori students, particularly Maori boys, perform below average. Pasifika students are also over-represented in the lower quartile, that is, the percentage of their total test population in the lower quartile is discrepant from New Zealand norms. Maori and Pasifika, together with those from non-English speaking backgrounds, are generally taken to represent those groups targeted in terms of efforts to raise achievement. The question of interest here is whether the project is benefiting targeted student groups.

Students who sat asTTle writing tests nominated their ethnic identification according to the categories available in asTTle (NZ European/ Pakeha, NZ Maori, Pasifika, Other), while the reading schools that employed the STAR test supplied ethnicity data from their records. The latter generally employed the Ministry of Education’s categorisation, however, given the relatively small sample size of some ethnicities, these data were grouped into the same four categories as in asTTle3. We prioritised a consideration of the data from groups of specific interest with respect to literacy achievement and, therefore to the project and the Ministry, namely Maori and Pasifika. The analyses are reported first for reading and then for writing.

---

3 NZ European/ Pakeha = NZ European/ Pakeha; NZ Maori = Maori; Pasifika = Samoan, Tongan, Fijian, Niue, Cook Island Maori; Other = Indian, Chinese, Japanese, Korean, Other Asian, Other European, Other.
Reading: The total sample of students with data at both points in time, from the five schools in the first intake that focussed on reading comprehension, was 1201. The students, by ethnic groups and by year are shown as a percentage of a year cohort in Table 2.5. Note the small numbers of Pasifika students in Years 4-6 where they account for only two to four percent of the student population.

Table 2.5a
Ethnic Group as a Percentage of Total Students by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Ethnicity</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NZ European/Pakeha</td>
<td>42</td>
<td>47</td>
<td>49</td>
<td>63</td>
<td>42</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Maori</td>
<td>26</td>
<td>24</td>
<td>24</td>
<td>22</td>
<td>18</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Pasifika</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>20</td>
<td>24</td>
<td>23</td>
<td>11</td>
<td>26</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2.5b
Sample Size of Ethnic Group by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Ethnicity</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NZ European/Pakeha</td>
<td>41</td>
<td>49</td>
<td>55</td>
<td>75</td>
<td>143</td>
<td>215</td>
<td>578</td>
</tr>
<tr>
<td></td>
<td>Maori</td>
<td>25</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>61</td>
<td>74</td>
<td>238</td>
</tr>
<tr>
<td></td>
<td>Pasifika</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>50</td>
<td>59</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>19</td>
<td>25</td>
<td>26</td>
<td>13</td>
<td>89</td>
<td>113</td>
<td>285</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>97</td>
<td>104</td>
<td>113</td>
<td>119</td>
<td>343</td>
<td>461</td>
<td>1237</td>
</tr>
</tbody>
</table>

To obtain a broad indication of whether the relative position of various ethnic groups changed during their participation in the professional learning project, the stanine distribution is explored. If student achievement remained at the same relative level at both points in time, the stanine would stay the same. So, any increase in stanine marks improvement over previous achievement levels.

The distribution of stanine scores by ethnic group across all years at Time 1 and Time 2 is shown in Figure 2.3. The graphs suggest that only NZ European and Other seem to have a raised stanine median score over time. The relative standing of ethnic groups remains largely the same over time except that the median stanine for Other, which at Time 1 was equivalent to Pasifika, has surpassed the Pasifika median stanine at Time 2. The distribution of scores in the lower quartile for Maori students has narrowed between Time 1 and Time 2. The distribution of scores for NZ European appears to have widened, due largely, it seems, to the distribution shifting upwards. For the Other group, the upper quartile has narrowed and the lower extended; a reversal of Time 1 as the median shifts upwards.
A test of whether the performance, as indicated by mean stanine, differed at Time 1 shows that there was a significant difference in performance by ethnic group ($F(3, 1197) = 67.38$, $p < .001$). Multiple comparisons (Bonferroni) showed that the difference is accounted for by NZ Europeans having a significantly different (higher) mean stanine than other ethnicities (mean stanines were: 6.10 (SD=1.91) for NZ European, 4.64 (SD=1.95) for Maori, 4.18 (SD=1.76) for Pasifika, and 4.55 (SD=2.16) for Other). There were no other significant differences between groups.

At Time 2, there was also a significant difference in performance across groups ($F(3, 1197) = 71.67$, $p < .001$). Multiple comparisons showed that, again, NZ European had a significantly different mean stanine from the other groups with Maori also having a significantly higher mean stanine that Pasifika (mean stanines were: 6.43 (SD=1.81) for NZ European, 5.05 (SD=1.99) for Maori, 4.40 (SD=1.66) for Pasifika, and 4.89 (SD=2.15) for Other).

Further analysis considered how the stanine distribution profile of each of the groups accords with the normal distribution, the New Zealand average 'profile'. The STAR data normative stanine profile is that, on average, 23% of students fall into stanines 1 to 3; 54% in the average band of stanines 4 to 6, and 23% in stanines 7 to 9. The following bar graphs show the stanine band profiles and the variations from the NZ
mean of the various ethnic groups at Time 1 (Figures 2.4 and 2.5) and Time 2 (Figures 2.6 and 2.7). The profiles most discrepant at Time 1 from this New Zealand average profile are the NZ European where most students were in the average band or above (only around 10% in the lower band) and the Pasifika where only 10% are in the upper band. At Time 2, the NZ European distribution has become even more positively skewed with an increased proportion in the upper band. The Pasifika group have lowered the proportion in the lower band at Time 2 with the increase showing in the average band but with virtually no change in the upper band. Other shows a drop in the percentage in the below average band and an increase in the other two bands between Time 1 and Time 2. Maori show a decrease in the below average and average band and a consequent increase in the percentage in the upper band to be almost equivalent to the New Zealand average at Time 2.

**Figure 2.4**

*Percentage in Stanine Bands by Ethnic Group at Time 1*
Figure 2.5
Percentage Difference of each Ethnic Group from the NZ Average at Time 1

Figure 2.6
Percentage in Stanine Bands by Ethnic Group at Time 2
As noted earlier, we were particularly interested in the relative progress of the various groups and whether there was any significant difference in rate or nature of progress. All ethnic groups made progress. Paired comparisons established that all groups made significant improvements from Time 1 to Time 2 in total score (the t values are shown in Table 2.6 below). These mean total scores should only be viewed within a group as, with STAR, the maximum total score varies with year level test and the relative proportions of different ethnic groups vary across year levels (as shown above in Table 2.5).

### Table 2.6
**Paired Sample t test of Mean Total Score at Time 1 and Time 2 by Ethnic Group**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean Time 1</th>
<th>Mean Time 2</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European/ Pakeha</td>
<td>50.63</td>
<td>55.23</td>
<td>-18.70</td>
<td>567</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Maori</td>
<td>39.83</td>
<td>45.59</td>
<td>-12.61</td>
<td>228</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pasifika</td>
<td>41.23</td>
<td>46.53</td>
<td>-8.60</td>
<td>130</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Other</td>
<td>41.59</td>
<td>47.55</td>
<td>-13.54</td>
<td>272</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

More significantly, as there are no norms to indicate expected gain in total score, all ethnic groups made significant gains in stanines (p < .05). Table 2.7 shows the stanine means for each group at Times 1 and 2 and the t values associated with a paired sample test of mean difference. An analysis of variance showed that there was no difference in the stanine gains made by each ethnicity (F(3, 1197) = .82, p > .05) so groups had similar rates of progress.
Table 2.7
Paired Sample t-test of Stanine Mean at Time 1 and Time 2 by Ethnic Group

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean</th>
<th>Mean</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NZ European/Pakeha</td>
<td>6.10</td>
<td>6.43</td>
<td>-7.11</td>
<td>567</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Maori</td>
<td>4.64</td>
<td>5.05</td>
<td>-5.09</td>
<td>228</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pasifika</td>
<td>4.18</td>
<td>4.40</td>
<td>-2.61</td>
<td>130</td>
<td>0.01</td>
</tr>
<tr>
<td>Other</td>
<td>4.55</td>
<td>4.89</td>
<td>-5.22</td>
<td>272</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

With respect to how each group fared relative to an expected zero change, all ethnic groups experienced an increase in the percentage of students in the upper stanine band (7-9) with NZ European/Pakeha increasing by around 11 percent and Pasifika with the smallest gain at around one percent. Likewise, all ethnic groups – especially Pasifika and Other - decreased their percentage of students in the below average stanine band of 1 to 3 (see Figure 2.8).

Figure 2.8
Percentage Stanine Difference by Band from Time 1 to Time 2

Although it is problematic to consider the performance on the various sub tests of STAR, given ceiling effects and large standard deviations in the normative data, the tests are nevertheless a useful way to get a rough idea of whether the gain in performance is similar on the different aspects of reading they represent. The difference between Time 2 and Time 1 scores on sub tests are considered for each group by year/test in the graphs below. Bear in mind that, for some groups, notably NZ European, who were already scoring high
Literacy Professional Development Project: Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning

on average on some of the subtests, there was less room for improvement. Also, note that the sample size of ethnic groups at some year levels is very small.

The results for Year 3 are illustrated in Figure 2.9 which shows that Other make the least gain compared to other groups in paragraph comprehension but the most in vocabulary. Maori students make the most improvement compared to others in three of the four subtests, namely, in word recognition, sentence comprehension and paragraph comprehension.

**Figure 2.9**
Year 3 Difference Between Time 2 and Time 1 for Each Subtest by Ethnic Group

At Years 4-6 shown in Figure 2.10, Maori again make the greatest gains except in vocabulary where NZ European gain most. Finally, at Years 7-8 the greatest gains would seem to have been made by Other students, although Pasifika make the most gains in paragraph comprehension (see Figure 2.11).
Literacy Professional Development Project: Identifying Effective Teaching and Professional Development Practices for Enhanced Student Learning

Figure 2.10
Year 4-6 Difference Between Time 2 and Time 1 for Each Subtest by Ethnic Group

![Bar chart showing differences between Time 2 and Time 1 for each subtest by ethnic group for Year 4-6.](image)

Figure 2.11
Year 7-8 Difference Between Time 2 and Time 1 for Each Subtest by Ethnic Group

![Bar chart showing differences between Time 2 and Time 1 for each subtest by ethnic group for Year 7-8.](image)

Writing: The total sample of students with data of established reliability at both points in time, from the two schools in the first intake that focused on writing was 278. Note that we report data only for students whose
writing was scored by the research team at both points in time. Table 2.8 shows the size of the sample by ethnic group and year. Ethnic group affiliation is not evenly distributed across the years. There are a lower than expected number of Maori in Years 4-6 and higher numbers than expected in Years 7-8. The reverse is true for the Other group.

This configuration is a result of the character of the two schools that are represented in our data to date. The low numbers necessitate that for the purposes of analysis the years be aggregated at this point in time (which, fortunately, with a common scale used in asTTle, it is feasible to do). The fact that we report analyses for only two schools is problematic as the nature of the school and its progress in terms of the professional learning contract become a potentially powerful mediator. A much larger sample size in terms of schools is necessary. Unfortunately, the total numbers in schools in our second cohort was low and these schools were far less diverse than cohort 1.

### Table 2.8
Sample by Ethnic Group and Year

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Year</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European Pakeha</td>
<td></td>
<td>16</td>
<td>10</td>
<td>6</td>
<td>51</td>
<td>54</td>
<td>137</td>
</tr>
<tr>
<td>Maori</td>
<td></td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>36</td>
<td>31</td>
<td>72</td>
</tr>
<tr>
<td>Pasifika</td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>13</td>
<td>8</td>
<td>19</td>
<td>10</td>
<td>9</td>
<td>59</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>30</td>
<td>23</td>
<td>29</td>
<td>100</td>
<td>96</td>
<td>278</td>
</tr>
</tbody>
</table>

The median scores and range for each group are shown at both time points in Figure 2.12. At Time1 ranges are common across groups except for the small Pasifika group where there is a truncated range.
There appear to be little differences in mean score at Time 1. Multiple comparisons with Bonferroni correction show no significant differences across groups. However, a test with a less strict criterion (Games-Howell) suggests that NZ European score higher than Pasifika (p < .05) and that the group, Other had scores significantly higher than Pasifika (p < .05) (see Table 2.9).

**Table 2.9**

*Multiple Comparisons (Games-Howell) of Total Score by Ethnic Group at Time 1*

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Ethnicity</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European/ Pakeha</td>
<td>Maori</td>
<td>19.46</td>
<td>8.99</td>
<td>.139</td>
</tr>
<tr>
<td></td>
<td>Pasifika</td>
<td>49.16</td>
<td>12.40</td>
<td>.008*</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>9.47</td>
<td>9.15</td>
<td>.730</td>
</tr>
<tr>
<td>Maori</td>
<td>Pasifika</td>
<td>29.70</td>
<td>13.62</td>
<td>.166</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>-9.99</td>
<td>10.75</td>
<td>.789</td>
</tr>
<tr>
<td>Pasifika</td>
<td>Other</td>
<td>-39.69</td>
<td>13.73</td>
<td>.043*</td>
</tr>
</tbody>
</table>
At Time 2, the range of scores has extended and an analysis of variance established that there was a significant difference between groups ($F(3, 274) = 9.08, p < .001$). The results of post-hoc multiple comparison (Games Howell) tests show NZ European to be different to Maori and Other to be different to Maori (see Table 2.10).

### Table 2.10
Multiple Comparisons (Games-Howell) of Total Score by Ethnic Group at Time 2

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Ethnicity</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European/ Pakeha</td>
<td>Maori</td>
<td>43.49</td>
<td>12.79</td>
<td>.005*</td>
</tr>
<tr>
<td></td>
<td>Pasifika</td>
<td>64.54</td>
<td>31.30</td>
<td>.230</td>
</tr>
<tr>
<td>Maori</td>
<td>Pasifika</td>
<td>21.04</td>
<td>32.33</td>
<td>.913</td>
</tr>
<tr>
<td>Other</td>
<td>Pasifika</td>
<td>-67.77</td>
<td>14.79</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>-88.82</td>
<td>32.17</td>
<td>.075</td>
</tr>
</tbody>
</table>

The extent of improvement within each ethnic group over time is shown in Figure 2.13. An analysis of variance showed an effect of time on scores ($F(1, 274) = 15.62, p < .001$). Post-hoc tests showed that only NZ European and Other made significant improvements ($t = 4.52$ and $6.7$, respectively, $p < .001$). Maori and Pasifika appear to make little progress but again, our data are drawn only from two schools so the character of the school and its overall progress on the professional development project are a significant factor.

### Figure 2.13
Total Score at Time 1 and Time 2 for Ethnic Groups
A consideration of the relative progress of each group in surface and deep features shows that similar gains were made in surface and deep features for all groups except Other where 76 points are gained in surface, compared to 48 in deep. All ethnic groups have higher scores in deep features than surface features except Other at Time 2. These differences in favour of deep features are significant except for Pasifika due in part to a low sample size. (see Tables 2.11 and 2.12).

Table 2.11
Paired Sample t-test of asTTle Surface Score and Deep Score at Time 1

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European/Pakeha</td>
<td>-22.93</td>
<td>-4.66</td>
<td>136</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Maori</td>
<td>-21.28</td>
<td>-3.53</td>
<td>71</td>
<td>.001</td>
</tr>
<tr>
<td>Pasifika</td>
<td>-27.41</td>
<td>-1.63</td>
<td>9</td>
<td>.138</td>
</tr>
<tr>
<td>Other</td>
<td>-20.61</td>
<td>-3.12</td>
<td>55</td>
<td>.003</td>
</tr>
</tbody>
</table>

Table 2.12
Paired Sample t-test of asTTle Surface Score and Deep Score at Time 2

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ European/Pakeha</td>
<td>-21.56</td>
<td>-4.67</td>
<td>136</td>
<td>.000</td>
</tr>
<tr>
<td>Maori</td>
<td>-20.25</td>
<td>-3.29</td>
<td>71</td>
<td>.002</td>
</tr>
<tr>
<td>Pasifika</td>
<td>-28.26</td>
<td>-1.29</td>
<td>9</td>
<td>.230</td>
</tr>
<tr>
<td>Other</td>
<td>6.60</td>
<td>1.17</td>
<td>58</td>
<td>.248</td>
</tr>
</tbody>
</table>

When considering the curriculum functions in more detail, Other students make the most gains in all dimensions of writing except audience where Pasifika students improve the most. The dimension where every group made the most progress was structure, although only the Maori group and NZ European group made significant improvements ($t = 8.87$ and $t = 12.30$ respectively, $p < .001$). The area of grammar was where all made the least progress. The small sample size for Pasifika means that their gains were unlikely to be statistically significant. Again, with only two schools there is the possibility of a confounding effect of school. In the school where significant progress was made in writing, there was no difference in progress by ethnic group (see Figure 2.14).
Figure 2.14
Total Score at Time 1 and Time 2 by Ethnic Group for One High Progress School
By Gender

Reading: The distributions of stanine scores for males and females were very similar. The only apparent difference was a slightly higher median for females at Time 1 (see Figure 2.15). Significance tests confirm that no significant differences (p < .05) exist between genders at Time 1 or Time 2. Males and Females both made significant gains in stanine scores between Time 1 and Time 2 (t = -7.76, p < .001 for males; t = -7.20, p < .001 for females). The rate of gain was not significantly different (p < .05) between genders.

Figure 2.15
Distribution of Stanine Scores by Gender at Time 1 and Time 2 (Reading)

Writing: The distribution for males and females seems similar, but females appeared to have slightly higher medians especially at Time 2 (Figure 2.16). Both genders exhibited a widening of the distribution of asTTle score at Time 2. The difference between genders was significant at both time points (t = 2.37, p = .018 at Time 1; t = 3.03, p = .002 at Time 2) with females outperforming males.

Both males and females made significant gains in asTTle writing score from Time 1 to Time 2 (t = 3.29, p = .001 for males; t = 5.877, p < .001 for females). Females made more gain than boys with a mean difference between Time 1 and Time 2 of 34 asTTle points compared to the male gain of 19 asTTle points, however, this difference in gain was not significant (p < .05).
Conclusions

Students made moderate progress in reading with an average effect size of 0.46 in the first year of the project and 0.49 in the second year. In writing, there was a large average effect size gain of 1.28. While the average effect sizes are a little lower than that for the national cohort, the relativity between gain in reading and writing is similar.

With respect to differential achievement by ethnic group, in reading, there was a difference in performance at both points in time. However, all groups made significant progress in terms of gains in total score and stanine level from Time 1 to Time 2. There was no difference in extent of gain. However, the stanine distributions resulting at Time 2 showed the NZ European distribution to be top heavy with disproportionate numbers in the highest stanines; the distribution for Maori to resemble the NZ normal distribution and that for Pasifika to have large numbers in the 4-6 band and relatively few in the highest stanines.

In writing the median score at Time 1 was barely different across ethnic groups with NZ European and Other significantly higher than Pasifika. At Time 2 there was a significant difference between groups with both NZ
European and Other significantly higher than Maori or Pasifika. They appeared to make significant progress where Maori and Pasifika did not. However, this may be a function of a very small sample of schools as in the school that made excellent progress, all groups progressed equally.

There were no significant differences by gender in reading achievement scores in this sample or differences in rate of progress. In writing, however, females scored significantly better at both points in time but both made significant progress and there was no difference in the rate of progress.

**References**


Chapter 3

Issues Associated with Tests of Student Achievement in Reading and the Use of These Data to Inform Decisions

National Guidelines expect that schools will have goals and targets for student learning outcomes, specifically regarding improved performance and that they will report the extent to which they have made progress towards these goals. For most schools this now involves an examination of student achievement data. This chapter raises some issues associated with efforts to track progress in reading. Historically, in the research literature, “very little has been reported about actual changes in student performance from year to year, let alone what changes are possible, or even likely, with different educational initiatives” (Yin & Brennan, 2002, p. 83). Yin and Brennan (2002) suggest that one reason might be that testing programs do not routinely report change at the student level or even at a more aggregated level. However, the other reason, they suggest, may be that literature has been critical of the use of difference scores to track progress.

There are a number of issues associated with the effort to evaluate performance in a chosen area. First there is the ongoing debate about whether growth can be demonstrated using the difference for individuals in performance at points in time. Then, potentially, there are issues with the instrument used to measure and whether it can indeed both measure what is required and show trends and rates of progress absolutely and normatively.

Use of Gain Scores to Measure Progress

How to best ascertain the effect of an intervention or progress towards goals is not straightforward. Historically, the measurement of change in the educational research literature has been an area of debate. The measurement of potential educational outcomes focuses on true difference scores, estimated using observed scores. In classical test theory the observed score is the true score plus measurement error. A simple difference score is generally what is used as a direct measure of change. There are two basic criticisms of the use of difference scores based on notions of unreliability or unfairness.

The first criticism concerns the fact that, based on classical test theory, the reliability estimate for the difference scores tends to be much lower than that for the component parts/scales (Cronbach & Furby, 1970). Even when the correlation co-efficient between the tests is relatively large and positive, the reliability of the

---

4 If one were to administer the same test to a person 100 times, because of chance errors, these scores would form a normal distribution around the true score. So the mean of this distribution is take as the true score and the standard deviation as the error of measurement. In actual practice we do not have true scores only scores from a single administration so the standard error of measurement is used to estimate the reasonable limits of the true score for persons with any given observed score.
difference scores tends to be smaller than the reliability coefficients for the component tests. So, there is some question around the potential low reliability of difference scores.

The second concern is that difference scores may not measure progress fairly across the distribution. The correlation between the obtained difference scores and where students were at Time 1, is often negative (Cronbach & Furby, 1970; Linn & Slinde, 1977). So, students with high Time 1 scores will, on average, have lower change scores than students with lower Time 1 scores. This negative correlation is mostly explained by the statistical phenomenon of regression to the mean. This is solely a statistical phenomenon, independent of any treatment effect that might occur. It happens when repeated measurements are made on the same variable and collected from the same individual. It is due to the properties of conditional expectation (this is the expectation given some other event has already occurred). The logic is that measurements that are far removed from the mean represent relatively rare events and the farther away, the rarer. Because rare events tend not to happen repeatedly over time, repeated measurements mean it becomes less rare and consequently closer to the mean (Bonate, 2000). Thus, difference scores confer advantages or disadvantages, based on Time 1 scores, and so do not allow comparison of change at different points on a common scale.

The question remains as to how to measure change or growth. Some (e.g. Cronbach & Furby, 1970) simply advise that different questions be asked! Others attempt to use measures of change that are ‘base free’, that is they use residualized difference scores or a regressed-score estimate of change (McNemar, 1956; Brennan, 2001). These make predictions of change based on regression (Feldt & Brennan, 1989).

However, recent work has questioned some assumptions that have underpinned the traditional negative views of difference scores. They have shown that difference scores can be reliable. One basis for the challenge concerns the assumption that underpins obtaining a conventional reliability coefficient, namely that the variance in the same test will be equal. An assumption of homogenous change means that there are not individual differences in growth so the variance of the true difference scores is near to zero. But, homogeneous change is rarely the case. As students get older, they tend to become more dissimilar so the variability of difference scores may change over time. So, in obtaining the reliability for difference scores, the assumption of equal variances may not be met. Thus, the low reliability of difference scores may be a result of large measurement error or a lack of variance in the true change (Yin & Brennan, 2002). And a small variance in true difference scores does not mean that the difference in a person’s true score from Time 1 to Time 2 is small (as the mean true difference score [mean T2 – mean T1] is a constant that gets subtracted from every person’s true difference score in the traditional reliability formula so whether the constant is small or large has no bearing on the magnitude of the variance in true difference scores). If the aim is to make some statement about the magnitude of change for individuals, then there is a need to measure the extent to which actual change scores are dependable.

---

5 This is because the classical formula for calculating reliability of difference scores is highly sensitive to the reliability of the test. It has as its numerator the average of the reliability coefficients for the component tests, minus the correlation between scores on two administrations of the test, that is the correlation between scores at the two points in time. The denominator is 1 minus the correlation between Time 1 and Time 2. For STAR, for example, the split half reliability is reported as 0.91, minus the correlation between T1 and T2 administrations, which was 0.89 in our data. So, $\frac{0.02}{1-0.89} = 0.18$. 
When an alternative formula (one that does not contain the traditional restrictions that are imposed for deriving a reliability coefficient for difference scores) is used to calculate the reliability of difference scores (Rogosa & Willett, 1983; Zimmerman & Williams, 1982b), reliability coefficients for these difference scores can be quite high (Rachor & Cizek, 1996; Williams et al., 1984; Zimmerman & Williams, 1982b, cited in Yin & Brennan, 2002).

There is, likewise, a challenge to the notion that difference scores are inherently unfair. Zimmerman and Williams have shown that the correlation between change and initial status can be positive, zero or negative, largely depending on the ratio of the standard deviations of the two tests. So, the traditional view that the correlation between Time 1 score and the difference score is always negative (regression to the mean) is incorrect because this is based on the assumption of equal variances for the two tests.

In a large-scale study of gains in the subtests of the Iowa Test of Basic Skills, Yin and Brennan (2002) showed the resulting calculations of difference score reliability using a traditional calculation of reliability and Miller and Kanes’ (2001) absolute-change “reliability” formula. While traditional reliability of difference scores was moderate (around 0.6), the absolute-change reliability was noticeably larger (around 0.8).

Yin and Brennan (2002) suggest that the literature around measurement of change using difference scores has focused on reliability-related issues, often based on theory rather than empirical data. Their conclusion is that difference scores can be reliable and are potentially very useful. Now, many researchers take the view that a gain score for an individual is useful, according to Zimmerman and Williams “at least as meaningful as other measures that contain a large error component” (1982b, p. 154, cited in Yin & Brennan, 2002).

**The Tools Selected and Their Ability to Measure Progress**

Appropriate tools need to be selected that are able to measure this progress in order to track progress over time and to view this progress relative to normative expectations or particular goals. As a tool, STAR allows performance to be viewed relative to normative performance levels at various points in time (for at least two points in the school year) for the years the test covers (3-9). However, there are no normative data for progress of individuals and groups. In the course of a school year, the retaining of the same stanine by an individual indicates that the student is continuing to progress at the same rate. Viewing the proportions of students in each stanine at different points in time allows changes to be tracked at an aggregate level.

Given that the ability to track progress accurately was a major consideration in this project, a decision was made for 2005 to trial the use of asTTle Reading. Also, given the project aimed to help teachers to use data to inform their instruction, we were keen to explore the diagnostic features of asTTle as a curriculum referenced test. In 2005, students in the reading research schools sat both STAR and asTTle Reading at the beginning and end of the year. We wanted to gather some data on how the same students performed on both STAR and asTTle Reading and how each test performed with respect to showing progress. Table 3.1 shows the number of students who had scores for both tests at two points in time. It also shows the sample size of students who, in 2004, completed a STAR test at two points in time.
Each test had good external reliability with moderate to strong correlations between time points for individuals. STAR scores at Time 1 correlated 0.89 with STAR scores at Time 2 for Years 4 to 6. The comparable figure for Years 7 and 8 was 0.907. The correlation between time points for asTTle was 0.779 for Years 4 to 6 and 0.746 for Years 7 and 8. There were moderate to strong correlations between the STAR and asTTle scores of a student at both time points. For Years 4 to 6 at Time 1 the correlation was 0.81 and, at Time 2, 0.77. For Years 7 and 8 the comparable correlations were 0.83 and 0.82 (See Tables 3.2 and 3.3).

**Table 3.2**
Relationship Between asTTle and STAR Scores (Years 4-6 N=198)

<table>
<thead>
<tr>
<th></th>
<th>aTotal1</th>
<th>aTotal2</th>
<th>aTotalDiff</th>
<th>sTotal1</th>
<th>sTotal2</th>
</tr>
</thead>
<tbody>
<tr>
<td>aTotal1</td>
<td>Pearson Correlation</td>
<td>.779(**)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>aTotal2</td>
<td>Pearson Correlation</td>
<td>-.399(**)</td>
<td>.265(**)</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>aTotalDiff</td>
<td>Pearson Correlation</td>
<td>.806(**)</td>
<td>.761(**)</td>
<td>-.126</td>
<td>.076</td>
</tr>
<tr>
<td>sTotal1</td>
<td>Pearson Correlation</td>
<td>.784(**)</td>
<td>.772(**)</td>
<td>-.077</td>
<td>.890(**)</td>
</tr>
<tr>
<td>sTotal2</td>
<td>Pearson Correlation</td>
<td>-.284(**)</td>
<td>-.209(**)</td>
<td>.131</td>
<td>-.509(**)</td>
</tr>
<tr>
<td>sTotalDiff</td>
<td>Pearson Correlation</td>
<td>.000</td>
<td>.000</td>
<td>.281</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).**
Table 3.3: Relationship Between asTTle and STAR Scores Years 7 & 8 (N=186)

<table>
<thead>
<tr>
<th></th>
<th>asTTle1</th>
<th>asTTle2</th>
<th>asTTleDiff</th>
<th>STAR1</th>
<th>STAR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>asTTle1</td>
<td>Pearson Correlation</td>
<td>0.746(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>asTTle2</td>
<td>Pearson Correlation</td>
<td>-.132</td>
<td>0.561(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.072</td>
<td>0.561(**)&lt;br&gt;Sig. (2-tailed)</td>
</tr>
<tr>
<td>asTTleDiff</td>
<td>Pearson Correlation</td>
<td>-.132</td>
<td>0.561(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.072</td>
<td>0.561(**)&lt;br&gt;Sig. (2-tailed)</td>
</tr>
<tr>
<td>STAR1</td>
<td>Pearson Correlation</td>
<td>0.825(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.607(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.000</td>
</tr>
<tr>
<td>STAR2</td>
<td>Pearson Correlation</td>
<td>0.805(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.607(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.000</td>
</tr>
<tr>
<td>STARDiff</td>
<td>Pearson Correlation</td>
<td>-0.252(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.001</td>
<td>0.607(**)&lt;br&gt;Sig. (2-tailed)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

The Issue of Lack of Correlation of Difference Scores

The correlation of the asTTle difference score with the STAR difference score (Time 2 - Time 1) was also calculated for Years 4-6. What was most striking from this exercise was the lack of correlation of the difference scores. The difference score of STAR and asTTle correlated only 0.13 at Years 4-6 and 0.009 at years 7 and 8. This raises the issue of the reliability of the differences scores. The reliability of the difference scores was tested using the formula (Yin & Brennan, 2002).

\[
\frac{\sigma^2_{1T} + \sigma^2_{2T} - 2\rho_{1T2}^2 \sigma^2_{1T2}}{\sigma^2_{1T1} + \sigma^2_{2T2} - 2\sigma^2_{1T2}}
\]

Where \(\sigma^2_{T1}\) is the variance at Time 1, \(\rho_{1T}\) is the test reliability at Time 1, \(\sigma^2_{T2}\) is the variance at Time 2, \(\rho_{T2}\) is the test reliability at Time 2, \(\sigma^2_{1T2}\) is the covariance between the tests at Time 1 and 2.

The reliability of the difference score for STAR = 0.246 (based on \(\rho_{T1} & \rho_{T2}\) being 0.91 – split half reliability). The reliability of the difference score for asTTle = 0.90 (based on \(\rho_{T1} & \rho_{T2}\) being 0.98 – Cronbach alpha). The larger reliability of asTTle is based on the higher reported reliability of the test. If STAR had the same test reliability, the reliability of differences would be 0.81. Note that the tests use different methods to calculate reliability. Using a correction for unreliability (Gulliksen, 1987) which estimates what the maximum possible correlation could be between variables, given both were measured without error, the correlation between STAR and asTTle difference scores was still low (\(r_{corrected} = 0.279\)).
Also noted was the fact that for both tests, the correlation between the difference score and the Time 1 score for Years 4-6 was negative (-0.399 for asTTle and -0.509 for STAR). This suggests that the rate of difference is not the same across the distribution, that those scoring lower at Time 1, have greater difference scores. Both the unreliability of the difference scores and the negative correlation between the difference score and the Time 1 score warrant further investigation.

**The Issue of Differential Movement at Different Points in the Distribution**

Aside from the statistical phenomenon of regression to the mean, an obvious issue to examine is whether there are other potential contributors to differential progress across the distribution. Regression towards the mean, to some degree or another, occurs with repeated measurements on the same individual. It is independent of any intervention. So, some form of control is needed like using more than a single administration to obtain baseline scores and ensuring highly reliable instruments.

However, the effect of differential movement at different points on the distribution could be attenuated as a result of a skewed distribution. Skewness refers to the fact that a distribution is not symmetrical or normal. A ceiling effect occurs when an upper limit on a test results in a negatively skewed distribution whereby the majority of scores are at the higher end of the test scale at or near the maximum value (top heavy with a tail to the left or low end). A negatively skewed distribution is helpful for identifying persons who are at risk, to distinguish those individuals with a lower true score than that of the general population (Cronbach, 1990). This is something STAR was designed to do, according to the teachers’ manual. However ceiling effects are problematic in distinguishing those at the upper end and in accurately measuring improvement. If someone received the maximum value of a test at time 1 testing then they will of course show no improvement whether or not their ‘true ability’ has improved. Generally this shows that the test is too easy. A normal distribution is “preferred when it is necessary to distinguish all along the scale” (Cronbach, 1990, p. 212).

A test of skewness in our 2004 data by year showed significant skewness at Time 2 for Year 3 (-0.77); at Time 2 for Year 5 (-1.04) and at both Time 1 and 2 for Years 6 (-1.24 & -1.52), 7 (-0.57 & -0.81) and 8 (-0.81 & -0.97). (Z =skewness/sqr(6/N)) (Hair, Anderson, Tatham & Black, 1998). With each of the two STAR tests (the Years 4-6 test and the Years 7-8 test), the skewness increases. The skewness of the 2004 data is illustrated in the series of graphs that follow. They show the frequency distribution of raw scores with the best-fit curve superimposed.
Figure 3.1

Year 3: Distribution of Total Scores at Time 1 and Time 2

Skewness
T1: -0.04
T2: -0.772 (Skewness significant >99%)
Figure 3.2
Year 4: Distribution of Total Scores at Time 1 and Time 2

Skewness
T1: 0.233
T2: -0.213
Figure 3.3
Year 5: Distribution of Total Scores at Time 1 and Time 2

Skewness
T1: -0.336
T2: -1.036 (Skewness significant >99%)
Skewness
T1: -1.243 (Skewness significant >99%)
T1: -1.521 (Skewness significant >99%)
Figure 3.5
Year 7: Distribution of Total Scores at Time 1 and Time 2

Skewness
T1: -0.568 (Skewness significant >99%)
T1: -0.810 (Skewness significant >99%)
Skewness
T1: -0.808 (Skewness significant >99%)
T1: -0.970 (Skewness significant >99%)

There appear to be ceiling effects with the use of STAR, which should be taken account of in relation to the notion of differing relative movement or progress happening at different points in the distribution. To investigate ceiling effects, we used the data from both 2004 and 2005, but as 2004 had a larger sample size,
these are shown. The following five graphs (Figures 3.7 to 3.11) illustrate the scores for each year of Years 4-8 in 2004 at Time 1 (Series 1 February/March) and at Time 2 (Series 2 October/November). The Y axis is the total score. The maximum possible total score for Year 4, 5 and 6 is 50, while for Years 7 and 8 the maximum is 80. The X axis shows individual student scores at Time 1 and Time 2. The bars on the time 2 scores indicate the standard error of measurement for the instrument (plus or minus 3).

**Figure 3.7**

*Year 4 2004 (N = 104) STAR Total Scores at Time 1 and Time 2*

**Figure 3.8**

*Year 5 2004 (N = 115) STAR Total Scores at Time 1 and Time 2*
Figure 3.9
Year 6 2004 (N = 119) STAR Total Scores at Time 1 and Time 2

Figure 3.10
Year 7 2004 (N = 343) STAR Total Scores at Time 1 and Time 2
The 2004 graphs, in particular, make it clear that a ceiling effect is particularly marked for Year 6, the final year of the test intended for Years 4-6. However, there is a cluster of scores at the maximum in other years too, particularly in Years 7 and 8.

While STAR shows unreliability of the difference score and a ceiling effect, asTTle results exhibit neither of these. Plots of individual student scores on asTTle at Time 1 and Time 2 show no ceiling effect at any level. There was still room for growth as illustrated in Figures 3.12 to 3.16, which plot the scores for both time points for individuals by year level. Again, the Time 2 score has the addition of a bar to indicate the standard error of measurement of the asTTle test (SEM = 15). A consideration of these plots suggests that there is variability in individual performance. At Year 4, for example, 10 students out of 64 or 15.6%, score less at Time 2; at Year 5, 14.5% and Year 6, 22%. So, while the trend is increased performance, a proportion of students across the distribution show a negative gain.
Figure 3.12
Year 4 asTTle Total Scores at Time 1 and Time 2

Figure 3.13
Year 5 asTTle Total Scores at Time 1 and Time 2
Figure 3.14
Year 6 asTTle Total scores at Time 1 and Time 2

Figure 3.15
Year 7 asTTle Total Scores at Time 1 and Time 2
However, while there is room for growth to occur, there is still an issue in the sense that, in our data, there is a negative correlation between the asTTle reading score at Time 1 and the difference score between Time 1 and Time 2. This indicates that, generally, those with lower scores at Time 1 are improving at a greater rate than those with high scores at Time 1. This same negative correlation exists for STAR. For asTTle, however, the correlation is greatest at Year 4 and declines to insignificance at Year 6, while for STAR the magnitude of the correlation increases (see Table 3.4).

Table 3.4
Correlation of Time 1 Score and Difference Score

<table>
<thead>
<tr>
<th></th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>All Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>asTTle</td>
<td>-.693**</td>
<td>-.460**</td>
<td>-.141</td>
<td>-.399**</td>
</tr>
<tr>
<td>STAR</td>
<td>-.269*</td>
<td>-.520**</td>
<td>-.779**</td>
<td>-.509**</td>
</tr>
</tbody>
</table>

These results of an apparent differential gain at different parts of the distribution would suggest that those using these data to make decisions about progress and/or efficacy of interventions, consider examining the data after correcting for this effect. Adjusting for regression to the mean renders the overall correlation between the asTTle difference score and the asTTle Time 1 score for Years 4-6 near to zero (r = -.083). Even with adjustment, the correlation remains significant and negative overall for STAR (-.330), probably due to the considerable ceiling effect. The correlation between the adjusted difference scores for STAR and asTTle remains insignificant.
Those whose Time 1 scores were in the upper part of the distribution had little opportunity, given the construction of the test, to demonstrate growth at Time 2. This constriction would impact both variance and the true difference score. The Supplementary Tests of Achievement in Reading (STAR) states in its teacher manuals that it is designed to be able to “help schools to evaluate a new programme or policy, by testing children before and afterwards” (Elley, 2001 p8; Elley, 2003 p8). However, the use of STAR to measure improvement in reading ability is problematic due to a considerable ceiling effect apparent in the test. The lack of reliability in the difference score in STAR may be attributable to this very marked ceiling effect. While STAR may be useful to identify underachieving students, clearly schools would be well advised to reconsider its use to show progress across the distribution. The demonstration of a ceiling effect would help explain the relative lack of progress reported for Year 6 in the research relating to two other recent projects investigating gains in reading comprehension (Davis, 2006; Lai et al., 2004).

References


Part B: Teacher knowledge, beliefs and practices

Chapter 4

Examining the role of teacher pedagogical content knowledge in literacy

In the research literature there is increasing consensus that teaching accounts for a significant amount of the variance in student achievement (e.g. Alton-Lee, 2003; Rowan, Correnti and Millar, 2002). There is agreement that the most powerful way to raise student achievement is to foster quality or excellence in teaching (Darling-Hammond, 2001). So, there has been a concerted focus internationally on investigating and promoting effective teacher practice. Teacher knowledge is seen as key to effective practice. This knowledge involves knowledge of the subject, knowledge of how to teach and knowledge of learners.

Teacher content knowledge is central. Shulman (1986, 1987) called attention to this fact when he called it a “missing paradigm” (1986) in the study of teaching, arguing further that teachers require a particular sort of content knowledge; they need to know the subject in a way that helps them to teach it to others. Shulman’s (1986) notion was that such pedagogical content knowledge “embodies the aspects of content most germane to its teachability” (p. 9). This includes not only the major topics but representations of knowledge (including, particularly, the transformation of subject matter for teaching) and an understanding of what makes learning easy or difficult, of student learning difficulties and strategies to deal with them.

In its simplest form, indication of content knowledge should be represented in qualifications like those for teaching or post degree qualifications. In her review of teacher quality and student achievement, Darling-Hammond (2000) concludes that “measures of teacher preparation and certification are by far the strongest correlates of student achievement in reading and mathematics” (p.2). However, teacher certification has been questioned as a measure of content knowledge as standards are seen to vary (Ingersoll, 1999) while Friedman (2000) concludes that an advanced degree does not strongly influence the quality of teaching. There are inconsistent and thus inconclusive results from studies attempting to relate teachers’ subject knowledge either to the quality of teaching or to student outcomes or both (Ashton & Crocker, 1987; Ball, Lubienski & Mewborn, 2001; Ball, 2000; Wilson, Floden & Ferrini-Mundy, 2001).

While there may be several explanations for the disparities in findings, one of the reasons suggested is the different measures of teacher’s subject matter knowledge reflecting how researchers have conceptualized or defined this knowledge (Choi & Ahn, 2003). In their work within the TQ-QT Project, Choi and Ahn (2003) developed a list of indicators of teacher-subject matter knowledge from existing studies. It is apparent that most studies use educational background variables, which are problematic to equate (major, degree, amount or level of coursework/credits, GPA) to measure this knowledge. The next most common measure is test scores, generally researcher-designed tests or state licensure tests, the latter often multiple-choice and likely limited in their ability to test higher-level knowledge. A few studies employed observation of practice as a
measure, for example of “knowledge of the content and organization of a specific mathematical topic” (Stein, Baxter & Leinhardt, 1990, p. 642, cited in Choi & Ahn, 2003). Using four case studies, Choi and Ahn (2003) illustrate how “variation in defining and measuring teacher-subject matter knowledge is caused by the abstract nature of the teacher knowledge characterization” (p. 11) and conclude that these different ways of measuring yield inconsistent relationships between teacher knowledge and student outcomes.

In the present study, teacher’s knowledge of teaching reading comprehension and writing to elementary students was the focus. Arguably, teachers’ pedagogical content knowledge of reading (and writing) is somewhat different to subject knowledge of mathematics or history or even English. Such knowledge has not been a major topic of enquiry, although interest in the professional knowledge associated with teaching reading is increasing (Phelps & Schilling, 2004). There has been some research concerning the knowledge of language needed to help children learn to read or, more particularly, to decode (e.g. McCutcheon et al., 2002; McCutchen & Berninger, 1999; Moats, 2000; Moats, & Lyon, 1996; Wong-Fillimore & Snow, 2002). The findings from this body of work largely concern teachers’ lack of knowledge of specific features of language, notably the features that constitute words. This knowledge is important to make instruction explicit for children. As Phelps and Shilling (2004: 7) illustrate neatly, the capacity to make good teaching decisions or moves, in teaching reading, rests “partly on the teachers’ knowledge of the subtleties of word and text structure”. Specific linguistic knowledge may be needed to teach decoding but awareness of process is also needed in order, for example, to give explicit instruction in the strategies for comprehending text.

With respect to content knowledge needed to teach writing, there is a dearth of research. It is clear that teachers need a different type of knowledge than that possessed by a competent adult writer. Clearly, they need to know about how texts work to achieve their communicative, rhetorical purposes, including knowledge of the features of text most commonly employed to support writing for a particular purpose. This involves a detailed knowledge of language and of text structures. Teachers also need to be able to make accessible to developing writers that which is implicit and often at a level below conscious thought; to unpack what they are doing as they engage in the writing process. For example, in order to teach developing writers, teachers need to know what strategies more expert writers use as they engage in the complex activity of writing.

As with any form of assessment, though, the purpose for finding out about teacher knowledge has an influence on the ways and means by which the investigation is conducted. Like assessment generally, research into teacher knowledge still appears to be driven by evaluative purposes; the major reason seems to be for evaluation of quality of output (from pre-service programs) or for establishing competency. Investigating teacher knowledge, however, can serve the important purpose of identifying what practising teachers need to know in order to better meet their student’s learning needs.

The reason for investigating teacher pedagogical content knowledge in this study was for teacher learning. Arguably, using evidence of knowledge or learning is part of developing an inquiry habit (Earl & Katz, 2002); data lead to more focussed investigations and better questions. The method of enquiry into teacher knowledge in this study was designed to provide teachers with the opportunity to reflect on what they knew and to identify gaps in that knowledge.
**Context**

The present study is part of research accompanying and informing a national professional development project in literacy. The project specifically focuses, depending on the school’s choice, on either reading comprehension or writing. The project aims, through the use of expert facilitators working with schools, to build knowledge and improvement at all levels and to be evidence-informed in doing this. The underpinning philosophy of the project is that school-based, job-embedded professional development be needs-based, that is it consider, in concert, both the needs of students in terms of progressing their literacy learning and what teachers need to know in order to address those needs. A measure of teacher pedagogical content knowledge was thus one of the types of evidence used to investigate teacher need. Another feature of the project was that capacity be developed within schools for professional learning and for the ongoing learning to be self-sustaining.

The research questions for this study concern both how to measure pedagogical content knowledge in a way that promotes teacher reflection and growth and the relationship between teacher pedagogical content knowledge and student achievement and progress in either reading comprehension or writing.

**Method**

**Participants**

For participating schools, this is a two year project and the data presented are from the first cohort, who began in February, 2004. Teachers in the research were from seven schools, largely urban, participating in the research. The teachers taught Years 1-8. In Years 1-6, students have a main classroom teacher although there may be some specialist teaching for remedial work (e.g. Reading Recovery) or for music, drama or physical education, depending on the composition of the school staff and individual school decisions about use of staff. At Years 7 and 8 there is a history of specialist teaching of technical subjects, technology and arts and crafts. So, the participants included some specialist teachers as school principals, generally, were of the view that literacy was everyone’s responsibility. As this was a school-wide professional development project, principals and senior management as well as classroom teachers participated.

In reading schools, there were 73 teachers who provided data at Time 1. Their teaching experience ranged from 0.1 to 40 years with a mean of 13 years. Around 90 percent percentage of teachers reported that they had been in their current school for two years or more. Data at two points in time for reading schools were available for 59 participants and their students. In writing schools, there were 43 teachers who provided data at Time 1. Their teaching experience ranged from 0.2 to 36 years with a mean of 14.5 years. Sixty-five percent of these teachers reported being in their current school for two years or more. Data at two points in time for writing schools were available for 35 participants and their students.

**Measures and Procedure**

Data with regard to student achievement and teacher knowledge were collected at two points in time, near the beginning as part of a needs analysis process and towards the end of the school year. With respect to the project’s aim to create self-sustaining learning in literacy within school communities, schools were seen to advance through three phases. Most schools were just entering phase 2 of the project when the second round of data was collected towards the end of a year.
Student achievement data
In reading, data were from a standardized test of reading, Supplementary Test of Reading (STAR) (New Zealand Council for Educational Research & Elley, 2001). There are three test versions for different year groupings each with an A and B form. All have common sub-tests, namely word recognition, sentence comprehension, paragraph comprehension and vocabulary. The test for Years 7 and 8 has two additional subtests while the common tests for Years 4-6 contain more items than the Year 3 test. In writing, data were obtained from a criterion referenced (to NZ curriculum) measure of writing, Assessment Tools for Teaching and Learning: Writing (asTTle) (Glasswell, Parr & Aikman, 2001) that has associated national normative data (Yrs 4-8). There are scoring rubrics for each of six writing purposes with criteria for different levels of achievement for seven dimensions of writing (audience, structure, content, language resources, grammar, spelling and punctuation).

Teacher knowledge
The primary source of data for teacher knowledge was a questionnaire that all teaching staff and senior management who were present completed at a staff meeting. Similar, parallel forms of this questionnaire were used as responses were gathered at the two points in time.

The major instrument for pedagogical content knowledge (PCK) in writing or reading comprehension was a description of a hypothetical classroom lesson involving either the teaching of comprehension or of writing (see Appendix A for an example). The reading comprehension scenario described what the teacher aimed to do in the lesson, then what she did in terms of introducing the text and guiding the student’s reading of it. Here, participating teachers were asked to rate (on a six point Likert scale) aspects of the scenario teacher’s practice, namely, the lesson aim, the activities the teacher organised and the feedback given to students and to justify by giving reasons for the rating selected. Respondents were also asked to identify effective and less effective teaching moves and give reasons for their choice. The writing scenario, while describing the context and the teacher’s aims, also contained a student’s writing sample. Participating teachers were asked to rate aspects of the scenario teacher’s practice (similar to the reading scenario) and give reasons, then to provide a general comment and specific feedback to the student whose writing sample was included.

It was reasoned that teachers would reflect on knowledge (and practice) both through responding to the questionnaire scenario and then, subsequently, through considering their responses in relation to those of their peers in discussion forums. In addition, they could relate the teacher actions and their responses and ideas to information obtained in professional readings, specifically with regard to effective literacy practice.

Data Analysis
Student Achievement
Reading achievement tests were scored by classroom teachers using the scoring sheets that accompany the tests. As judgements are not really called for in scoring STAR, an accuracy check of addition of sub tests; of correctly locating the stanine corresponding to the score and time of year and of actual recording of data was made of 20 percent of every teacher’s class. If any errors were found (there was at least one error in three classrooms), scripts from the whole class were checked. Given that three different STAR tests were used (one for Year 3; one for 4-6 and one for 7-8), each with different maximum total scores, a ‘mean item score’
was calculated by dividing each student’s score by the amount the particular test was out of. An effect size for a class was calculated as class mean item score at T2 minus that at T1, divided by average class SD (T1 SD+ T2 SD / 2) and this was used as the measure of gain.

For writing, the research team scored all scripts at Time 1 and scripts from all schools, bar one, at Time 2. In this latter school, a sample of 230 scripts, approximately a third of each class, was scored. The resulting level of reliability with teacher scoring was at an unsatisfactory level so only researcher data were included in analyses for consistency. As asTTle writing is scored on a common scale, gain was calculated from total scores at Time 1 and Time 2 (although there are slightly different average gains for different years shown in the normative sample).

**Teacher Knowledge**

With respect to teacher pedagogical content knowledge, for scenario items where ratings were asked for, descriptive statistics were calculated. The majority of responses to the scenarios, including reasons for ratings, however, were open-ended responses and these qualitative data were coded and frequency counts made. For the reading scenarios, effective and less effective moves were identified by expert practitioner consensus. The reason given for selecting the move was used to check that the respondent had understood why that move was effective or less effective. Identifying a move appropriately with an adequate reason that explained why it was such, scored two points. Likewise, where reasons for ratings (for example of the aim for the lesson; the activities for students and the feedback given to students) were given, those responses that scored two points identified the crucial problematic aspect of the practice. One point was scored for responses that were partially correct but did not identify the crux of the issue. For example, a respondent may have identified that the aim was not shared with the group but failed to note that the aim was not related to the substance of the lesson. Or, a respondent may have identified an effective or ineffective move but the reason given for identifying it did not confirm that the teacher clearly knew why it was effective (or ineffective).

For the writing scenario, in terms of both the overall comment and the specific feedback comments to the student whose writing sample was included, the five dimensions of interest were: whether the feedback was evaluative; whether it was accurate in terms of writing function and lesson aim; whether it was specific to writing function or lesson aim; whether there was indication of action for writer to take, and the level of text to which the feedback was directed.

Then, the evaluative nature of the comment, for example, was further coded in terms of: confusing or mixed messages or contained inaccuracies; personal response or non- specific positive; evaluation or direction implicit; evaluation explicit; positive plus direction, and evaluation plus direction. Comments that were explicitly evaluative and/or indicated a direction to proceed scored two points, while those that were a personal response or contained only implicit evaluation or direction scored only one.

In addition to evaluative quality, feedback comments were coded in terms of whether they were specific to the writing function (e.g. to recount) or the lesson aim, or simply generic; whether they were accurate in terms of the function and aim, and whether the comments contained an indication of action that the writer should take to improve the piece. Appropriate responses scored one. Finally, the level of text at which
comments were directed was coded from deeper features of text (rhetorical/audience, structure, content and language resources) through to surface features (grammar, spelling and punctuation). A reference to deep features in the feedback scored one point. A total score was calculated for the writing scenario.

Reliability of coding of responses was obtained by a second rater coding all questions for a 20 percent sample of questionnaires. The percentage agreement between raters was then calculated. For most questions, acceptable levels were obtained, namely, greater than 75 percent agreement. However, for the questions where reliability was less than this, the second rater coded responses from all of the teachers. These were compared with the first rater’s coding and where disagreements occurred these were resolved through discussion. On approximately two percent of responses, agreement could not be reached. The first author reviewed these and made the final decision.

Findings
Teacher pedagogical content knowledge was examined through the scenario responses at each point in time and over time. Responses to the scenarios are presented first for reading and then for writing. To examine the relationship of teacher knowledge to student progress, teacher scores from the reading scenario scores are related to progress for students in their class on STAR in terms of effect size. Writing scenario scores for teachers are considered in relation to the gain scores on asTTle for their class of students.

The Reading Classroom Scenario
The responses involved both the identification of two effective and two ineffective moves and ratings of teacher moves as described in the scenario.

Identifying and rating moves: Time 1 and Time 2
At Time 1, while all effective moves were identified at some point, the largest percentage of responses recognised the effective moves of establishing and using prior knowledge and of providing information to fill gaps in knowledge (total of 37.6%) and of setting a purpose for reading (26%). At Time 1, a small percentage of responses were classified as ‘no response’ (4.5%) or as wrongly identifying a less effective move as effective (6.5%), or were vague, non-specific responses (5%). Table 4.1 shows these categories of response for effective moves and allows a comparison with less effective moves.

<table>
<thead>
<tr>
<th>Category of Response</th>
<th>Effective</th>
<th>Less Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly recognised</td>
<td>84.3</td>
<td>62.9</td>
</tr>
<tr>
<td>Wrongly identified</td>
<td>6.5</td>
<td>5.2</td>
</tr>
<tr>
<td>Insufficiently specified/other</td>
<td>4.5</td>
<td>7.1</td>
</tr>
<tr>
<td>New Move</td>
<td>0</td>
<td>11.0</td>
</tr>
<tr>
<td>No Response</td>
<td>4.5</td>
<td>13.6</td>
</tr>
</tbody>
</table>
Respondents were not as successful in identifying the less effective moves. This is illustrated by the fact that there was a greater number of ‘no response’ (14%) and the fact that 11 percent specified an additional new move that was not present (although the instructions clearly asked respondents to identify a less effective move that the teacher made). There were similar proportions of responses (as when identifying effective moves) that wrongly identified effective moves as less effective ones (5%) and those that were vague and generalised (5%). At Time 1, the ineffective moves with the largest percentage of responses were the fact the teacher gave non-specific feedback in summary of the lesson (18%); that the teacher intended to wait too long before returning to an issue (16%) and the teacher let students read too big a chunk of text with no focus or guidance (12%).

There were some similar patterns in responses at Time 2 to a seemingly comparable scenario with respect to identifying effective moves. Table 4.2 summarises these. The ‘no response’ rate remained much the same but non-specific responses fell to two percent. However, the percentage of responses that incorrectly identified a less effective move increased to 23 percent with teachers commonly identifying the retelling (which builds comprehension but, in this case, did not relate to the lesson learning aim) and the context within which this was done, working with partners, as effective moves. (Note: The activity the teacher arranged with partners was, in addition, a move they were asked to rate). Again, the largest percentage of responses appropriately identified using and relating to prior knowledge or ensuring students had the necessary background information to assist comprehension (40%). Another sizeable category of responses concerned helping student to use cues to predict (20%).

**Table 4.2**

<table>
<thead>
<tr>
<th>Category of Response</th>
<th>Effective</th>
<th>Less Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly recognised</td>
<td>68.6</td>
<td>71</td>
</tr>
<tr>
<td>Wrongly identified</td>
<td>23.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Insufficiently specified/other</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>New Move</td>
<td>0</td>
<td>7.4</td>
</tr>
<tr>
<td>No Response</td>
<td>6.2</td>
<td>17.3</td>
</tr>
</tbody>
</table>

With respect to identifying the less effective moves at Time 2, again, as at Time 1, there was a larger category of no response (17%). Smaller percentages of responses than at Time 1 wrongly identified an effective move (3%) or were insufficiently specific (1%) or specified an additional move that was not present (7%). The largest percentage of appropriate responses concerned identifying where the teacher had students read a largish chunk of text with little guidance (24%), while giving feedback not related to the learning aim was identified in 14 percent of responses. A similar percentage noted that the drawing activity did not relate to the reading so did not function to assist or check comprehension.
In addition, at Time 2, before identifying the moves, respondents were asked to rate, and give reasons for rating, teacher moves in relation to three aspects of the lesson; moves considered by experts as likely to be ineffective. These included the aim for the lesson, the activities provided and the feedback given to students. Table 4.3 shows the mean ratings for these teacher moves (scale is 6 point where 3= slightly ineffective and 4 = slightly effective and 5 effective).

Table 4.3
Mean Ratings for Teacher Moves

<table>
<thead>
<tr>
<th></th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>4.1 (1.4)</td>
</tr>
<tr>
<td>Activities</td>
<td>4.5 (1.1)</td>
</tr>
<tr>
<td>Feedback</td>
<td>2.9 (1.4)</td>
</tr>
</tbody>
</table>

The aim, while seeming clear and appropriate, bore no relation to the ensuing description of what went on in the lesson. The mean rating was 4.1 (4= slightly effective on a six point scale) with a range of 1 to 6 (SD 1.4). About 20 percent of respondents thought it a clear, specific aim, although 12 percent thought it not specific enough and 10 percent, while willing to rate it, offered no response. About 15 percent of reasons given for rating related not to the aim but to another segment of the lesson. Some found fault with the aim in that there was no associated success criteria to indicate how the aim was to be achieved (14%) while a few noted that it was not shared with the group (5%). Only 11 percent diagnosed that the aim and the substance of the lesson were not related.

The activity, which involved working with a partner, was described by many (47%) as a useful, enjoyable, inclusive oral activity to clarify understanding. Clearly this is in line with the number who identified this as an effective move. It was rated at 4.5 (between slightly effective and effective; SD 1.1) and, again, the full six point scale was used by respondents. However, a number of respondents perceived flaws. Around 10 percent accurately noted that the paired activity, a retelling, while a useful comprehension strategy, was not related to the learning aim of making a visual summary. Then some responses (9%) indicated that a few teachers thought the activity may have been reasonable in terms of relating to the aim if there had been guidance on how this should be done. Another seven percent also implied they thought it was aligned with the lesson aim but the outcome may not be as the teacher did not check by having students report back. Small percentages of respondents noted other issues with the activity such as a lack of modelling by the teacher on how to do this retelling (4%).

The final rating concerned feedback and it referred to the teacher giving a generalised positive comment at the end of the lesson. This was rated on average at 2.9 (3 = slightly ineffective; range 1 to 6, SD 1.4). Around 26 percent of respondents gave as a reason for their rating that the feedback was non-specific, general praise and a further 17 percent that it was limited because it was based on insufficient evidence, was given too soon or was not individualised. A third of the reasons offered showed that respondents recognised that the feedback was poor in that it did not relate to the lesson aims or desired outcomes, while nine percent thought it lacking as there was no element of feed forward.
The total scores for teacher knowledge at both points in time were compiled from the responses to identifying the effective and less effective teaching moves, as these were common to both points in time. The mean at Time 1 was 5.43 (SD 2.3) and, at Time 2, 5.60 (SD 1.9). There was not a significant increase in total score over time.

Relationship between teacher knowledge and student progress
Student progress in reading for individual classrooms was considered by calculating an effect size. Effect sizes ranged from 0 to 1.3 with a mean of .46 (SD .26). (Note: The mean effect size for gain in reading in the total national sample was 0.87). There was, in reading, no significant correlation between a teacher’s pedagogical content knowledge level at Time 2 and their student achievement gain ($r = .21, p = .17$).

The Writing Classroom Scenario
The responses to the scenario involved both ratings of teacher moves as described and reasons for ratings, together with feedback on the student writing sample.

Rating moves and providing reasons
With respect to rating of moves, at Time 1, teachers were asked to rate only the quality of the aim of the teacher in the scenario while, at Time 2, they were asked to rate the aim and, in addition, the way the teacher introduced the lesson and the paired activity the teacher used. In all cases, the whole rating scale (1 to 6) was utilised. The mean ratings are shown in Table 4.4 and were above slightly ineffective (3) and slightly effective (4).

<table>
<thead>
<tr>
<th>Aim</th>
<th>3.6 (1.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4.2 (1.4)</td>
</tr>
<tr>
<td>Activities</td>
<td>4.3 (1.4)</td>
</tr>
</tbody>
</table>

The reasons teachers gave for their chosen rating showed that, at both Time 1 and Time 2, nearly half of the teachers could see the non-specificity of the aim for the lesson. With respect to the question at Time 2 regarding the way the teacher introduced the writing task, under half recognized the inappropriateness of the teacher actions in that the modelling was neither related to the aim of the lesson or to the writing purpose, recount. As to the activity done in pairs, less than a third recognised that it did not relate either to the lesson aim or that it may not, necessarily, without guidance, relate to the writing purpose. Nearly 60 percent of responses considered it a useful oral activity to gain more information or to get feedback.

Feedback to student: Comparison of Time 1 and Time 2
The giving of written feedback to a student is a potentially powerful lens through which to view growth in teacher content knowledge in writing, as it involves first diagnosing a problem in text and then seeing a solution that the student could conceivably enact, both of which mean retrieving, then applying, content knowledge about writing. The questions that could be asked of these data include what were the features that teachers focused on and were there changes in the nature and quality of their feedback to the student over time? The expectation is that feedback at Time 2 would contain more evaluation and direction, that it be
more specific to the writing purpose and aim, and that more comments would contain an indication to the student of action to take. Finally, comments may, over time, should show a trend away from the commonly noted focus on mechanics.

The quality of teacher feedback on writing showed modest increases from Time 1 to Time 2 (See Table 4.5). For example, in terms of the evaluative quality dimension, there was a decline in the largest subcategory, that of implicit comments - where it was simply implied that the writing was lacking - from 68 percent at Time 1 to 52 percent at Time 2. Correspondingly, comments containing an explicit evaluation and/or direction and comments specifically linked to the writing purpose or lesson aim, increased at Time 2.

**Table 4.5**
Percentage of Responses in Evaluative Quality Categories of Teacher Feedback at Time 1 & Time 2

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit comment</td>
<td>67.6</td>
<td>52.2</td>
</tr>
<tr>
<td>Explicit direction</td>
<td>5.9</td>
<td>11.9</td>
</tr>
<tr>
<td>Linked to aim/purpose</td>
<td>17.7</td>
<td>22.4</td>
</tr>
</tbody>
</table>

Most noticeably, at Time 2, teachers were giving feedback at higher levels of text. Figure 4.1 shows this shift graphically where 1 represents rhetorical concerns like taking account of audience; 2 includes structural features; 3 content related feedback; 4 language resources to achieve purpose; 5 syntax related and 6 mechanics like spelling and punctuation. The proportion of comments in these categories changed. The change in proportions was statistically significant, that is the number who went from lower levels of text to higher was significantly different from the number who went from higher to lower ($p = .004$, McNemar test).

**Figure 4.1**
Percentage of Feedback at Different Levels of Text at Time 1 and Time 2
There was no significant increase over time in the total score for teacher knowledge of writing \((t = .087, p > .05)\). With a small sample size, difference between schools is a factor to consider, as there may be considerable growth in one school and not in another.

Relationship between teacher knowledge and student progress

The pedagogical content knowledge scores of individual teachers (Years 4-8) for the writing scenario were considered in relation to the writing progress of students in their class (gain in total scores) (Note: In the national sample, the mean effect size for gain in writing was 1.28, considerably larger than in reading). The level of knowledge of teacher at Time 2 was considered in relation to the progress of students by class. Teacher knowledge was significantly related to student progress \((r = 0.77, p =.004)\). Those teachers with higher levels of pedagogical content knowledge had students who made greater progress.

**Discussion**

This was a novel approach that aimed to conceptualise what form pedagogical content knowledge might take in reading comprehension and writing, then to measure this using responses to hypothetical scenarios. Others (e.g. Phelps & Shilling, 2004) have designed multi-choice test items to examine linguistic knowledge related to teaching reading. Their aim was largely to evaluate teacher knowledge, for example, the relative standing of graduates from different programmes. The decision to use the scenario format with open-ended responses was related to the use to which the data were to be put, namely, as a formative assessment tool for teacher learning and one that could be employed to assist teacher reflection. Previous research work (Timperley, Parr & Higginson, 2003; Timperley & Parr, 2005) had pointed to the need for practitioners to perceive a need for the professional learning offered and to see it as contributing to their professional knowledge. The analyses of responses to the scenarios give a picture of what teachers readily recognise as effective and less effective moves and where difficulties arise in their efforts to identify such moves. The data from such a tool have been successfully used by literacy facilitators as a starting point for professional learning and discussion within the school community.

This study wanted to view teacher pedagogical content knowledge in reading and writing in terms of its contribution to variance in student achievement. The reading test while designed locally and containing established components of reading ability was, unlike the writing test, not curriculum referenced. Further as a previous chapter illustrates, the opportunity for growth in reading for some students in particular year levels was severely curtailed. A truncated range affects the likelihood of obtaining a significant correlation. The results were mixed and lend support to the argument of Choi and Ahn (2003) that differences may well be a function of the way in which knowledge is measured. In reading, there was no relationship between level of teacher knowledge of reading and gains in student achievement. While the lack of a significant relationship may be a function of small sample size or of the likely measurement error associated with measuring a construct like teacher content knowledge (Choi & Ahn, 2003), the lack of a relationship in reading could be partly explained by the nature of the task. In the reading scenario, unlike in the writing scenario, teachers had support in that they only had to recognise the required pedagogical content knowledge rather than retrieve it from memory and then apply it as was required in giving feedback to the student in the writing scenario. Perhaps this contributed to the second explanation for the lack of a relationship, namely, the fact that there was not a great deal of variation in participant’s ability to identify moves. Clearly, more work
needs to be done in terms of constructing a scenario that requires teachers to diagnose a situation, retrieve the knowledge they have of, say, reading comprehension strategies, then apply this knowledge to instruction. Such a scenario may more faithfully represent pedagogical content knowledge in teaching reading comprehension and may provide more variability in responses.

How knowledgeable a teacher was with respect to what was important in teaching writing, however, related significantly to the extent of progress that students made in terms of writing achievement. This finding is encouraging. The notion was that considerable pedagogical content knowledge was required in order to give quality feedback on writing. The hypothesis that pedagogical content knowledge in writing comprises knowledge of how texts work in relation to purpose and the structures and language features that are associated with particular communicative purposes and contexts, a notion which was explored in this study through examining the quality of feedback, needs further exploration.

References


Chapter 5
What Are We Supposed To Be Learning?
Students’ Understanding of Their Writing Lessons

There is considerable evidence that students who understand learning goals and either receive or self-generate feedback about their progress in reaching them are more successful learners than those who do not. The evidence comes from the convergence of two different perspectives, one with its origins and emphases on the development of self-regulatory learning processes both within and among individuals (e.g. Butler & Winne, 1995; Zimmerman, 2001) and the other with greater emphasis on the development of those learning processes through effective classroom teaching associated with formative assessment (e.g. Black & Wiliam, 1998). Considered in combination, these perspectives have the potential to provide teachers the necessary understanding of the links between student cognition, in particular the development of self-regulation, and daily teaching actions. From these perspectives we identify the key features of the teaching environment that provide the conditions likely to promote student self-regulated learning. In this chapter, we report on the extent to which these conditions were present in the classrooms of teachers involved in the literacy professional development contract at three different time periods and the impact on student understanding.

Claims about the effectiveness of self-regulated learning (SRL) are captured in Butler and Winne’s (1995) extensive review of the literature, which they began with the statement, “Theoreticians seem unanimous - the most effective learners are self-regulating” (Butler & Winne, 1995, p. 245). It is difficult to present a single definition of the concept of self-regulated learning because at least seven different theoretical perspectives have emerged (Zimmerman, 2001). However, Butler and Winne’s (1995) portrayal captures many of the key elements when they describe self-regulated learners as those who, “… judge performance relative to goals, generate internal feedback about amounts and rates of progress towards goals, and adjust further action based on that feedback” (p. 258). It is, in their view, a deliberate, judgmental, adaptive process.

Much of the research on self-regulation, particularly that coming from a cognitive constructivist perspective, has examined how these self-regulatory processes develop independently of any particular qualities in the learning environment (e.g. Son & Metcalfe, 2000; Theide & Dunlosky, 1999). The ability to self-regulate requires sufficient cognitive maturity for individuals to set goals, to be aware of and to monitor their own learning processes. Some psychologists in this tradition explain self-regulation in terms of a discrepancy-reduction model which involves monitoring the discrepancy between current knowledge and a desired state, with learners continuing to engage in the learning process until the discrepancy reaches zero (Dunlosky & Hertzog, 1997; Son & Metcalfe, 2000; Theide & Dunlosky, 1999). From this perspective, knowledge of the learning goals and the ability to monitor progress towards them accurately is central.

Zimmerman (2000), in adopting a more socially-oriented cognitive perspective, has been critical of the strong focus of the above research on what he calls singular traits related to cognition alone and proposes that to understand self-regulation is to understand how learners constantly adjust to the changing personal, behavioural, and environmental conditions in typical learning situations. Rather than self-regulated learners reducing performance discrepancies reactively against a set standard, he proposes that they also act
proactively and increase “performance discrepancies by raising goals and seeking more challenging tasks” (p. 14). This more dynamic view of SRL is probably more reflective of constantly evolving classroom situations with their myriad of learning opportunities and potential goals.

This social cognitive perspective has more to offer in identifying the kinds of teaching and learning environments that might lead to developing self-regulation. Rather than relying on maturational processes, these theorists identify social processes in dynamic learning environments (e.g. Butler & Cartier, 2004). Zimmerman (2000) proposes that self-regulated learning emerges from a process of observation and emulation of a model, which then develops into self-control and self-regulation. From this perspective, teachers, presumably, could foster self-regulation in their students by providing the model and the conditions to promote self-regulatory strategies.

A more deliberate teaching–learning relationship is evident in the Vygotskian perspective, which understands the learning process in essentially socio-cultural terms. Through guided interactions with more skilled others, important concepts and skills in the external environment become internalised, and as learners are scaffolded into gaining greater control of the ideas, they are increasingly able to guide, plan and monitor their own activities. Thus, “for every individual at any point in time, there will be a mix of other regulation, self-regulation and other automatized processes.” (Gallimore & Tharp, 1990, p. 186). The points at which a teacher is likely to be most effective in developing self-regulation are those that focus on aspects of skills that are emerging but still require the assistance of others for their mastery.

These more socially oriented perspectives on self-regulated learning are evident in the literature on formative assessment. However, the research on formative assessment has a more direct focus on the role of classroom teaching activities and their impact on students. The parallels between the two research perspectives are obvious however, when looking at the learning processes involved, as identified in the seminal review by Black and Wiliam (1998). These authors suggest that the core of the activity of formative assessment lies in the sequence of two actions.

The first is the perception by the learner of a gap between a desired goal and his or her present state (of knowledge, and/or understanding, and/or skills). The second is the action taken by the learner to close that gap in order to attain the desired goal. (p. 9).

These authors calculated that by the end of secondary school, the achievement of students whose teachers helped them to understand what it was they were supposed to be learning, fostered self-assessment and gave appropriate feedback, showed an additional one to two years progress in comparison to their peers.

It appears from this literature that fostering effective self-regulation and realizing the benefits of formative assessment requires teachers to help their learners understand the learning goals and to provide the opportunities for them to seek and receive feedback on progress towards those goals. In this study, we examined the extent to which teachers provided these conditions for their students, and further, whether the students were able to utilize them in ways that fostered SRL.
Goals and Feedback

In this section, we consider the literature on the qualities of goals and feedback likely to foster SRL. In its purest form, SRL involves learners formulating their own goals. However, we are taking the position that young students need guidance in this, and it is the role of teachers to make these goals clear to students so that they know the purpose of classroom learning tasks and activities. Understanding what is to be learned is fundamental to successful task interpretation (Butler & Cartier, 2004). Specific rather than general goals have been found to be more effective in focusing students’ attention, developing greater commitment and allowing more directed feedback (Bargh, Gollwitzer, Lee-Chai, Barndollar & Trotschel, 2001; Kluger & DeNisi, 1996; Locke & Latham, 1984).

Learning goals need to be distinguished from performance goals because they are differentially effective in promoting self-regulated learning. Learning goals focus on understanding how to tackle new problems and learn new things, such as learning to formulate an argument, whilst performance goals focus on grades and tend to have a strong focus on performance being associated with ability, such as doing better than one’s peers. Newmann and Schwager (1995) found that a focus on performance goals developed maladaptive questioning patterns and poorer problem-solving ability than those focused on learning goals.

Closely aligned to learning goals is the power of mastery learning, which involves the learner having an understanding of what success in that task might look like and receiving instruction and feedback directly related to it. Research reviews assessing the effect of mastery learning have found an average effect size of 0.82 (Guskey & Gates, 1986), which is among the largest average effects reported for a teaching strategy (Kulik & Kulik, 1989).

Although feedback features prominently in all the above accounts, it has highly variable effects. The right type of feedback, therefore, is important. Feedback about the personal qualities of the learner invites a focus on social relationships rather than cognitive processes and can be detrimental to the development of self-regulated learning (Brophy, 1981; Kluger and DeNisi, 1998).

Outcome feedback, sometimes called knowledge of results, which gives binary information about whether or not a particular response is correct, can also be problematic in developing SRL. Typically, this type of feedback does not carry sufficient information to guide a learner about how to self-regulate. More cognitively oriented process feedback helps students identify cues indicative of progress towards particular goals, monitor task engagement and assess the value of those cues in achieving task success (Butler & Winne, 1995).

Feedback can also promote either minimal or deep learning. Feedback that promotes minimal learning is typically focused on the correctness of content in a domain and usually contains insufficient information to affect the development of knowledge construction. Feedback directed at deep learning, on the other hand, triggers other forms of cognitive processing, such as assembling ideas, searching for relationships and developing knowledge with which to elaborate information (Balzer, Doherty, & O’Connor, 1989). Thus, the most efficacious type of feedback is that which promotes self-monitoring, directing, and regulating activities associated with the learning goal at these deeper levels. It fosters autonomy, self-control, self-direction and self-discipline (Zimmerman, 2000).
Receiving information about progress towards a goal, however, is of little benefit unless a third condition is also met, that is, the feedback is linked to a corrective strategy. This type of feedback is sometimes referred to as feedforward. A key condition of SRL is that learners adjust future actions in response to feedback, so the quality of the information contained in the feedback or in the feedforward is fundamental to developing effective processes (Black & Wiliam, 1998). Zimmerman (2000) found that the adjustment was not necessarily deliberative, but may also occur incidentally, but can still be an effective learning strategy.

Goals, feedback and feedforward are closely interrelated. Locke and Latham (1990) identified that goals and feedback work both prospectively and retrospectively. Goals can inform learners about the level of performance to be attained so that they can direct and evaluate their actions and efforts accordingly, while feedback allows the learners to set reasonable goals and to track their performance in relation to their goals.

How goals are defined also plays a role in determining the feedback cues to which learners attend. Learners judge progress towards goals by selecting and monitoring cues they perceive to be relevant to achieving a particular goal. If the writing purpose is to give an account of a ball game and the teacher has a specific learning goal, for example learning about temporal markers, the students will use the knowledge they have about the order of events to select cues to monitor and generate internal feedback. They will then use this feedback to work out whether their writing efforts are meeting the learning goal. If the teachers’ learning goals are unclear or inaccessible to the students, they are likely to construct alternative goals, for example, writing a whole page, and use these to assess their progress (Butler & Winne, 1995). Alternatively, if teachers state one kind of learning goal but give feedback that is misaligned to those goals, the learners are likely to be confused about what it is they are to self-monitor and how best to make progress.

In this empirical study of the relationship between teaching strategies associated with fostering SRL and student’s understanding, we analyzed a selection of writing lessons in elementary classrooms (Years 4 – 8). We did not attempt to categorize the “whole” lesson to determine the overall style of teaching or whether particular processing strategies were explicitly taught as have other authors studying SRL (e.g. Bolhuis & Voeten, 2001) but rather focused on the key conditions discussed above. During each lesson we interviewed a sample of students to find out what sense they made of the learning goals, the feedback and feedforward information provided by their teachers in order to understand the relationship between the features evident during the lesson and the students’ understanding.

**Method**

This analysis includes 20 writing lessons for 15 different teachers in Year 2 – 8 classrooms. The lessons were recorded using audio equipment and later transcribed. They are divided into three sets. The first set of 9 observations took place prior to any professional development with respect to the LPDP contract. They were all from Otago / Southland schools. The second group of eight observations were undertaken between four and ten months after the professional development project began. Six of these teachers were from the Northland / Auckland area who had been involved in the professional development for between six and ten months but their earlier practice had not been recorded. Two of the teachers were from the Otago / Southland area whose practice had been recorded in the first set, so they are included in both observation sets. The third set of observations was for three teachers from the Auckland / Northland region after two years of
involvement in the LPDP contract. These teachers are the only ones remaining in the writing research schools in the Auckland / Northland area. Their lessons were also included in the second set of observations.

In addition to the observations, teachers’ were asked to rate their confidence in undertaking particular aspects of lesson activities related to what happened in those lessons. These items were in a questionnaire format and were completed at approximately the same time as the observations took place, that is, at the beginning of the contract, towards the end of the first and second years. The two teachers re-observed only four months after the beginning of the project were not asked to fill in a questionnaire at that time, so no confidence ratings are reported. Demographic information, such as years of experience was also requested on these questionnaires. The observation protocols and questionnaire items are detailed below.

**The Lesson Observation**

Prior to the lesson observation, the teachers were asked to fill in a sheet that provided the observer with the lesson aims, where the lesson fitted in relation to other lessons and the intended lesson organisation. With their permission, the teachers’ instructions and interactions with students were audiotaped for each writing lesson that lasted for approximately 45 minutes. Nearly all lessons followed the same format – a whole class introduction, followed by independent writing or group work.

Some students’ responses were audible during teacher–student interactions depending on their distance from the teacher. During the second part of the lesson when the students were working independently, between three and nine students per class, depending on the time available, were interviewed. The part of the interview schedule used in this analysis consisted of the following five questions. The purpose of each question is noted in brackets after the question.

- What are you supposed to be doing? (Purpose – general introduction to get students talking)
- What are you learning about writing while you are doing this? (Purpose – to find out if they were aware of the writing aims for the lesson)
- Can you tell me what a good [type of writing working on] looks like? E.g. … what a good argument looks like? (Purpose – to find out if they know the criteria for success / mastery goal)
- What does your teacher tell you to work on in your writing? (Purpose – to find out students’ understanding of the feedback received)
- Who are you doing this writing for? (Purpose – to find out their understanding of audience)

The transcripts of the teachers’ instructions and interactions during the lessons and the students’ responses were divided into three parts. The first two parts included the extent to which the lesson aims and criteria for success were shared with and understood by the students. The third part related to the type of feedback given and the students’ understanding of it.
Questionnaires

For each of these parts of the lesson, teachers had previously indicated their confidence by rating the following items on the questionnaires. This section of the questionnaire read as follows.

How confident do you feel with respect to the following statements:
I know how to develop the important learning objectives for a writing lesson
I know how to make the criteria for successful learning in writing clear to my students
I know how to provide feedback about their writing to students in ways that support further learning

Each item was rated on the following scale of 1 – 6:

<table>
<thead>
<tr>
<th></th>
<th>definitely not confident</th>
<th>moderately unconfident</th>
<th>slightly unconfident</th>
<th>slightly confident</th>
<th>moderately confident</th>
<th>highly confident</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Results

The results for each set of observations are reported in sequence. The observations that took place prior to the professional development are reported first, followed by the observations that occurred after the professional development had been underway for at least four months. Finally, the analysis of the last three lessons is reported.

Observations Prior to the Professional Development

The analysis progresses through each of the three parts of the observed lesson. These parts include the extent to which the learning aims were shared, the students’ understanding of the criteria for success and the feedback received.

Students’ understanding of learning aims / intentions

The analysis of the learning aims and how well the students understood them are presented in Table 5.1. The table is organised according to how well the students understood what they were supposed to be learning about writing. The teachers whose students had the least understanding are at the top of the table and those teachers whose students had the most understanding at the bottom.

As can be seen from Table 5.1, whether or not a teacher was a literacy leader, or whether they felt confident to develop important learning objectives, were unrelated to how well the students understood the learning aims of the lesson. Similarly, neither experience in teaching nor year level of students was related to how well the students understood the lesson aims, so these attributes are not included in the table.

The learning aims, as written on the pre-observation sheet by all teachers, were focused on deeper rather than surface features of writing (See Table 5.1). The specificity of the aims as written was difficult to code reliably and so all aims are presented in full. Several aims were expressed in terms of activities rather than what students were to learn, but in all cases the learning aim could be assumed from the activity description. The apparent clarity with which the aims were written did not appear to relate to how well they were conveyed to the students and were not, therefore necessarily, a good predictor of whether the students would understand what it was they were supposed to be learning.
Table 5.1
Teachers’ Lesson Aims, Confidence Ratings in Developing Them and Students’ Understanding Prior to Professional Development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence rating - aims</th>
<th>Aims as written</th>
<th>Shared with students</th>
<th>Students’ responses¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>To help children to start their stories using an interesting beginning</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>Writing with the audience in mind as preparation for speeches later in term.</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Using English matrices from NZ curriculum exemplars “Audience Purpose” (impact and voice) at Levels 1iii, 2 and 3.</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>To introduce to the students the concept of persuasive writing</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>Introducing a new form of poetry</td>
<td>Partly</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Introduction of new poetry style</td>
<td>Yes. Not written.</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>Children will write a poem entitled “I like words like: … using interesting sounding – lip-smacking words</td>
<td>Partly. Not written.</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>Language choice in poetry. Brainstorming adjectives related to a topic / theme and then finding more interesting synonyms to “paint a picture with words”</td>
<td>Yes. Not written.</td>
<td>3</td>
</tr>
</tbody>
</table>

¹ Note: Coding for student understanding
1. Surface features only e.g. length, punctuation and spelling (may include some undefined “better writer” in the future, learning to write good stories
2. Some mention of lesson aims (less than half responses)
3. Mostly focused on lesson aims (more than half responses)

A better predictor of student understanding than the written learning aims was the clarity with which the aims were conveyed to the students during the lesson. When clearly conveyed, they were typically not stated immediately on starting the lesson, but rather the teachers began by linking prior knowledge about writing to the observed lesson through questioning the students. For example, Teacher 9 began the lesson (the fourth in a series on writing an argument):

Teacher: What are the main parts of an argument? The main things we are working on. [Child’s name?] One of them?

Student 1: Make sure your opinion is what you want to say

Teacher: Say what you want to say, you write your opinion. [Child’s name], what else?

Student 2: You give your reasons
Teacher: What do we give reasons for?
Student 2: So they know why or why not to agree with us or agree with it

The lesson introduction then continued with the introduction of the topic about which they were to write their argument, that is, whether a loved story character “Greedy Cat”, who was very overweight, should go on a diet. The students’ opinions and reasons about the topic were then organised to cover the order in which arguments are put with the strongest reasons first and the weakest last. The students’ interview responses indicated that they were very clear that this was what they were supposed to be learning.

Another lesson on writing arguments, but one in which the students did not understand the learning aims so well began in a different way (Teacher 4 - initial lesson on writing an argument).

Teacher: I want you to talk about this question in pairs, “Should schools start at 10:00 and finish at 2:30? You might want to modify it a wee bit.

…. [Discussion occurred about preferred stating and finishing times followed by paired discussion of opinions.]

Teacher: I want you to give me one positive about school starting at 10:00 and finishing up when you like it to and one negative. Who’s going to give me one? [child’s name]?
Child: I’ve got a negative. You wouldn’t learn as much.

The lesson introduction continued to focus on identifying positive and negative reasons and possible starting and finishing times rather than their connection to writing an argument. When interviewed, these students were unsure of what it was they were learning about writing.

None of the three teachers giving lessons on recounts (teachers 1, 2 and 3) shared the learning aims with their students. All began with the topic of the recount with most of the lesson focused on motivating the students to recall the event through reading them a story on the topic, telling them of their own personal experiences and asking the students to share in pairs. They were then told to write their story of what happened to them. When the students were asked by the researcher what they were learning about writing, their answers in all three classes referred to story length, getting better at spelling, and neatness.

Students’ understanding of mastery criteria
The extent to which students understood the criteria for success in relation to the lesson aims was closely related to how well they understood the learning aims / intentions themselves. Those teachers who were more explicit about the lesson aims, were mostly explicit about what constituted criteria for success in that form of writing and so the sequence of lessons in Table 5.2 is in the same order as Table 5.1, that is, lessons for which students had a poor understanding of the criteria for success are at the top of the table. The teachers’ confidence ratings in making the criteria clear to the students (column 2 of Table 5.2) were unrelated to the extent to which students understood the criteria.
Table 5.2
Teachers’ Confidence in and Explicitness when Sharing Mastery Criteria and Student’s Understanding of Criteria Prior to Professional Development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence rating</th>
<th>Criteria explicit in task introduction</th>
<th>Criteria explicit through activities/individual assistance</th>
<th>Students’ responses ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>No. General instruction to write first sentence</td>
<td>No. Mostly on mechanics &amp; prompting content related to topic.</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>No. Writing a story about a topic. Focus on generating content</td>
<td>No. Focused on topic of story and getting something written.</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>No. Focus on generating content.</td>
<td>No. Focus on generating content.</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Partly. Not written. In opening sentence should express an opinion</td>
<td>Minimal reference to learning aim. Unclear.</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Partly. Not written. Students told to note describing words to write poem from video.</td>
<td>Yes. Most assistance focused on qualities of poem form.</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Yes. Not written. Gave rules for poem form.</td>
<td>Partly. Focus on word quality within poem form.</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>Yes. Not written. Focused on generating appropriate vocab for poem</td>
<td>Yes. Focused on generating appropriate vocabulary for poem.</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Yes. Not written. Focused on generating appropriate vocab for poem.</td>
<td>Yes. Focused on generating appropriate vocabulary.</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>Yes. Written. Oral instructions focused on criteria for success developed with students.</td>
<td>Yes. All assistance focused on criteria for success as written.</td>
<td>5</td>
</tr>
</tbody>
</table>

¹ Note: Coding of students’ understanding
1. Focus on surface features
2. Some mention of deeper features unrelated to lesson aims but most focus on surface features
3. Some or all responses partially related to learning aim
4. Fewer than half responses related to learning aim
5. Half or more responses related to learning aim

Teachers 1, 2 and 3 focused their instruction time on generating content about the topic of their recounts and did not disclose what would constitute effective writing on that topic. As indicated above, they all read a story to the students on the topic, recounted their own experience and provided an opportunity for the students to talk about their experience of the topic in pairs prior to the writing. While these strategies can all be effective in motivating students, at no time were the criteria for effective recounts divulged to the students during the introductory activities, which took up most of the lesson time. Teacher 1 modelled a first sentence for the students, which was consistent with her aim “To help children to start their stories using an interesting
beginning”, but all her references during the modelling were related to the mechanics of constructing words, not to the qualities of interesting beginnings (her lesson aim). Her only references to the qualities of interesting beginnings during the lesson occurred when the students began to write and she gave an instruction to write their first sentence and to “think of an interesting beginning to your story”. Her assistance to individuals as they were writing included one suggestion to “start off with a bang” and three specific wording suggestions for starting. She also suggested that two students use some speech in their first sentence and told two others that they should not start with “once upon a time” because that beginning was for fairy tales.

She also gave many other suggestions during these activities that were not specific to the beginning sentence but did constitute strategies students could use to meet implicit criteria for successful recounts. These included repeated references to writing about attributes of the situation (e.g. visualising it, what they could hear, what they could smell, the temperature), attributes of the writer (e.g. what they thought, felt and did), together with urging students to write something down, to write about the topic with two general references, to making it interesting for the audience although how they might do this was not elaborated. Most of the individual assistance was focused on helping the students with the mechanics of writing. Given all this information, it was not surprising that the students believed that good writing of the type they were doing was neat, had a title, capital letters and full stops.

Although the analysis of this teacher’s lesson is presented in some detail, the patterns were evident in other lessons that focused on content rather than the writing aim with students giving similar responses about what successful writing of the type they were doing looked like.

In contrast, teacher 9 had worked during previous lessons to develop criteria for success in writing an argument with the students and displayed these on the board. The criteria included:

- You need an opening statement that gives your opinion
- Reasons for your opinion and examples to support your reasons
- Put your reasons in order from strongest to weakest.

The lesson instruction was all focused on the students’ progress in relation to the criteria in terms of developing their arguments about Greedy Cat’s diet. When interviewed, the students all referred to these criteria when describing what effective persuasive writing looked like.

Students’ understanding of feedback and feedforward

The analysis of feedback focused on the oral feedback given during the lesson itself to individual students and to the whole class when sharing their work at the end of the lesson. Students’ understanding of the feedback / feedforward was judged in terms of their response to the interview question, “What does your teacher usually tell you to work on in your writing?” An analysis of the oral feedback given to students, the teachers’ confidence in giving feedback in ways that support further learning and the students’ responses are presented in Table 5.3.
### Table 5.3
Instances of Teachers' Oral feedback, teachers' confidence in giving feedback and students' responses prior to the professional development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence Rating</th>
<th>Oral feedback</th>
<th>Students’ responses</th>
</tr>
</thead>
</table>
| 1       | 4                | Non-specific praise (x2)  
Praise for using speech (x1) starting with an action (x1) | Doesn’t talk to us about improving work (x2), Length (x1)  
Says “fabulous” or “interesting” stories (x1). |
| 2       | 5                | Comment on having something similar to others (x3) and something different from others (x1).  
Praise for expressing feelings (x1) and using descriptive language (x1) | Mechanics (x4)  
Make it interesting (x1)  
Can’t remember (x1). |
| 3       | 4                | Praise for starting sentence with something different (x2), long sentences with reference to audience (x1), including attributes of situation (x3) and emotion (x1).  
Told too many “ands”. | Mechanics (x4)  
Reference to content inclusion (x4) |
| 4       | 5                | Praise for making good points (unspecified) (x6), that sentence states opinion (x6), for making a good start (x2) and finishing (x1).  
Asked for reasons.  
Prompted to cross ‘t’s” and dot “i’s” and that expressed opinion a bit vague. | Spelling and working out hard words. |
| 5       | 4                | Non-specific praise (x5)  
Reminded need 4 adjectives (x1).  
Prompted to find better words (x3) and take out surplus words (x1) | Make sure it makes sense (x2)  
Punctuation (x2)  
Good words (x1) |
| 6       | 5                | Non-specific praise (x7)  
Praised for using good / better words (x5), writing something different (x1), writing about what like doing (x1).  
Prompted to think of better word (x3) | Work independently (x1)  
Prompts writing process (x2)  
Describing words (x1)  
Reference to different genre (x1)  
Reference to rules for a writing competition (x1) |
| 7       | 4                | Non-specific praise (x12), Praise for word quality (x18), different words from others (x2), gave reason for liking word (x1) and working hard (x1) | Fix up words and spelling (x3)  
Don’t know (x1)  
Have a go (x1)  
Make more sense (x1) |
| 8       | 3                | Specific praise with reasons related to learning aim (x4), other specific praise (x1) | Procedures for spelling (x9)  
Punctuation (x2)  
Unspecified check (x1)  
Make sense (x1) |
| 9       | 5                | Specific praise with reference to success criteria (x3).  
Prompts for story sense (x2) and that students check work against success criteria (x1) | Included feedback on a range of aspects of writing, mechanical, structural and content. |
Once again, the students’ responses about their understandings of what their teachers told them to work on were predictable from the feedback they received from their teachers and the extent of their understanding was unrelated to the teachers’ confidence ratings in giving feedback in ways that support further learning.

Teacher 9, whose students were writing the argument about Greedy Cat’s diet stopped the students a few minutes into the lessons and asked volunteers to read out their beginning sentences. When they did so, he asked the other students to give feedback based on the criteria previously developed. Some of his questions and comments included:

“Now, has C. got a great opening statement?” ....
“Do you know what he’s thinking, what his opinion is?” ....
“Has he given good reasons?” ....
“Why are they good reasons?” ....
“Yes, his reasons support why Greedy Cat doesn’t need to go on a diet”

Not surprisingly, these students talked about feedback related to the deeper features of writing, associated with particular lesson aims. In contrast, the students of teachers (Teachers 1, 2, 3 & 4) who gave non-specific praise and focused on spelling in their written feedback were more likely to refer to mechanics when asked what their teachers told them to work on. The students of those teachers in between these two extremes, who made some reference to using better words in their poetry (Teachers 5, 6 & 7) were more likely to refer to deeper features of writing such as making sense, although these were rarely in relation to the lesson aims.

Observations following professional development (one year or less)
The lessons analysed in this section were for those teachers who had been involved in the professional development project from between four and ten months. The only repeat observations of the same teacher were for Teachers 1 and 3 who had been involved with the project for only four months. These two teachers had been working on recounts since the previous observation. They did not complete new questionnaires. All other teachers had been involved with the project for approximately nine months.

Students’ understanding of learning aims / intentions
In these lessons, all the teachers explained the learning aims of the lesson reasonably clearly to the students, as a result of which none of the students referred only to mechanics when asked what they were learning about writing. Some gave very sophisticated answers, such as, “We are learning to describe a person without naming that person by giving enough detail so the person reading it can guess.” (Student of Teacher 15).
Table 5.4
Teachers’ Lesson Aims, Confidence Ratings in Developing Them and Students’ Understanding Following 4 – 12 months of Professional Development

<table>
<thead>
<tr>
<th>Teacher &amp; Year level</th>
<th>Literacy Leader</th>
<th>Confidence rating - aims</th>
<th>Aims as written</th>
<th>Shared with students</th>
<th>Students’ responses¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Yr 3-4</td>
<td>No</td>
<td>5</td>
<td>Making a piece of writing better by adding or replacing certain words</td>
<td>Yes.</td>
<td>2</td>
</tr>
<tr>
<td>11 Years 7-8</td>
<td>No</td>
<td>4</td>
<td>Initial observation sheet had series of activities, not aims. As shared with students, “We are learning to write a descriptive passage that creates a “picture” in the reader’s mind i.e. imagery.</td>
<td>Yes. Written and read but embedded in activity.</td>
<td>2-3 (2 DK, 2 partial, 3 coherent with aims)</td>
</tr>
<tr>
<td>12 Yr 7-8</td>
<td>No</td>
<td></td>
<td>To identify and use illiteration</td>
<td>Yes. Written</td>
<td>3</td>
</tr>
<tr>
<td>13 Years 7-8</td>
<td>No</td>
<td>4.5</td>
<td>Writing attention grabbing, descriptive informative orientations in the context of a recount</td>
<td>Yes. Written</td>
<td>3</td>
</tr>
<tr>
<td>1 Years 2-3</td>
<td>Yes</td>
<td>NA</td>
<td>Learning to do a recount – structure and order of events</td>
<td>Yes. Written</td>
<td>3</td>
</tr>
<tr>
<td>3 Yrs 4-5</td>
<td>Yes</td>
<td>NA</td>
<td>The ending sentence of a recount</td>
<td>Yes. Written.</td>
<td>3</td>
</tr>
<tr>
<td>13 Yrs 4-5</td>
<td>No</td>
<td>6</td>
<td>To think about the reader (audience) and purpose when writing character descriptions.</td>
<td>Yes. Oral</td>
<td>3</td>
</tr>
<tr>
<td>15 Yrs 5-6</td>
<td>Yes</td>
<td>6</td>
<td>The opening sentence in a description to tell reader about the big picture of what describing</td>
<td>Yes. Oral</td>
<td>3</td>
</tr>
</tbody>
</table>

¹Note: Coding for student understanding

1. Surface features only e.g. length, punctuation and spelling (may include some undefined “better writer” in the future, learning to write good stories
2. Some mention of lesson aims (less than half responses)
3. Mostly focused on lesson aims (more than half responses)

The changes in teacher one’s lesson illustrates some of the changes. In the observation four months later, teacher 1 coincidentally was teaching recounts with a focus on structure and order of events. The context was a birthday party. In this lesson she used the context to outline a structure for the recount, revising the qualities of beginning sentences, then instructing the students, “… so we’re writing down everything that happened in the order that it happened.... You can’t have the thing that happened last at the top can you because it won’t make sense.... And what comes at the end?” After structuring some students’ stories of birthday parties into beginning, middle and endings with the whole class, she then set the students the task of writing about a birthday party they had been to. All assistance given to the students was focused on the learning aims and mastery criteria.

When interviewed, the students in both classes were able to articulate the learning aims with a focus on the deeper features of writing. They spontaneously indicated how much they enjoyed writing. They were less clear in their responses to questions about feedback, however, with less than half the students mentioning
features of recounts. It appeared that the students’ understanding was closely related to the specificity of the different attributes of the lesson.

Teacher three showed similar changes in the repeat observation of the lesson on recounts. In her introduction during this second lesson, she began by focusing on the writing aim rather than the topic of the writing.

Teacher: First of all, can someone tell me what a recount is
Child: It’s when you have an opening sentence and then you have to have the events that ...
Teacher: Stop there, remember we talked about what a recount is and there’s a sentence that sums it all up. Would you like to tell me [child’s name]
Child: …. It’s writing something that’s already happened to you so someone else can read it.

Student’s understanding of criteria for success
Both teachers 1 and 3 continued their lessons by revising the structure of recounts. Teacher 3 focused the discussion ending with the qualities of the final sentence, the aspect on which the lesson was focused. She read some of the students’ work checking it against the agreed criteria for the final sentence. When the students were interviewed they all said that they were learning to write recounts and clearly articulated the features of recounts. One also referred to the audience and another more vaguely about learning how to do it if he wanted to be a writer.

As can be seen in Table 5.5, most of the students from both classes mentioned criteria closely related to the learning aim and focused on the deeper features of writing. Overall, the more consistent the teacher was in relating the different features of the lesson to the learning aim, the clearer the students were about what they were learning.
Table 5.5
Teachers’ Confidence in and Explicitness when Sharing Criteria for Writing Related to Aims and Students’ Understanding of Criteria Following four – twelve months of Professional Development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence rating</th>
<th>Criteria explicit in task introduction</th>
<th>Criteria explicit through activities / individual assistance</th>
<th>Students’ responses¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>Partly – but non-specific “to make the description more interesting”.</td>
<td>Partly – but not specific enough to provide guidance for students</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>4.5</td>
<td>Yes. Written and re-stated but only partly aligned to the learning aim. SC focused on identifying and using language features and re-crafting for fluency and effect.</td>
<td>Yes, but links to LI and task unclear. Fluency equated with punctuation. Language features equated with creating a picture.</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>No but very transparent in task introduction</td>
<td>Yes. Repeated activities reinforcing one another.</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>Yes for some aspects, rest to be made into explicit criteria next lesson.</td>
<td>Mostly but indicated would make into explicit SC next lesson.</td>
<td>5</td>
</tr>
<tr>
<td>1 (repeat observ)</td>
<td>NA</td>
<td>Yes. Written and explained.</td>
<td>Mostly. Integrated content with structure of recount but focused on first sentence only.</td>
<td>5</td>
</tr>
<tr>
<td>3 (repeat observ)</td>
<td>NA</td>
<td>Yes. Written. Oral instructions focused on criteria for success.</td>
<td>Yes. All assistance focused on criteria for success as written.</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>Partly in that not written as success criteria but all oral activities, instructions and discussion explicitly focused on criteria for success</td>
<td>Yes. Consistently referred to descriptive features.</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>Partly in that not written as success criteria but all oral activities, instructions and discussion explicitly focused on criteria for success</td>
<td>Yes. consistently referred to desirable features opening sentences of a description</td>
<td>5</td>
</tr>
</tbody>
</table>

¹Note: Coding of students’ understanding
1. Focus on surface features
2. Some mention of deeper features unrelated to lesson aims and most focus on surface features
3. Some or all responses partially related to learning aim
4. Fewer than half responses related to learning aim
5. Half or more responses related to learning aim

Students’ understanding of feedback / feedforward received
Feedback was less consistent with the deeper learning aims than other aspects of the lessons (Table 5.6). While it is understandable that the written feedback would, at times, refer to mechanics, an exclusive focus on mechanics does not tell the students what they should be working on to achieve particular learning aims.
### Table 5.6
Oral and written feedback, teachers confidence in giving feedback and students’ responses following between four months and one year of professional development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence Rating</th>
<th>Oral feedback</th>
<th>Students’ responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>Praise for: Removed / replaced words – no reason (x4)</td>
<td>Punctuation (x1) Write good stories (x1) DK (x1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Replaced words – reason “more interesting” (x4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Word qualities e.g. rhyming (x3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adding words – with reason (x2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adding humour (x2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use of conjunctions (x1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Added details (x1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggestions for change: Avoid repetitive structures (x1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reduce length of introduction – reason more interesting (x1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Make simile more precise (x1)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4.5</td>
<td>NA (students working on writing task on own)</td>
<td>Spelling (x2), Handwriting (x1) Form letters properly (x1) Changing words to make them better (x1)</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>Only in form of confirmation of responses</td>
<td>Language features (x2) (most unintelligible on tape).</td>
</tr>
<tr>
<td>13</td>
<td>No rating available</td>
<td>Only short time available. One piece of specific next steps. One piece of more general next steps.</td>
<td>Vocabulary, recrafting, planning, audience related to LI (x4) Tense in recounts (x2) Punctuation (x1)</td>
</tr>
<tr>
<td>1</td>
<td>No rating available</td>
<td>Non-specific praise (x3), praise for qualities of first sentence related to learning aim (x8), ending (x1). Prompted to make sure sentence made sense (x1) Word corrected (x1), told not to press so hard (x1), punctuation (x2)</td>
<td>Non-specific writing related (x2) Not sure (x1) Punctuation with reference to sense (x1) Making sense (x1) Sounds out words (x1)</td>
</tr>
<tr>
<td>3</td>
<td>No rating available</td>
<td>No feedback except individual students prompted to include why (x2), what (x2), when (x2), who (x2), read sentence to make sure it made sense (x1).</td>
<td>Features of recount (x4) General reference to recount (x1) Expression (x1) Mechanics (x2)</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>NA - None given during lesson</td>
<td>All referenced to deeper features of either character description or audience.</td>
</tr>
<tr>
<td>15</td>
<td>5-6</td>
<td>Specific praise and asking for more information related to success criteria.</td>
<td>All apart from one related to deeper features of SC.</td>
</tr>
</tbody>
</table>
Patterns of achievement
In case there is concern that a focus on the deeper features of writing will leave students without the mechanical skills to write well, the results from two schools which have shown improvement in their asTTle scores, show that as scores for deeper features have improved as a result of focusing on them, so have mechanics. For example, in the school in which we undertook repeat observations, achievement gains showed an effect size of 1.04 over four months. Surface features improved as much as the deeper features. In another school which showed gain in asTTle scores, the effect size over a period of six months was 1.14 with surface features also improving as much as the deeper features.

In the only other school for which we have asTTle scores, students writing did not show improvement over the year. The teachers in this school had students who were less clear than others about the learning aims and success criteria related to the lesson. Becoming more explicit about what it is teachers want the students to learn and the criteria for success therefore is no guarantee that students will immediately understand. Making the needed changes is a complex process and in the next section we analyse two lessons that show some of the complexities involved and the importance of making clear links between the different aspects of lessons.

Coherence and leaps of inference
In this section, two lessons are analysed to illustrate some of the more subtle aspects of coherence between the features of lessons that superficially have all the analysed components, but to establish how they were linked took considerable interpretation on the students’ behalf. Both lessons involved re-crafting writing, which may have made the teacher’s task more difficult and required them to draw on a greater depth of content knowledge than presenting the initial introduction. In each of these lessons, students were confused about how they might undertake the desired re-crafting effectively.

Teacher 11 stated the lesson aims and success criteria clearly, but the links between them and the tasks the students were asked to complete were not obvious and, thus, confusing for some. For example, in this lesson, the third in a series, the stated learning intention during the lesson was “To write a descriptive passage that creates a ‘picture’ in the reader’s mind” with most of the lengthy introductory activities asking students to read a short description supplied by the teacher, draw a picture and identify the language features used in the description. Towards the end of the introduction, the teacher briefly reviewed the purpose of different forms of punctuation.

The written success criteria which she then reviewed with the students stated, “We will know we have achieved this [the learning intention] when we can:

- Identify language features within own and other’s writing
- Can incorporate some language features into own writing and can talk about these
- Can re-craft text for fluency and effect.”

The relationship between the learning intention and criteria for success and the activities the students were asked to engage in was not explained to the students at this time. After the introduction, the students were given the following task. “You’re going to go away now and you’re going to go back and you’re going to see what punctuation and how you’ve organized your writing that you’ve done thus far, and see how you can re-craft and improve on it, because when we re-craft, we want fluency and effect out of it, don’t we? When
you’re re-crafting, we need to have fluency, but we also want your passage to be effective, don’t we? So we’re looking at our language features here. We’re only going to be successful with this when we do incorporate some language features. Otherwise, instead of “the red cliffs”, it will come across as just “the cliffs”. And we need fluency and that’s where our punctuation comes in.”

Although the links between identifying and using language features and the learning intention of creating a picture, punctuation and fluency, and re-crafting may have been obvious to the teacher, she left it to the students to infer these links rather than making them explicit. Not surprisingly, when the students were asked what they were learning about writing descriptive passages in the interviews, their answers were mixed. One student reiterated the learning intention and two referred to use of language features. One other was less clear referring to “putting things in paragraphs” with two others indicating they were “Not sure”. Similarly, when asked about the features of a good description, responses were again mixed. Three students responded “Don’t know” or “Too hard”. Another student referred to, “Descriptive stuff like metaphor and personification” but appeared to have little idea about how they might be used to effect. Two other students referred more directly to the learning aim or some related aspect, such as, “like put the picture in your head, like what you’re writing about, so you can see it.”

The students in teacher 10’s class were also learning to re-craft their descriptions from the day before. The lesson began with a statement of the learning intention “To make our writing better by adding or replacing certain words.” After eliciting from the students how they might make a piece of writing more descriptive by deleting words, making the words more interesting, rewriting the sentence or making them into more complex sentences, the students were then, in groups, set the task of re-crafting a short paragraph written by their teacher. No criteria were given for judging whether the substitutions, additions or deletions were effective beyond a number of general statements and questions by the teacher relating to “making the passage more interesting”. Implicit in the task instruction and assistance given to students, success consisted of joining sentences to make them more complex and substituting supposedly more interesting words. No specific reasons were given for doing so.

Students were then set an independent task of editing their own work “to make it better”. Again, no specific criteria were given for success in this activity, so we analysed the transcripts for implicit messages in the task instruction, assistance given to individual students and feedback. Success was essentially defined in terms of changing or adding words without any specific reason given for doing so (14 references) and making longer sentences (3 references). Three references were to adding more information. Additional criteria were also mentioned as the teacher assisted individual students. None of these were mentioned more than once and included adding an explanation mark, using a different type of simile, putting in a movie character, putting in something people can relate to or, removing a word, not starting two sentences the same way, adding humour and making a simile more precise. Again, no reasons were given for doing so. One student received non-specific praise and two others were praised for using “good language” and “good adjectives”.

Not surprisingly, when students were asked what they were learning, they mostly referred to process, independent of criteria, such as, “Describing a lot and breaking up sentences” (in direct contradiction to the teachers’ emphasis on joining sentences), “Put in describing words and make it make sense”, and “Fix the words”. One referred to more accurate spelling. When students were asked about success criteria and
feedback, any references to deeper features were of a general nature, such as “write good stories” and “make our stories make sense”, or the specifics of surface features such as neatness and punctuation.

Observations Following Two Years of Professional Development

Only three teachers were able to be observed at this time due to school attrition and the timing of the writing of this report. These three teachers (10, 13 and 15) had already been observed at the end of their first year of involvement in the contract. Two teachers (10 and 15) included re-crafting in their lessons, an aspect of practice that had proved problematic in the earlier observations. The three tables describing lesson aims, success criteria and feedback of the observed lessons at the end of two years are presented together because all these aspects of the lessons were closely aligned.

Table 5.7
Learning aims, teachers’ confidence ratings and students’ responses after two years of professional development

<table>
<thead>
<tr>
<th>Teacher &amp; Year level</th>
<th>Confidence rating - aims</th>
<th>Aims as written</th>
<th>Shared with students</th>
<th>Students’ responses¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>No specific aims written. Each student had written feedback on previous attempt at writing an argument</td>
<td>Students each had a “where to next” in their books.</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>Plan our ideas using key words and to use ideas that will benefit the reader</td>
<td>yes</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>Identify emotive / opinion language in models to incorporate these in our pre-written paragraph</td>
<td>yes</td>
<td>3</td>
</tr>
</tbody>
</table>

¹ Note: Coding for student understanding

1. Surface features only e.g. length, punctuation and spelling (may include some undefined “better writer” in the future, learning to write good stories
2. Some mention of lesson aims (less than half responses)
3. Mostly focused on lesson aims (more than half responses)
Table 5.8  
Success criteria teachers’ confidence ratings and students’ responses after two years of professional development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence rating</th>
<th>Criteria explicit in task introduction</th>
<th>Criteria explicit through activities / individual assistance</th>
<th>Students’ responses ¹</th>
</tr>
</thead>
</table>
| 10      | 4                 | Not written. Recalled main features of an argument and attributes of audience. Analysed example of argument that had insufficient elaboration. Explicitly identified elaboration, rhetorical questions and imperatives as reworked example with students. | Elaboration (x5)  
State opinion more clearly (x3)  
Use emotive language (x2)  
Rhetorical question (x1)  
Make sure claims are accurate (x1) | 5 |
| 13      | 5                 | Not written. Constantly referred to need to engage audience and gave reasons for planning. | Key words to identify reasons (x3)  
Clarifying reasons underpinning argument (x3)  
Reference to knowing audience (x1) | 5 |
| 15      | 6                 | Not written. Used short written argument to identify powerful / emotive vocabulary. Built word bank from students’ vocabulary. Re-crafted own writing using word bank. | No individual work but all class work aligned to identifying powerful / forceful words to make an argument stronger. | 5 |

Coding of students’ understanding
1. Focus on surface features
2. Some mention of deeper features unrelated to lesson aims and most focus on surface features
3. Some or all responses partially related to learning aim
4. Fewer than half responses related to learning aim
5. Half or more responses related to learning aim
Table 5.9
Oral feedback, teachers confidence in giving feedback and students’ responses following between four months and one year of professional development

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Confidence Rating</th>
<th>Oral feedback</th>
<th>Students’ responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6</td>
<td>Rhetorical question, Elaboration (x2) Give reasons Better picture of opinion (x2) Opinions need elaboration (x3) Change words to make them more emotive Reasons link to opinions</td>
<td>Keep to the point in each paragraph Main points need elaboration (x2) Not as many adjectives because gets confusing Spelling</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>No specific feedback given – planning emphasis resulted in teacher helping students to organise their ideas.</td>
<td>Audience (x3) Reasons underpinning argument (x3) Structure Planning ideas It’s persuasive and pulls you in (x2)</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>Boring and forceful words (x3) Must make sense</td>
<td>Lots of different things Elaboration Planning to work out the main point Putting more verbs into writing to make better sentences.</td>
</tr>
</tbody>
</table>

 Whereas in the earlier description of teacher 10’s lesson, the criteria for judging whether a particular change would improve a students’ writing were unclear, in this lesson they were much clearer. She had given the students the option to argue for or against keeping birds in cages. After working through some criteria related to arguments, the teacher then asked a student if she could read out his work.

**Teacher:** “Dear Editor, I belong to the KBIC: Keep Birds in Cages Company. How would you feel if birds became extinct?” [pause] What’s that an example of?

**Student:** Rhetorical question

**Teacher:** Mmm. “Firstly, birds survive much longer in cages. They have fresh water and food every day and also get cleaned. The birds don’t need to hunt; it is very easy.” Is there a reason and an elaboration there?”

Her comments were all aligned to the criteria for effective arguments. As she circulated among the students while they were editing their work, all comments related to criteria she had previously covered in the lesson. For example, after reading a student’s minimal elaboration of a reason, she said, “If you’re saying that they are dying, you need to explain. What happens to cause them to die?” And to another student she said, “Good, you’ve explained your point. What’s your next point?”
With respect to the use of emotive language, she said to one student,

**Teacher:** I know you’ve got some fabulous ideas about emotive words. Let’s go right back to the beginning. Well instead of saying, “We think it is wrong not to keep birds in cages”, how could you change that to make it stronger, putting in a word that would make the cage sound really nice.

**Student:** Not to be ... to be protected in cages

**Teacher:** Well done, OK.

The students responses indicated that they were much clearer about the lesson aims, success criteria and feedback. Re-crafting was also the focus on Teacher 15’s lesson but in this case the focus was on vocabulary and audience. As with Teacher 10, her comments with respect to recrafting were aligned with the learning intentions. As she indicated in her introduction, “We are identifying opinion vocabulary, the vocabulary that tells the reader this is what we think.” She then read a brief argument about too much homework that began “Enough is enough”, and asked, “What makes it opinion vocabulary? What gives us an impression of how the writer is thinking?” When she had finished reading the short piece on homework she asked, “What’s the difference?” A student replied, “They use emotive language like forceful words.” She then took a statement from a student’s writing, “If you buy more sports equipment for our school we will have healthier students” and worked with it to make it more powerful. The final product read, “If you buy a range of sports equipment for our school there will be a dramatic increase in student health levels.”

The student’s interview responses at both times indicated that they understood all aspects of the lesson in both lessons. A similar pattern was evident in the students responses to Teacher 13.

**Conclusion**

In this chapter, we have focused on the aspects of classroom instruction that are identified in both the self-regulated learning and formative assessment literatures as likely to be most powerful in providing the conditions for students to self-regulate their learning and become successful writers. Our focus was on the interrelationship between teaching and student learning because, as Butler and Cartier (2004) argue, teachers influence students’ task interpretation by virtue of how they structure learning environments. Aspects of the learning environment in this chapter included providing students with clear learning goals for instruction, criteria for successful task accomplishment and feedback that promoted further learning. We were not looking to see if teachers were explicitly training meta-cognitive SRL strategies as such, but rather examining whether the instructional environment became more effective throughout the professional development in providing the conditions that would allow students to develop self-regulatory processes. Given the importance of these conditions in promoting learning success, it is reasonable to assume that such conditions would be desirable in all learning situations. It appeared that the professional development was effective in assisting teachers to create these conditions for learning with an increasing number of teachers showing proficiency over the period of the professional development.

Fundamental to the self-regulatory process is students’ understanding of learning goals and what it means to master them because without such an understanding they are unable to monitor their progress or to generate
relevant internal feedback. In many of the initial observed lessons, despite the teachers having written substantive learning aims, these were not conveyed to the students and it was difficult for the observer, let alone the students, to deduce from activities during the lesson what was to be learned or what mastery of this learning might look like. In the space of each individual lesson, links between teacher actions and students’ understanding were able to be powerfully demonstrated.

When the learning and mastery goals of the instruction were unclear, students typically focused on surface features of the writing task or had some general learning aim such as becoming a better writer. It was as if students adopted these features as a default position when they were unclear about more sophisticated concepts of writing giving weight to Butler and Cartier’s (2004) contention that, “If task interpretation is absent or faulty, learning is derailed” (p.1735). It is well known that struggling writers often think of writing as a task that involves correct spelling and grammar, with the product neatly presented (e.g. Graham, Schwartz, & MacArthur, 1993; Wong, 1999), yet it was these features that were unwittingly promoted in many of the observed writing lessons prior to the professional development.

It could be argued that to become self-regulating, students should have the freedom to select their own goals, rather than meeting those pre-determined by their teachers. But from these observed lessons, it appears that until students have an in-depth understanding of what it means to be a writer for a particular purpose, learning is more likely to be progressed if teachers assist students to understand the necessary learning goals and mastery criteria so that they can monitor their learning progress.

This study has clearly shown the link between student understanding and instructional processes. It has also shown that rather than assume this understanding, it is vital to check the students’ perspectives. The problem with relying on an analysis of teaching practice in order to judge effectiveness was highlighted in the example of the teacher who appeared to present the learning aims and mastery criteria and provide learning-related feedback but her students failed to understand them. A more detailed analysis of this lesson revealed the mixed messages and unclear criteria evident during the instructional process.

Further, this study suggests that if teachers convey a learning aim that is understood by their students, they are also more likely to articulate mastery criteria and provide feedback that relates to these qualities. Indeed, integration and alignment was more important than just having a stated learning aim by itself. The ability to achieve either the separate parts or their integration did not appear to be related to self-reported confidence or teaching experience. Rather, it would seem that considerable knowledge of both the writing process and how texts work for particular purposes is necessary.

It appeared that one of the most difficult tasks was to provide students with clear criteria and processes against which to re-craft their writing. It was evident by the end of the second year that one teacher at least, who was struggling to master this skill at the end of the first year, showed much greater success at the end of the second year.

If students have experienced low achievement and have found the task of writing confusing, the results from this study indicate that those patterns of achievement can change if teachers reveal the secrets of how to be successful. Some students are able to work out from their teachers’ implicit or confused messages what it is
they are supposed to be learning and what mastery looks like and consequently have access to the conditions that will allow them to self-regulate their learning. But, for many, these features need to be made more explicit to help them understand what they are working towards and to realise that writing is more than getting the punctuation and spelling right in long and neatly presented pieces.

References


Chapter 6

Teachers, Schools and Using Evidence: Considerations of Capacity

If the key to better learning is better teaching (Darling- Hammond, 2000), then the key to better teaching is better understanding of the nature of expertise (Sternberg & Horvath, 1995). This notion is clearly reflected in international work identifying the features of quality literacy teachers and teaching. What constitutes effective literacy practice or what effective practitioners of literacy know or do has been expressed in slightly different terms in numerous publications (e.g. Allington, Johnston & Day, 2002; Block & Mangieri, 2003; Pressley, Allington, Wharton-McDonald, Block & Morrow, 2001; Wray, Medwell, Fox & Poulson, 2000). A logical stance to take, though, is that effective practice is not something that is not absolute but is achieved by knowledgeable, committed teachers who tailor and adapt their practices to the ongoing needs of their learners to achieve outcomes of a high standard across heterogeneous groups of students (Alton-Lee 2003).

Having an evidence-informed and well-articulated knowledge about what one is teaching (the content), about how to teach and about one’s students underpins effective teaching. Knowledge of the learner involves identifying patterns of strengths and weaknesses; looking backward at what has been done, to assess the effectiveness of instruction in terms of rate and extent of progress, and looking forward to work out what to teach next. Knowing a student’s present level allows a teacher to work on learning goals just a little ahead of independent performance, in the region of sensitivity where the skills and knowledge are perhaps in embryonic form, to build on what the learner is just starting to do (Vygotsky, 1978). The teacher has to design graduated support so that the goal is attainable and so there is a match among the task requirements, the skill of the learner and the support provided.

Ongoing assessment informs and guides instruction (Black & Wiliam, 1998; Crooks, 1993; Torrance & Pryor, 1998; Tunstall & Gipps, 1996), allowing better or more accurate decisions to be made (Stoll, Fink & Earl, 2003). Closely analysed evidence about the learning of students allows deliberate adjustments to a classroom teaching program in order to better meet the needs of students. A careful matching of task and instruction to student competence levels was a feature of the outstanding first grade teachers in Pressley and colleagues’ studies (2001). Resources and materials should be matched to established learner needs and evaluated against these needs. Effective practitioners consider resources or programmes in relation to the needs of their students and test using evidence of achievement whether these work in their context. Research suggests that when schools engage in in-depth analysis of assessment information to assist them to modify their program, then student achievement is raised (Newman, King & Rigdon, 1997; Symes & Timperley, 2003).

Despite assessment being viewed as integral to school programs, there are questions about the preparedness of schools and teachers to engage in the process (Earl & Katz, 2002; Gusky, 2003; etc). The capacity of teachers to be evidence-informed about their learners in terms of the skills associated with collecting, using and interpreting evidence of student achievement has been raised in several recent reports associated with different research evaluations of innovations (Parr, Aikman, Irving & Glasswell, 2003; Timperley &
Wiseman, 2002; Timperley, Parr & Higginson, 2003). Routine use, particularly of student achievement data, in relation to honing practice to better meet the needs of students, may not be part of the professional canon or skill-set. Research documenting the implementation of ready-made literacy materials into primary schools suggests that evidence of need does not necessarily inform the choice of materials; that changes in recommended or routine use are seldom evaluated and that the use of student achievement data is rarely used to make decisions about the veracity of the materials (Parr, Aikman, Irving & Glasswell, 2003). Similarly, in a project where schools set literacy goals and then designed, implemented and evaluated in an action research type cycle, a classroom initiative to monitor progress towards these goals, virtually none of the schools studied collected evidence that could be considered adequate, in that an independent judgement could be made on the basis of it (Timperley, Parr & Higginson, 2003).

Such undertakings may require some shifts in beliefs about the value of evidence and in professional norms and discourse (Annan, Lai & Robinson, 2003). At the most fundamental level, schools and their teachers may hold a different theory about the form that evidence might take or how it should be used (Timperley & Parr, 2005). There are also other possibilities to be investigated with respect to preparedness to make use of evidence, namely, whether teachers and their leaders possess the requisite knowledge and skills. In this study we aimed to investigate three aspects of knowledge that, arguably, are necessary in order to use assessment evidence. These are knowledge of how to interpret what data show and what to do with this information, in the sense of how it might be applied to inform instruction and the more abstract knowledge of how and why this might operate, knowledge of the principles of evidence-informed decision making. The study examined the nature and extent of the knowledge of both data interpretation and use and of these principles. Signs of any changes in knowledge as a result of engaging in a literacy professional development project where there was a focus on professional learning around how to work with and utilise data were examined. In addition, the study explored the relationship between facility in these aspects and student achievement in terms of progress.

**Method**

**Participants**

Nationally, there were two intakes of schools, cohort one beginning at the start of the school year in February and the second, smaller cohort, starting in July of the same year. Research schools, numbering seven and six, respectively, were drawn from each cohort. Teachers of Years 1 to 8 from thirteen schools, largely urban, participated in the research. As this was a school-wide professional development project, principals and senior management as well as classroom teachers participated.

For participating schools, this was potentially a two year project although there was an option to complete only a year. In cohort 1, there were data available from 95 teachers at both Time 1 and Time 2, the beginning and end of the first year of the project. Data at three points in time were available for only 33 participants from cohort 1 as two large schools completed the one year option for the project. For cohort 2, there were data at two points in time from 22 teachers. As a group, the teachers were experienced, around a half had over 10 years experience while only a small percentage 10-15 were early career teachers with up to two years experience. Over 60 percent of the total sample held a degree as their basic qualification.
Procedure and Materials
In the first year, data were collected near the beginning and end of the school year. With respect to the project’s aim to create self-sustaining learning communities within schools specifically with respect to literacy, schools were seen to advance through three phases. Most schools were just entering phase 2 of the project when the second round of data was collected towards the end of their first year. A third round of data collection took place at the end of the second year on the project for cohort 1.

Teacher Knowledge
The primary source of data for teacher knowledge was a questionnaire that all teaching staff and senior management who were present completed at a staff meeting. Similar forms of this questionnaire were used as responses were gathered at the three points in time. At Time 1 and Time 2 for cohort 1, there were two types of scenarios. Scenario one aimed to examine the specific knowledge required to interpret, then use student achievement data and it was employed at all time points (three points for cohort 1 and two for cohort 2).

The second type of scenario involved examining more abstract, conceptual knowledge of the principles that operate in evidence-informed decision making.

Interpreting data
The first type of scenario (see Appendix B for an example - the Time 2 scenario for reading) asked respondents to interpret hypothetical data from a small class of students (N = 20) from either a reading or a writing assessment. These hypothetical assessments were similar in components to the STAR and asTTle tests schools would use to gather student achievement information in the project. Basically, the data contained subtest results and total raw scores for individuals and the class. For each subtest and the total score, the national mean and expected range and the class mean and range were given. Respondents were asked to interpret the class data for a junior colleague. At both Time 1 and Time 2, comparable questions were asked. The questions concerned what the data showed with respect to a specific aspect of reading comprehension or writing; what the main points were that could be taken from these data and, importantly, what advice they would give about what the colleague should do, given these data. At Time 1 for cohort 2 and at Times 2 and 3, this latter aspect was examined by asking respondents to rate the usefulness of the data in terms of using the results for teaching and to explain the rating.

In addition, teachers were asked to self-report their level of confidence in analysing data from reading or writing tests using a six point Likert scale.

Principles of evidence-informed decision-making
The second type of scenario (see Appendix B for an example- the Time 2 scenario for reading schools) was deliberately constructed so as not to meet certain criteria for evidence-informed practice on a number of major dimensions. The scenarios employed at Time 1 and 2 were not directly comparable. The Time 2 scenario was a revised version that addressed all the major dimensions of evidence-informed decision making so this scenario is employed to focus on the relative strengths and weaknesses in teacher’s knowledge at the higher level of principles.
The scenario was deliberately constructed so as not to meet criteria for evidence-informed practice on a number of dimensions.

Part A of the scenario related to the identification of students’ needs. In this scenario, these needs were based on teachers’ expressed concerns about reading comprehension/writing cohesion rather than using achievement data to determine the extent to which reading comprehension/writing cohesion was a problem. This basis for identifying needs also made comparison with outcomes difficult to determine.

Part B focused on the needs / programme match. The basis for the decision to adopt a particular programme (in this case a “tape assisted reading programme” or “peer tutoring: writing”) was that the students had enjoyed the programme in a neighbouring school. As evaluators, we considered this to be insufficient evidence of a programme that was likely to address comprehension/cohesion difficulties.

Part C addressed meeting processes. In the scenario, the focus of meetings was to address self-identified organizational issues, rather than challenging and addressing any concerns that the leadership might have of teachers’ professional knowledge about this, or any other approach to reading/writing. While organizational issues are important, it is unlikely that the processes described would enhance teachers’ professional knowledge of reading/writing instruction, which was a major focus of the literacy professional development initiative.

Part D focused on the match between the initial need (comprehension/cohesion) and the assessment used to determine progress and the decision to continue with the programme. The scenario specifically mentioned improvement in “reading fluency”/ “amount written” as the basis for the decision to continue with no mention of data on comprehension/cohesion – the originally identified need.

Teachers were asked to rate how effective they thought the school in the scenario seemed to be in their approach to the four areas listed above. A rating scale of 1 – 6 was used, where 1 represented ‘highly ineffective’, 2 represented ‘moderately ineffective’, 3 represented ‘slightly ineffective’, represented ‘slightly effective’, 5 represented ‘moderately effective’, and 6 represented ‘highly effective’. For some analyses, this rating scale was reduced to three points for the sake of clarity with ratings of 1 and 2 renamed as “low”, ratings of 3 and 4 renamed as “medium” and ratings of 5 and 6 renamed as “high”. Reasons were asked for all ratings.

It was reasoned that teachers would reflect on knowledge (and practice) both through responding to the questionnaire scenario and then, subsequently, through considering their responses in relation to those of their peers in discussion forums. In addition, they could relate the teacher actions and their responses and ideas to information obtained in professional readings, specifically with regard to effective literacy practice. Further, as an integral part of the professional development in literacy programme, they would work with student achievement data to understand student needs and to inform teaching.

**Student Achievement Data**

In reading, data were from a standardized test of reading, Supplementary Test of Reading (STAR) (NZCER & Elley, 2001). There are three test versions for different year groupings each with an A and B form. All
have common sub-tests, namely word recognition, sentence comprehension, paragraph comprehension and vocabulary. The test for Years 7 and 8 has two additional subtests while the common tests for Years 4-6 contain more items than the Year 3 test. In writing, data were obtained from a criterion referenced (to NZ curriculum) measure of writing (asTTle) that has associated national normative data (Yrs 4-8). There are scoring rubrics for each of six writing purposes with criteria for different levels of achievement for seven dimensions of writing (audience, structure, content, language resources, grammar, spelling and punctuation).

**Data Analysis**

The analyses of the data were based on the following research questions:

**Scenario 1**

1. To what extent are responses accurate and referenced to data? (Responses were coded in terms of the accuracy of the interpretation, given the data and whether the interpretation given was referenced to the data).
2. What is the nature of change, as a result of participation in the project where use of evidence is emphasised, in responses to the questions about the scenario?
3. What is the relationship between knowledge of data interpretation and use, confidence in interpreting data and student achievement progress? (The analyses in order to answer this question involved calculating a score for data analysis and use. Basically, for each question, accurate but non specific or answers that used information selectively, neglecting perhaps to mention major points, scored one point, while an accurate, data referenced point scored two points (maximum score 3 points even if two data referenced points). Then, student achievement progress scores were calculated for each teacher’s class. Gains in raw score were used).

**Scenario 2**

1. To what extent did respondents, as a whole, agree that the dimensions depicted in the scenario represented relatively ineffective evidence-informed practice? This question was answered by examining the ratings given to the different aspects of the scenario by all respondents grouped together.
2. To what extent were the reasons given for particular responses based on evidence-focused attributes of the scenario? In order to answer this question all reasons for ratings were coded with the coding categories reflecting these attributes.

Reliability of coding for both scenarios was established by calculating the percentage of codes that are coded the same between markers. Where one coder allocated a code to a point and the other did not, this was counted as a disagreement. Agreement ranged from 71% to 89%. For each question where the level of agreement on a sample of 20 percent of questionnaires fell below 80 percent, all responses to that question were double coded and disagreements between coders discussed and resolved.
Findings

The level of teacher knowledge with respect to data interpretation obtained from responses to scenario one is considered first; then responses to the second scenario that examined knowledge of more general principles underpinning the use of evidence are presented.

Scenario 1: Knowledge of Data Interpretation and Use

The responses to the three questions (with respect to a specific aspect; the main points and what to do with the information) were coded using similar categories that allow us to see the extent to which responses were accurate and specific in that they did not simply refer in a generalised way to trend but responded in a manner that indicated they had based the interpretation on evidence. The questions are considered separately, then overall patterns noted.

The first question asked for what the data for a specific aspect of reading or writing showed. Figure 6.1 shows that a similar percentage did not respond at both points in time while the percentage responding with an answer that was either inaccurate or muddled fell by half between Time 1 and Time 2. Similarly, the percentage of teachers giving a generalised interpretation with no reference to data fell to 10%. Most significantly, the percentage referencing their points to data rose from 37% to 63%.

Figure 6.1
Question 1 Percentage of Teachers Responding to Question 1 by Category at Time 1 and Time 2

For question 2 that asked what main points should be taken from the data, a similar pattern of changes over time is apparent. This time the no response category declines a little while the inaccurate responses decline more markedly (from 23% to 3%). Again, the generalised responses declined, but less so than in response to Question 1. This time, the data referenced responses more than double. Figure 6.2 shows graphically this increase in the use of data to back up points.
So, the questions around data, one about a specific aspect in the data and the other with respect to the main points to be taken, yielded similar patterns of data at both points in time. There were similar changes in the patterns of response for both questions across time.

The final question asked about what the young colleague should do, given these results. The no response was similar to those for the other questions as was the level of generalised response although the latter did not fall off as much at Time 2 as it did in responses to the other two questions. Perhaps there is a tendency in giving advice to err on the side of generality. The interesting aspect of the responses to this question are that at Time 1, over half of the teachers were able to give advice that was referenced to data, higher than in their replies to either of the other two questions. And, by Time 2, this level of data referenced response had risen to 70%, the highest rate for any category in response to any of the questions. (Figure 6.3 shows these trends graphically).
At Time 3, a year after the Time 2 data collection point, only 33 teachers from cohort 1, who also responded at Time 1 and Time 2, remained (as explained previously, this was due, in the main, to two large schools from cohort 1 taking the one year option). For both questions concerning data interpretation, the gains made at Time 2 were largely maintained. The percentages of no response and inaccurate response continued to fall a little. Reference to data was maintained at the high level of Time 2. However, the level of generalised responses tended to increase but not to the Time 1 level.

Growth in data interpretation knowledge
A total score for each teacher was obtained according to how their responses were categorised. At Time 1 the average score was 3.62 while, at Time 2, this had increased to 4.85. There was an overall significant increase in knowledge of data interpretation and use (t = 5.089, p<.01). The self-report item in the questionnaire regarding confidence to interpret data from tests <of reading or writing> correlated significantly, although modestly, with score on the data interpretation scenario at Time 2 (r = .474, p< .001).

Relationship between data interpretation and use and student progress
However, for both reading and writing, neither the extent of teacher progress nor the level of competence of a teacher with respect to data interpretation at either time point (Time 1 or Time 2) or the amount of gain in teacher’s knowledge of data interpretation was significantly related to the degree of progress of students in that teacher’s class.

Scenario 2: Knowledge of Principles of Evidence-informed Decision Making
A scenario in the form that tapped the principles of evidence-informed decision-making was only presented at one point in time. This was at Time 2 for the North Island schools and at Time 1 for the South Island. The responses for both cohorts have been combined as the relativity patterns were similar for both cohorts and
the sample size for the South Island is small. The first question considered was the extent to which respondents, as a whole, agreed that the dimensions depicted in the scenario represented relatively ineffective evidence-informed practice. The ratings given by respondents for each dimension are summarized in Table 6.1, which indicates the percentage of respondents that allocated each rating for each of the four dimensions.

### Table 6.1

**Percentage of Respondents Giving Each Rating for the Four Dimensions of Evidence-Informed Decision Making**

<table>
<thead>
<tr>
<th>Scenario Question</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>A) How student needs were identified</td>
<td>2.1</td>
</tr>
<tr>
<td>B) The basis for adopting the programme</td>
<td>8.5</td>
</tr>
<tr>
<td>C) Identifying issues at the meeting</td>
<td>4.9</td>
</tr>
<tr>
<td>D) The basis for the decision to continue with the programme</td>
<td>9.3</td>
</tr>
</tbody>
</table>

For the question regarding ‘How student needs were identified’, most school leaders/teachers responded positively with the most common response being ‘moderately effective’. Only a third thought that identifying student needs based simply on teacher observation was ineffective. The mean rating for this question was 4.1, which was above the mid-point and equates to a rating of ‘slightly effective’. Responses for Question B “The basis for adopting the programme” appear reasonably evenly spread from ‘moderately ineffective’ to ‘moderately effective’. More respondents rated this aspect of the scenario negatively than the other three parts. The mean rating for this question was the lowest of the four questions at 3.5. Question C. “Identifying issues at the meeting” is heavily weighted towards positive responses with the most common response being ‘highly effective’ with a mean rating of 4.6. Question D, “The basis for the decision to continue with the programme” has a relatively even spread of responses except for the most common response of ‘moderately effective’. The mean rating for this question was 3.8.

The next analysis considered the extent to which the reasons given for particular ratings were based on evidence-focused attributes of the scenario. It also allowed a consideration of the level of congruity between rating and reason. Ratings were grouped into high, medium and low. Responses to each dimension of the scenario are reported separately. In order for a reason to appear in a table, more than two people needed to have nominated it (Reasons nominated by only one or two people are included in the “other” category in the tables).

The majority of respondents (56.6%) stated that the school’s method of identifying student needs was inadequate (see Table 6.2). This suggests that a high proportion of teachers in our sample do have knowledge of evidence informed practice even though the ratings may not always suggest this - as exemplified by the
small number of respondents (9.6%) who claimed the method was inadequate, despite rating the method as moderately to highly effective. It should be noted also that some teachers (8.8%) responded that the test used to identify the need was appropriate because they thought the running records discussed in Part A of the Reading scenario included a test of comprehension. However, over a fifth of respondents (22.8%) believed that teacher observation alone was effective in identifying student needs.

Table 6.2
Percentage of Low, Medium & High Rating Groups Nominating Reasons for Rating the Dimension: How students' needs were identified.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Low ratings</th>
<th>Medium ratings</th>
<th>High ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate / insufficient information</td>
<td>19.1</td>
<td>27.9</td>
<td>9.6</td>
</tr>
<tr>
<td>Appropriate to identify a need</td>
<td>0</td>
<td>0.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Appropriate process / test</td>
<td>0</td>
<td>2.9</td>
<td>19.9</td>
</tr>
<tr>
<td>Appropriate test (assumed test also included comprehension)</td>
<td>0</td>
<td>0</td>
<td>8.8</td>
</tr>
<tr>
<td>Staff input / communication effective</td>
<td>0</td>
<td>0.7</td>
<td>2.9</td>
</tr>
<tr>
<td>Effective to focus on specific area in depth</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>Other</td>
<td>0.7</td>
<td>2.2</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Teachers were better at identifying a lack of appropriate evidence with the second question concerning the basis for adopting the programme, than for the question relating to how student needs were identified. Table 6.3 shows that around 70% of respondents gave appropriate negative reasons for their rating. Again, around a fifth of teachers (19.2%) believed that the notion of adopting a programme based on the reportedly increased enjoyment of students at another school was an effective basis for decision-making.

Table 6.3
Percentage of Low, Medium & High Rating Groups Nominating Reasons for Rating the Dimension: The basis for adopting the programme.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Low ratings</th>
<th>Medium ratings</th>
<th>High ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not based on best practice / research</td>
<td>11.1</td>
<td>21.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Poor needs / programme match</td>
<td>11.9</td>
<td>8.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Cannot assume that what works for one group will work for others</td>
<td>3.0</td>
<td>8.9</td>
<td>0</td>
</tr>
<tr>
<td>Programme insufficiently individualized</td>
<td>1.5</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>Based on what worked elsewhere</td>
<td>0</td>
<td>2.2</td>
<td>17.0</td>
</tr>
<tr>
<td>Based on perceived benefits for students</td>
<td>0</td>
<td>0.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Appropriate involvement of staff</td>
<td>0</td>
<td>0.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Other</td>
<td>2.2</td>
<td>0.7</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Respondents were positive about the collaborative nature of the meeting as portrayed in the scenario (see Table 6.4). These positive ratings match the perceptions expressed in the interviews following syndicate meeting observations (referred to in Chapter 10 of this report). In the interviews teachers were generally highly satisfied with syndicate meetings that focused on organisational issues with collaborative, supportive
leaders. A non-threatening approach was given higher priority than perhaps a more challenging approach that may have promoted professional learning.

**Table 6.4**

Percentage of Low, Medium & High Rating Groups Nominating Reasons for Rating the Dimension: Identifying issues at the syndicate meeting.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Low ratings</th>
<th>Medium ratings</th>
<th>High ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concerns about focus on organizational issues only</td>
<td>12.2</td>
<td>12.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Not a data-based decision</td>
<td>0.8</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Collaborative, non-threatening approach</td>
<td>0.8</td>
<td>6.9</td>
<td>44.3</td>
</tr>
<tr>
<td>Discussion of issues prioritized / regular time / general positive</td>
<td>0</td>
<td>0.8</td>
<td>12.2</td>
</tr>
<tr>
<td>Other</td>
<td>0.8</td>
<td>3.1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Around two-thirds of teachers gave a negative reason for their rating of the dimension regarding the basis for continuing the programme (Table 6.5). Like Question A, a small number of respondents (3.8%) responded that the test used to identify the need was appropriate because they thought the running records discussed in Part D of the Reading scenario included a test of comprehension. However, a disparity between rating and reason exists again with 12.8% of respondents giving a negative reason with a rating of moderately or highly effective. It is important that caution is exercised in interpreting ratings. It may be that some teachers perceive the situation in terms of the “glass half full” while others perceive the situation in terms of the “glass half empty”. Ratings and reasons need to be combined to obtain an overall picture.

**Table 6.5**

Percentage of Low, Medium & High Rating Groups Nominating Reasons for Rating the Dimension: Basis for decision to continue with programme.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Low ratings</th>
<th>Medium ratings</th>
<th>High ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment not matched to need</td>
<td>23.5</td>
<td>12.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Assessment not adequate (on its own) (no reason given)</td>
<td>2.3</td>
<td>11.4</td>
<td>8.3</td>
</tr>
<tr>
<td>Inadequate depth of analysis</td>
<td>3.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Appropriate assessment for making continuation decisions</td>
<td>0</td>
<td>1.5</td>
<td>17.4</td>
</tr>
<tr>
<td>Appropriate assessment (indication that assumes comprehension assessed)</td>
<td>0</td>
<td>0</td>
<td>3.8</td>
</tr>
<tr>
<td>Students improved</td>
<td>0</td>
<td>0</td>
<td>6.1</td>
</tr>
<tr>
<td>Other</td>
<td>0.8</td>
<td>3.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Discussion

Studies suggest that effective gathering and use of evidence or assessment makes a difference to student achievement. Simply receiving targeted, specific feedback about their students reading ability led to significant shifts in student achievement in schools, even before the commencement of professional development intended to help teachers hone their practice to meet the needs of the students as shown in the data (McNaughton et al., 2004). Informed of the strengths and needs of their students, they were able to make shifts in practice that were associated with increased student achievement. Similarly, teachers who examined student data together and worked out what its implications were for teaching and who collectively took responsibility for deciding how best to help those underachieving, difficult to move students, had higher achieving students than schools where such a collective examination, diagnosis and problem-solving cycle did not operate.

The argument advanced here is that there are several levels on which knowledge is needed in order to use evidence to inform decisions. One is that a teacher has to believe in the utility of evidence and have a sense of the principles that guide evidence-informed decision-making. Our teachers showed only modest recognition of these principles. However, their generous ratings at times mask a likely appreciation of the ineffective nature of the move. Although most teachers in the sample had been involved in examining the data related to their class during the year, it is likely that key concepts related to evidence-based practice had not been generalised beyond the immediate uses to which it was being put in the professional development context. These findings are a timely reminder that generalisation of concepts and skills to a new situation cannot be assumed. Rather generalisation needs to be specifically taught with the skills related to the examination of student data for immediate teaching purposes then applied to the use of evidence to decide on the introduction of new programmes.

On another level, teachers need to have the skill to draw inferences from the data, to understand what it means. The teachers in our study became more adept at this aspect. There is, however, a large element of pedagogical content knowledge required to interpret the data in relation to classroom instructional practice. Clearly, it is difficult for individual teachers or groups of teachers to take the information from a diagnostic test and be able to make teaching decisions from it if they have insufficient pedagogical content knowledge (the knowledge about the subject from the point of view of how to teach it). The diagnostic information that students score poorly on the curriculum function, structure, when writing for one of various purposes for writing (say an argument) is not particularly useful information if the teacher does not know how language works at a text and local level to provide coherence and cohesion. If, for example, you administer asTTle and secure some expert assistance in scoring (as teachers need considerable knowledge of writing to score using the rubrics), and you find students have a gap with regard to structuring an argument, then teachers need specific knowledge of how argument texts operate to work with this information.

Hence, it is not surprising that there is no direct relation between scores on data interpretation and student progress. The use of data at the classroom practice level is likely to be mediated by the teacher’s pedagogical content knowledge. Unfortunately, we had only one item that indirectly tapped this aspect. It asked what the young colleague should do with the information or how useful the information was for teaching and why. And, this item did not require the respondent to be specific, to demonstrate adequate content knowledge of teaching reading or writing. And, data presented elsewhere in this report show that the progress of students
in writing in particular, is significantly related to the level of pedagogical content knowledge their teachers have (Parr & Timperley, 2006). Teachers with higher levels of knowledge had students who made more progress. So, we suggest that using evidence involves knowledge of more than simply understanding what the data show. To apply this knowledge to teaching practice requires considerable knowledge of the subject from the point of view of teaching it. It appears as if this knowledge, as illustrated in both research cases where evidence use was related to improved student outcomes, is best sourced from the combined efforts of the professional learning community within a school.

References


Chapter 7

Teacher Perceptions of Their Role and Learning

In this chapter we present teacher self-report in response to questions about their role and their learning. As these are responses with respect to several different dimensions of teacher perception, the chapter is in sections. The sections deal with (i) perception of influence and satisfaction with student achievement; (ii) with teacher confidence with respect to aspects of practice that centrally require pedagogical content knowledge and, (iii) finally their judgment about aspects of the professional development project in relation to their learning and support for this learning. Relevant literature is discussed where appropriate within each section and details of method also presented there.

Perception of Influence on and Satisfaction with Student Achievement

There has always been a debate in the literature about the potential influence of education on the child. Views in the research community have moved from the somewhat deterministic stance of Coleman and colleagues (1966) who maintained that, given genetic and home factors, schools had a small amount of the variance in student achievement that they could affect. Now, there is a consensus internationally that teachers (and the quality of their teaching) have, relatively, the greatest potential effect on student achievement. For example Alton-Lee (2003) concludes that teachers can account for up to 58 percent of the variance in student achievement. Clearly, however, if teachers do not believe that they can make a tangible difference or that this difference is small, then their level of effort or the extent to which they feel responsible for, and satisfied with, student achievement may be affected. We were interested in teacher perception of how much they were both influential in student achievement and satisfied with that achievement and whether this perception changed as a result of participation in the professional development project. Previous work (Timperley, 2003) has shown that, as teachers were able to effect positive change in student achievement, their attitudes as to their potential influence changed.

Method

Teachers, literacy leaders and principals were asked to judge the relative contribution of major influences on a child’s achievement, basically that of the child him/herself, the home, the teacher and the school. They did this by allocating each a portion of a total 100 percent. Both cohorts, North and South Islands, have data for two points in time. As the patterns for these cohorts were similar at both points in time and the South Island was a small sample size, they are combined for purposes of analysis (N = 107). There were data from only 27 teachers from cohort 1 available at the end of the two year period.

Teachers were also asked to rate their level of satisfaction with student achievement (on a 6 point Likert scale from 1 disagree to 6 agree). There were responses available from 110 teachers.

Findings and Discussion

Over the year period of the professional development, the amount of influence that teachers (N = 107) reported that they thought they had significantly increased (t = -2.50, p < .05). As Figure 7.1 shows, the
average percentages allocated to both home and child each fell a few points; that for the school stayed constant, while the percentage of influence attributed to the teacher increased. However, it is still averaging around 40 percent, considerably below the figure suggested in the research literature.

**Figure 7.1**
Relative Proportions Allocated to Influences on Achievement at Time 1 & 2

The Time 3 data show a continuation of this trend. The 27 teachers for whom we have data at three points in time had a higher average percentage for teacher influence at Time 1, namely, 39% and this had reached 46% at Time 3. For this group, the amount attributed to home influence fell more markedly than that attributed to the child.

The numbers of respondents are small at the school level but the different patterns by school are interesting. Three different patterns are shown in the figures below. Figure 7.2 shows School X that rates the child initially as the most influential and the teacher slightly more influential than the home. By Time 2, teacher influence has gained markedly and home and child influence dropped. Figure 7.3 (School Y) shows a pattern where, initially, the home was viewed as more influential than the teacher. At Time 2 this reverses, while the child influence remains constant. Figure 7.4 shows School Z where teachers continue to believe that home is more influential than they are (teacher and school combined).
Figure 7.2
Relative Proportions Allocated to Influences on Achievement at Times 1& 2: School X

Figure 7.3
Relative Proportions Allocated to Influences on Achievement at Times 1& 2: School Y
With respect to the question about how satisfied teachers were with the achievement of their students, the mean rating increased over time. For the 110 teachers with responses to this question at both Time 1 and Time 2, the mean rating significantly increased ($t = -4.98$, $p < .000$) from 4.10 to 4.60. Satisfaction with achievement rose in all schools bar one where it remained constant (interestingly, achievement in that school remained constant!). The lowest level of satisfaction at Time 2 was reported from the school with the lowest average level of achievement, while the highest level of satisfaction was reported by the school with the high levels of achievement and of gains.

### Self-reported Confidence re Dimensions of Pedagogical Content Knowledge

#### Method
First, teachers rated their level of confidence, using a six-point Likert scale (from 1 definitely not confident to 6 highly confident), in response to items concerning practice where a high level of pedagogical content knowledge was involved. There were 10 such items at Time 1 and seven at Time 2 with a core of six items common at both points in time. These common items asked them to rate, for example, their ability to construct appropriate learning objectives for lessons in reading comprehension or writing (objectives); to make the criteria for successful learning clear to students (criteria); to develop activities to support learning objectives (activities); to teach strategies that support reading comprehension or the features of text associated with different genre (features); to assess reading comprehension or writing (assess), and to give students feedback in ways that support learning (feedback). Those items at Time 1, not asked at Time 2, were about teachers’ ability to cater for a range of needs (range); their knowledge of whether progressing well for age (progress); knowledge of the characteristics of readers or writers at different points in their development (characteristics), and a more specific item around teaching inferential comprehension in reading (inferential) or structure in writing (structure). The additional item at Time 2 asked about confidence in ability to interpret and use achievement data (interpret).
For the self reported confidence item ratings, the items were first analysed in terms of their inter-relatedness. Descriptive statistics were obtained, then comparisons made of means at a single point in time, then over time.

**Findings and Discussion**

**Writing**

At Time 1 the items were significantly inter-correlated ($p < .001$), suggesting that these items are likely to represent aspects of an underlying construct, namely, pedagogical content knowledge in terms of writing. Correlations ranged from .42 to .79.

The mean ratings shown in Table 7.1 suggest, however, that they were not equally confident in their knowledge base enabling them to do certain things. The means of the item ratings range from 3.61 to 4.14 (where 1 is ‘definitely lack confidence’ and 6 represents ‘highly confident’).

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress</td>
<td>4.14</td>
<td>1.17</td>
</tr>
<tr>
<td>Feedback</td>
<td>4.09</td>
<td>0.97</td>
</tr>
<tr>
<td>Objectives</td>
<td>3.95</td>
<td>0.97</td>
</tr>
<tr>
<td>Criteria</td>
<td>3.91</td>
<td>0.95</td>
</tr>
<tr>
<td>Activities</td>
<td>3.93</td>
<td>1.03</td>
</tr>
<tr>
<td>Range</td>
<td>3.61</td>
<td>0.95</td>
</tr>
<tr>
<td>Features</td>
<td>4.01</td>
<td>1.17</td>
</tr>
<tr>
<td>Structure</td>
<td>4.04</td>
<td>0.98</td>
</tr>
<tr>
<td>Characteristics</td>
<td>3.66</td>
<td>0.97</td>
</tr>
<tr>
<td>Assess</td>
<td>3.93</td>
<td>1.02</td>
</tr>
</tbody>
</table>

A multivariate analysis of variance found that there was a significant difference amongst the mean ratings of the items ($F(9,31)=3.35$, $p = .006$). Post-hoc tests were performed to see where this difference lies. Given multiple tests, a Bonferroni correction was applied whereby $p < .001$. The indication was that teachers were significantly less confident about catering for the range of needs in their class than they were confident about knowing whether students were progressing well for their age or giving feedback to students.

By Time 2, it is arguable that teachers now knew more about what they needed to know in order to teach either reading comprehension or writing more effectively to their students. They may have been judging their confidence to do certain things from a different vantage point. Self-report confidence ratings are problematic to interpret because the relationship between increased knowledge or confidence and increased rating is not necessarily linear. The old adages, ‘you don’t know what you don’t know’ or ‘when you know, you realise what you didn’t know’ may have currency.
At Time 2 the relationship between the seven items certainly appears less straightforward. There seems no longer to be a single factor or construct. However, there are some logical groupings. Objectives, Activities, Criteria and Feedback appear to have been moderately inter-correlated. However, no significant correlation existed between Activities and Criteria, which might be expected if activities are constructed to meet learning objectives with their associated success criteria. Assess and Interpret were significantly correlated together, as one might expect but had very weak correlations with the other items. The Features item has weak correlations with all items although a significant correlation exists with Feedback. This makes perfect sense in writing, as one has to know the features of writing associated with particular purposes in order to provide feedback.

At Time 2 teachers, on average, reported levels of confidence ranging from 4.65 for their ability to interpret and use data to 5.15 for their confidence in setting learning objectives (see Table 7.2). This is a similar range to Time 1. A multivariate analysis of variance found that there was no significant difference amongst the mean ratings of the items (p>.05). They did not report significantly different levels of confidence by Time 2.

### Table 7.2
Mean Rating and Standard Deviation for the Seven Items at Time 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>5.15</td>
<td>0.70</td>
</tr>
<tr>
<td>Activities</td>
<td>4.89</td>
<td>0.61</td>
</tr>
<tr>
<td>Criteria</td>
<td>5.11</td>
<td>0.65</td>
</tr>
<tr>
<td>Assess</td>
<td>4.69</td>
<td>1.00</td>
</tr>
<tr>
<td>Feedback</td>
<td>4.84</td>
<td>0.67</td>
</tr>
<tr>
<td>Features</td>
<td>4.97</td>
<td>0.69</td>
</tr>
<tr>
<td>Interpret</td>
<td>4.65</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Analyses were undertaken to view Time 1 and 2 ratings in relation to one another. The six core items that were the same at Time 1 and Time 2 were significantly correlated except for Criteria (see Table 7.3). This suggests that there was a more than chance relationship in terms of the ratings that individuals gave to an item at each point in time. However, the significant correlations are moderate not high.

### Table 7.3
Correlation between Time 1 and Time 2 for parallel Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>.359(*)</td>
</tr>
<tr>
<td>Criteria</td>
<td>.037</td>
</tr>
<tr>
<td>Assess</td>
<td>.630(**)</td>
</tr>
<tr>
<td>Feedback</td>
<td>.537(**)</td>
</tr>
<tr>
<td>Activities</td>
<td>.435(*)</td>
</tr>
<tr>
<td>Features</td>
<td>.547(**)</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
A further analysis considered the mean ratings given to items and whether these varied over time. It is clear that mean ratings were higher at Time 2 than Time 1, which suggests that the opportunity to learn more about teaching and assessing writing, in this case, was associated with greater confidence. A repeated measures analysis of variance (ANOVA) showed that rating significantly increased over time (F(6,28)=11.45, p<.001). All items received significantly higher ratings.

Reading

As for writing, the correlation matrix for the reading items indicates, at Time 1, the items were significantly inter-correlated (p < .01), suggesting that these items are tapping related dimensions of a construct. Correlations ranged from .35 to .71.

The mean ratings shown in Table 7.4 suggest, however, that they were not equally confident in their knowledge base regarding aspects of teaching reading comprehension. The means of the item ratings range from 3.36 to 4.33. The highest ratings were given to items concerning knowing whether students were performing well for their age and giving feedback to students in order to progress learning while the lowest ratings were given to an item asking about confidence in assessing reading strategies.

Table 7.4
Mean Rating and Standard Deviation for the 10 Items at Time 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress</td>
<td>4.33</td>
<td>1.08</td>
</tr>
<tr>
<td>Feedback</td>
<td>4.14</td>
<td>0.89</td>
</tr>
<tr>
<td>Objectives</td>
<td>3.88</td>
<td>0.97</td>
</tr>
<tr>
<td>Criteria</td>
<td>3.82</td>
<td>0.86</td>
</tr>
<tr>
<td>Activities</td>
<td>3.99</td>
<td>0.87</td>
</tr>
<tr>
<td>Range</td>
<td>3.86</td>
<td>0.95</td>
</tr>
<tr>
<td>Strategies</td>
<td>3.73</td>
<td>0.98</td>
</tr>
<tr>
<td>Inferential</td>
<td>3.71</td>
<td>1.03</td>
</tr>
<tr>
<td>Characteristics</td>
<td>3.52</td>
<td>1.07</td>
</tr>
<tr>
<td>Assess</td>
<td>3.36</td>
<td>1.13</td>
</tr>
</tbody>
</table>

A multivariate analysis of variance found that there was a significant difference amongst the mean ratings of the items (F(9,64)=7.69, p<.001). Post-hoc tests were performed to see where this difference lies. Given multiple tests, a Bonferroni correction was applied whereby p < .001. Teachers felt more confident about knowing whether students were progressing well for their age than for all other items, save feedback and activities where there was no significant difference in the rating. Similarly, they felt more confident about giving feedback to students in order to help them learn than for other items (except developing activities, setting objectives and catering for the range of needs in their classroom for which there was no difference in confidence rating).

At Time 2, unlike the results for the writing items, there were still significant inter-correlations for all items. Teachers may see knowing about reading in order to teach it as a more homogenous entity than knowing
about writing in order to teach it. Correlations ranged from .420 to .763. When a larger sample size is available, factor analyses may be undertaken to investigate possible factors or groupings of items.

At Time 2 teachers, on average, reported levels of confidence ranging from 4.68 for feedback to 4.79 for arranging activities (see Table 7.5). The range has narrowed since Time 1 and the standard deviations decreased. A multivariate analysis of variance found that there was no significant difference amongst the mean ratings of the items (p>.05). They did not report significantly different levels of confidence about what they knew about teaching aspects of reading.

### Table 7.5
**Mean Rating and Standard Deviation for the Seven Items at Time 2**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>4.75</td>
<td>0.94</td>
</tr>
<tr>
<td>Activities</td>
<td>4.79</td>
<td>0.87</td>
</tr>
<tr>
<td>Criteria</td>
<td>4.68</td>
<td>0.99</td>
</tr>
<tr>
<td>Assess</td>
<td>4.72</td>
<td>0.96</td>
</tr>
<tr>
<td>Feedback</td>
<td>4.71</td>
<td>0.89</td>
</tr>
<tr>
<td>Teach</td>
<td>4.75</td>
<td>0.84</td>
</tr>
<tr>
<td>Interpret</td>
<td>4.40</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Analyses were undertaken to view Time 1 and 2 ratings in relation to one another. The six core items that were the same at Time 1 and Time 2 were significantly correlated except for Criteria (see Table 7.6). This suggests that there was a more than chance relationship in terms of the ratings that individuals gave to an item at each point in time. However, as in writing, the correlations are moderate, not high and, for one item, how to make criteria for successful writing clear to students, there is no relationship.

### Table 7.6
**Correlations between Time 1 and Time 2 for similar Items**

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>.409(**)</td>
</tr>
<tr>
<td>Criteria</td>
<td>.233</td>
</tr>
<tr>
<td>Assess</td>
<td>.542(**)</td>
</tr>
<tr>
<td>Feedback</td>
<td>.359(**)</td>
</tr>
<tr>
<td>Activities</td>
<td>.332(*)</td>
</tr>
<tr>
<td>Strategies</td>
<td>.392(**)</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

A further analysis considered the mean ratings given to items and whether these varied over time. It is clear that mean ratings were higher at Time 2 than Time 1, which suggests that the opportunity to learn more about teaching and assessing reading, in this case, was associated with greater confidence. A repeated
measures analysis of variance (ANOVA) showed that rating significantly increased over time (F(6,51)=20.08, p<.001). All items received significantly higher ratings.

An analysis was undertaken to explore the relationship between a teacher’s mean confidence score at Time 2 and the effect size gain for student achievement. There was no significant relationship for reading comprehension teachers or for writing teachers.

**Conclusion**

The overall conclusion from the self reported confidence in aspects of pedagogical content knowledge is that teachers significantly increased in their confidence that they had the knowledge and ability to: construct appropriate learning objectives for lessons in reading comprehension or writing (objectives); to make the criteria for successful learning clear to students (criteria); to develop activities to support learning objectives (activities); to teach strategies that support reading comprehension or the features of text associated with different genre (features); to assess reading comprehension or writing (assess), and to give students feedback in ways that support learning (feedback). Not only did teachers rate themselves as more confident but the range in ratings decreased over time. However, the additional item at Time 2, concerning analysing and using data received the lowest rating from teachers at both reading and writing schools, suggesting that there is still professional learning that needs to happen in this area.

**Teachers’ Attitudes towards Professional Development after One Year**

Motivation is always important in professional learning and development situations. This is particularly so when the professional development takes place over a longer period of time as is the case in this literacy project. Teachers are more likely to feel motivated if they are improving their skills, feel they are being supported in their new learning and do not feel overwhelmed.

Therefore, in March 2005, literacy leaders and teachers in the Northland / Auckland schools who were at the school in 2004 were asked to fill in a 10 item questionnaire that sought to determine reactions and motivation to the project so far. A similar questionnaire was administered to such teachers in the South Island schools after a year on the project. The items were selected to tap into a range of possible responses and included both positively-oriented and negatively-oriented items. The items and the order in which they were presented follows:

- I have learnt how to make my literacy teaching practice more effective;
- I am more confused about how to teach literacy;
- I feel the team I usually work in is more focused on improving student learning;
- I feel support is there if I need it;
- My time has been wasted;
- My use of information from student assessments is more effective;
- I am expecting to learn more about teaching literacy this year;
- I feel I am a more knowledgeable teacher about literacy;
- I feel overwhelmed by what I am being asked to do;
- I am concerned that when the project finishes the changes we have made will not be sustained.
A six-point rating scale was used, with a rating of 1 representing “Strongly disagree”, 2 representing “Mostly disagree”, 3 “Slightly disagree”, 4 “Slightly agree”, 5 “Mostly agree” and 6 representing “Strongly agree”.

Analysis of Responses
This initial analysis sought to ascertain the order in which the items were ranked and the average rating given to each item. In order to determine the ranked position of the negative items, a reverse rating was used. For example, if the original rating was 2, it was given a reverse rating of 5. The wording of the item has been changed to reflect the positive orientation of the reverse rating. For example, “My time has been wasted”, has been reworded to “My time has not been wasted”.

We then compared the pattern of responses evident in the different schools and sought to determine whether the pattern of responses was related to the relative gains in student achievement over the previous year.

Aggregated responses for schools
Overall, the responses in all schools were very positive. Positively oriented items received high ratings and negatively oriented items received low ratings. A ranking of items and the overall average and range are presented in Table 7.7 with the reverse rating of negative items noted in the right hand column.

Table 7.7  
Average Rating and Range of Ratings of Survey Items (All Schools)

<table>
<thead>
<tr>
<th>Item</th>
<th>Average rating (including reverse rating of negative items)</th>
<th>Range</th>
<th>Original rating of negative items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time has not been wasted</td>
<td>5.43</td>
<td>2-6</td>
<td>1.57</td>
</tr>
<tr>
<td>Expecting to learn more</td>
<td>5.17</td>
<td>2-6</td>
<td></td>
</tr>
<tr>
<td>Less confused about how to teach</td>
<td>5.14</td>
<td>1-6</td>
<td>1.86</td>
</tr>
<tr>
<td>Teaching practice more effective</td>
<td>5.15</td>
<td>4-6</td>
<td></td>
</tr>
<tr>
<td>Feel I am a more knowledgeable</td>
<td>5.11</td>
<td>3-6</td>
<td></td>
</tr>
<tr>
<td>Use of assessment information</td>
<td>4.88</td>
<td>2-6</td>
<td></td>
</tr>
<tr>
<td>Team more focused</td>
<td>4.86</td>
<td>1-6</td>
<td></td>
</tr>
<tr>
<td>Support is there if I need it</td>
<td>4.92</td>
<td>2-6</td>
<td></td>
</tr>
<tr>
<td>Lack of concern about sustainability</td>
<td>4.43</td>
<td>2-6</td>
<td>2.57</td>
</tr>
<tr>
<td>Do not feel overwhelmed</td>
<td>4.29</td>
<td>1-6</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Responses for individual schools
The next analysis considers the pattern of ratings for individual schools and the extent to which this pattern is associated with changes in student achievement over the previous year. In Table 7.8, items are presented in the same order as in the previous table with an average rating for all items also calculated. Reverse scores for negative items have been used in this table with wording changed to reflect a positive orientation.
Table 7.8
Average Rating of Survey Items for Individual schools

<table>
<thead>
<tr>
<th>Item</th>
<th>School</th>
<th>School</th>
<th>School</th>
<th>School</th>
<th>School</th>
<th>School</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Item</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Time has not been wasted</td>
<td></td>
<td>5.67</td>
<td>4.50</td>
<td>5.75</td>
<td>5.75</td>
<td>6.00</td>
<td>5.70</td>
</tr>
<tr>
<td>Expecting to learn more</td>
<td></td>
<td>4.83</td>
<td>4.33</td>
<td>5.25</td>
<td>5.50</td>
<td>4.75</td>
<td>5.40</td>
</tr>
<tr>
<td>Less confused about how to teach</td>
<td></td>
<td>5.58</td>
<td>3.83</td>
<td>5.50</td>
<td>4.50</td>
<td>5.75</td>
<td>5.60</td>
</tr>
<tr>
<td>Teaching practice more effective</td>
<td></td>
<td>5.42</td>
<td>4.33</td>
<td>4.75</td>
<td>5.50</td>
<td>5.50</td>
<td>5.40</td>
</tr>
<tr>
<td>Feel I am a more knowledgeable</td>
<td></td>
<td>5.50</td>
<td>4.33</td>
<td>4.75</td>
<td>5.00</td>
<td>5.75</td>
<td>5.40</td>
</tr>
<tr>
<td>Use of assessment information</td>
<td></td>
<td>5.25</td>
<td>4.00</td>
<td>4.25</td>
<td>4.50</td>
<td>5.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Team more focused</td>
<td></td>
<td>5.08</td>
<td>4.00</td>
<td>4.00</td>
<td>5.25</td>
<td>5.50</td>
<td>5.50</td>
</tr>
<tr>
<td>Support is there if I need it</td>
<td></td>
<td>4.75</td>
<td>4.67</td>
<td>5.50</td>
<td>5.50</td>
<td>5.00</td>
<td>5.20</td>
</tr>
<tr>
<td>Lack of concern about sustainability</td>
<td></td>
<td>4.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.10</td>
</tr>
<tr>
<td>Do not feel overwhelmed</td>
<td></td>
<td>4.58</td>
<td>3.17</td>
<td>5.00</td>
<td>4.75</td>
<td>5.25</td>
<td>4.60</td>
</tr>
<tr>
<td>Average rating for all items</td>
<td></td>
<td>5.08</td>
<td>4.15</td>
<td>5.05</td>
<td>5.05</td>
<td>5.28</td>
<td>5.29</td>
</tr>
<tr>
<td>Effect size in achievement gains</td>
<td></td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

The average ratings for all items did not appear to be related to relative gains in student achievement over the previous year. The highest average rating was given by teachers where no gains in student achievement were made over the year with the two schools with average gains giving the lowest ratings. A Spearman rank order correlation between average rating across all items and the effect size in achievement gains for each school was not significant.

Comment
Facilitators and school leaders have reason to be very pleased that the atmosphere created in these research schools is one in which literacy leaders and teachers are giving very positive ratings to the time spent, the amount learned, and their anticipation of learning more.

Conclusion
This chapter has examined teacher self report of attitudes and learning, largely over a year of professional development. The data presented suggest that teachers have changed their attitudes with respect to the amount of influence they perceive they have on students’ learning. They are more satisfied now with their students’ achievement in reading comprehension or writing. They report higher levels of confidence on all items intended to tap pedagogical content knowledge. After a year on the project, positive ratings were given to the time spent, the amount learned, and practitioners’ anticipation of learning more.

References

PART C: Facilitating Teacher Learning

The Literacy Professional Development Project has a number of distinguishing features. One of these features is the strong outcomes focus. The contracted outcomes, as outlined in the introduction to this report, included:

- Student achievement
- Development of teachers’ pedagogical content knowledge
- Transfer of understanding of literacy pedagogy to practice
- Professional learning communities

A second distinguishing feature relates to process. A needs analysis approach was taken that involved developing an understanding of teachers’ learning needs through classroom observations and their responses to a scenario designed to determine the depth of teachers’ pedagogical content knowledge. An overview of the schools’ use of data was obtained through interviews with the principal and literacy leader. A major challenge for facilitators was to combine the contractual requirements and the needs analysis approach into meaningful learning opportunities for teachers and school leaders. Not surprisingly, integration of these features was a new and major challenge for most facilitators. The next four chapters examine different aspects of this challenge.

- Chapter 8 outlines a case study illustrating how the four outcomes and needs analysis approach can be integrated in ways that promote teacher learning
- Chapter 9 identifies the associated facilitation challenges from the facilitators’ perspective.
- Chapter 10 examines how teacher learning can be within professional learning communities by examining the evolution of teacher meetings over the two years.
- Chapter 11 identifies ways to promoting teacher learning through structured episodes
Chapter 8
A Case Study: Promoting Teacher Learning Using a Needs-Analysis and Inquiry-based Approach

Note: This case has been written in conjunction with Cherry Bertanees, Facilitator, Literacy Professional Development Project.

This case study describes how one facilitator combined the challenges of meeting the four project outcomes with a needs analysis approach to professional learning. This case description is not intended to imply that the processes described are the only or the most effective ones. Rather, it is intended to illustrate how the outcomes and processes might be integrated effectively. The case analysis is based on a set of principles that have been identified in the professional development research literature as effective in promoting teacher learning. The following eight principles provide an overview of this literature.

Principle 1: Goals and student performance. Professional development should be driven by analysis of the differences between goals and standards for student learning and student performance (Fullan, 1993; Howey & Collinson, 1995; Pink & Hyde, 1992). Such analyses will define what teaching professionals need, rather than want to learn.

Principle 2: Teacher involvement. Professional development should involve teaching professionals in the identification of what they need to learn and, when possible, in the development of the learning opportunity and the process to be used (Borko & Putnam, 1995; Newmann & Wehlage, 1995). This engagement increases teaching professionals’ motivation and commitment to learn (Hodges, 1997); affirms their strengths and enhances their sense of efficacy; empowers them to take instructional risks and assume new roles and responsibilities (Pink, 1992); increases the likelihood that what is learned will be meaningful and relevant to particular contexts and problems (Pink & Hyde, 1992). However, simply asking teachers what it is they need to learn is not useful because they are rarely able to identify these accurately (Jones & Hayes, 1980). While they have no difficulty identifying problems, dilemmas, concerns and wants, these tend to describe symptoms of needs that must be diagnosed more thoroughly and interpreted more broadly.

Principle 3: School based. Professional development should be primarily school based and integral to school operations (Feiman-Nemser, 1983; Grossman, 1992; Guskey, 1995a; Joyce & Showers, 1995; Little, 1993; Little & McLaughlin, 1993; Louis, Marks, & Kruse, 1996; Smylie, 1995). More recent work, however, has highlighted many of the pitfalls of basing professional development in schools. These pitfalls need to be taken into account by facilitators. A recent review by Corcoran, Fuhrman, and Belcher (2001), for example, found that when decisions about professional development were primarily school-based, “school staff members paid lip service to the use of research” and “were more interested in designs that drew on research about practices that they already felt were ‘good’ than in designs that were producing results”. (p.81) Other researchers have highlighted the potential for negative micropolitics to develop within schools that may impact adversely on the progress of the professional development.
Principle 4: Collaborative problem solving. Professional development should provide learning opportunities that relate to individual needs but for the most part are organized around collaborative problem solving (Fullan, 1991; Guskey, 1995a; Hargreaves, 1994; Huberman, 1995). This principle is based on the assumption that although learning needs are individual they are best met through focused and supportive interactions with others.

Principle 5: Continuous and supported. Professional development should be continuous and ongoing, involving follow-up and support for further learning. This support should include sources external to the school that can provide necessary resources and an outside perspective (Fullan, 1993; Guskey, 1995a; Hodges, 1996; Pink & Hyde, 1992).

Principle 6: Information rich. Professional development should incorporate evaluation of multiple sources of information on outcomes for students and an understanding of the teaching processes that are involved in implementing the lessons learned through professional development (Guskey, 1995a; Joyce & Showers, 1995; Little, 1993; Tillema & Imants, 1995).

Principle 7: Theoretical understanding. Professional development should provide opportunities to engage in developing a theoretical understanding of the knowledge and skills to be learned (Eraut, 1995; Feiman-Nemser & Parker, 1992; Fullan, 1991; Joyce & Showers, 1995; McDiarmid, 1994; Tillema & Imants, 1995).

Principle 8: Part of a comprehensive change process. Professional development should be integrated with a comprehensive change process within the school that deals with impediments to, and facilitators of, student learning (Guskey, 1995a; Little, 1993; Smylie, 1995).

In summary, this body of research has identified that effective professional development engages professionals in coherent, sustained, and reflective professional learning. Professional learning should focus on individual, collegial, and organizational improvement and promote continuous inquiry into, and improvement of, teaching practice and student learning.

The School

The decile 5 school of 96 students had a staffing profile that included the principal who taught part time, the principals’ release teacher and three other full time teachers. Two of these teachers were nominated as literacy leaders. A Resource Teacher of Learning and Behaviour and part-time reading recovery and ORS-funded teacher, made up the staffing allocation of 6.4 teachers. It is located in a small rural New Zealand town. The students’ ethnicities comprised 79% NZ European, 14% Maori with a small percentage of Samoan, Asian and Indian students. A new principal arrived one term prior to the start of the project after a series of principal appointments. During this term, staff had engaged in a number of professional development activities, but did not feel that they were making much headway in terms of impacting on student learning. The staff decided that writing should be the focus for this project because an analysis of the students’ writing, using writing Exemplars, indicated low levels of achievement for many students.
Data Collection for Case Study

Data were collected through the needs analysis tools used in all research schools (see Appendix C), together with a follow-up visit from a member of the research team four months after the school began its involvement with the contract. On both visits, all data were collected by both the researcher and facilitator. These initial data collection activities included interviews with the principal and two literacy leaders (Years 2/3 and 4/5 teachers) and classroom observations of the literacy leaders and the teacher who acted as the principal’s release teacher. During the observations between six and nine students were interviewed in groups of three from each class. All observations and interviews were audiotaped and transcribed. Consent forms were signed by the principal, all teachers and the interviewed students.

All staff completed the questionnaire and scenario (Appendix A) after school on the same day and the students completed an asTTle writing sample two weeks later. Details of the research data collected are provided in the relevant section of this case report. After the initial data collection, the facilitator kept a diary of the times she visited the school and the principal provided a list of associated school-based activities. Four months later, the same member of the research team revisited the school with the facilitator and interviewed the principal and two literacy leaders. It was decided not to involve the teacher who acted as the principal’s release teacher because she was leaving the school. The interview questions focused on the impact of the needs analysis process. The researcher and facilitator undertook a second observation of the Year 2/3 and 4/5 teachers’ writing lessons and re-interviewed their students. Following these observations, students were given a similar asTTle writing task. Both sets of asTTle scripts were marked by members of the research team who had no other involvement with this part of the research.

As the analysis of each set of data became available it was processed with the staff. Classroom observations were conducted on the first visit, so the implications were discussed and analysed with staff on the next visit to the school by the facilitator. When the results from the scenario became available, they were used to confirm the initial analysis from the classroom observations. Finally, the asTTle data on students’ writing were analysed to allow a more in-depth look at what needed to be taught. A timetable of visits, the main facilitator activities and who was involved are provided in Appendix C.

Classroom observations

Initially three classroom observations from Years 2 – 6 were undertaken of a 45 minute writing lesson. Prior to the lesson the teacher was asked to complete a form that asked for the lesson aims and how this lesson fitted in planned units. During the observation the teachers’ interactions with the students were audiotaped for 45 minutes and an average of six students were interviewed by the researcher during the lesson. Typically this occurred when the students were writing independently at their desks following the lesson introduction. The students were selected through teacher nomination to represent the range of writing abilities in the class. Student interview questions and the rationale for each included, “What are you supposed to be doing?” (as an introduction to get students to talk), “What are you supposed to be learning about writing while you are doing this?” (to find out if they knew the learning aims), “What does a good [genre named] look like?” (to find out if they understood what success looked like) and “What does your teacher tell you to work on in your writing?” (to find out their understanding of teacher feedback).
All teachers assigned the same lesson topic for the children to write about that they believed would be of interest to the students. They also chose a similar structure for the lessons but had different aims. These aims, as written on the pre-observation sheets, are listed in Table 8.1.

### Table 8.1 Lesson Aims

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 5/6</td>
<td>Writing with the audience in mind from English plan. Will choose one to use for presentation for speeches later in the term.</td>
</tr>
<tr>
<td>Year 4/5</td>
<td>Using English Matrices from N.Z. Curriculum Exemplars. The ‘Audience Purpose’ (impact and voice) at Levels 1iii, 2 and 3.</td>
</tr>
<tr>
<td>Year 2/3</td>
<td>To help the children to start their stories using an interesting beginning – to write using detail – Personal experience</td>
</tr>
</tbody>
</table>

Teachers had carefully prepared the observed lessons with the lesson formats following a structure presented at a recent professional development seminar. In summary, the topic was introduced through class discussion of students’ prior experience (of the topic) and the teachers recounted their own personal experience of a similar situation. The teacher then read a short story relevant to the topic, invited students to share their experiences and followed this activity with a paired discussion. The Year 2/3 teacher also modelled the first sentence of the story from her own experience. A brief writing time of between 5 and 15 minutes was followed by sharing of some students’ writing examples with the class for a few minutes at the end of the lesson. An analysis of the audiotape showed the following features:

- The learning aims of the lessons were not shared with the students;
- No explicit criteria were given to the students against which they could monitor the quality of their work;
- The implicit criteria for success (established through teachers’ presentation of the task, teacher modelling, and feedback provided) included a wide range of criteria which rarely related directly to the stated lesson aims;
- Most feedback was in the form of non-specific praise;
- Most of the lesson was spent motivating students to write a “story” with little time spent actually writing.

An analysis of the written feedback indicated that all feedback was focused on mechanics. Student interviews established that their understanding of what they were to learn and their perceptions of the feedback they received focused on surface features.

[Note: This analysis was not available to the facilitator at the time of giving feedback to the teachers.]

Feedback from the teacher-related data: The theory

Research on student learning has consistently shown that effective feedback can be a powerful way of promoting learning (Hattie, 1999). Given the similarity of adult and student learning processes (Donovan, Bransford, & Pellegrino, 1999), there is every reason to believe that feedback has the same potential to promote professional learning. However, there is also considerable evidence that under particular circumstances, feedback can be ineffective or even have a negative impact (Deci, Koestner, and Ryan, 1999).
The rationale underpinning the feedback process in this case was Robinson’s (1993) problem based methodology. The main tenants in relation to this case included:

1. Those giving and receiving feedback need to have a shared understanding of the data related to the practice that is the subject of the feedback so that they are in a position to discuss and contest the interpretation and implications each draws from that data. This means that teachers should be able to recognise facilitator descriptions of their practices and be in a position to contest their accuracy.

2. Personal theories (including beliefs and the practices that arise from them) determine practice. Therefore, achieving changes in practice often requires changes in the underpinning theories. These theories need to be the teachers’ “theories-in-use” and understood in the context of actual practice and observations of it. This is because articulating personal theories that are disconnected from practice leads to the identification of “espoused” theories, which are more likely to be driven by what is believed to be desirable rather than by the reality of practice.

3. Any recommended changes to practice need to be based on a shared understanding of these theories-in-use so that their implications for any change in practice can be negotiated. Attempts to introduce new practice are strongly influenced and distorted by existing theories and in the absence of such theory engagement change is likely to be interpreted in terms of existing theories and lead to superficial rather than fundamental change that is unlikely to be sustained (Darling-Hammond, Bransford, & LePage, 2005).

4. Facilitator’s judgements about teachers’ theories, practices and consequences need to be supported explicitly by relevant evidence, particularly when they are likely to be contentious.

The theory in practice

As no analysed data from the needs analysis were available by the time of the facilitator’s second visit, she decided to give the teachers feedback on the observations of the three Year 2 – 6 classes. The feedback process began by summarising the key teaching practices observed during the previous visit on a large newsprint sheet on the refrigerator in the staffroom. Teachers’ practices and the beliefs underpinning them were presented diagrammatically. Given that all the observed lessons were fairly similar because they were based on a lesson presented in an earlier professional development seminar, it was possible to include all practices within the same summary that is provided in Figure 8.1. At this time the “Beliefs” box was left blank.

**Figure 8.1**

**Describing Teachers’ Practices**

<table>
<thead>
<tr>
<th>Beliefs</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Teachers’ practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Children shared experiences</td>
</tr>
<tr>
<td>- Teacher read story for motivation</td>
</tr>
<tr>
<td>- Children wrote for 5-15 minutes</td>
</tr>
<tr>
<td>- Teacher responded positively [to children’s efforts]</td>
</tr>
</tbody>
</table>
The facilitator then asked the teachers to identify their beliefs (personal theories) that led them to use these practices rather than other possible alternatives. The temptation in this type of situation is for facilitators to infer teachers’ beliefs that underpin their practices, then develop a change strategy without making either the beliefs or the inferences explicit. The danger of such an approach is that teachers’ personal theories are bypassed and the focus is shifted to the facilitator’s preferred recommended practice with the implications frequently misunderstood because teachers interpret these recommendations from the viewpoint of their existing theories. Change, if it occurs at all, becomes superficial rather than deeply understood. In this case, the facilitator assisted the teachers to express their beliefs that underpinned their practice and recorded them in the top box of the diagram. The evolving diagram is illustrated in Figure 8.2.

**Figure 8.2**

**Teachers’ Identified Beliefs**

<table>
<thead>
<tr>
<th>Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Children need to be inspired / motivated</td>
</tr>
<tr>
<td>• Children need to share experiences to drive their writing and stimulate others</td>
</tr>
<tr>
<td>• Sharing writing gives other children ideas for their writing</td>
</tr>
<tr>
<td>• Sharing writing gives children an audience.</td>
</tr>
</tbody>
</table>

The beliefs were written above the practices on the chart to acknowledge that practices are based on them, rather than the other way around.

**Levers for sustainable change**

If teachers (and their leaders) are generally satisfied with the current situation, then they are unlikely to be motivated to change it although they may perceive that it is desirable to improve some aspects of practice in minor ways in order to “keep up” or because “there is always room to improve”. In these situations, teachers are rarely motivated to initiate and sustain fundamental change because there is no obvious reason for them to do so. Initiating change then becomes the responsibility of the facilitator, rather than the teachers. The change process, therefore, is unlikely to be sustained once the facilitator leaves. It is only when teachers are clear about the problem to be solved and become motivated to continue to seek new knowledge and skills that the change process becomes the responsibility of the teachers and more likely to be sustained. It is at this point, therefore, that issues of sustainability need to be addressed, not in the closing stages of professional development contracts.

Prior to making any recommendations or suggestions for change, it is important to determine whether concerns held by the facilitators about observed practice, are shared by the teachers. If the concerns are shared, they can act as a motivator for teachers to seek the necessary new knowledge and skills with the assistance of the facilitator. If not, it is unlikely that any suggestions for change by the facilitator will be acted on by the teachers because there is little reason to do so apart from doing what experts tell them to do. The teachers themselves need to drive the change process through some dissatisfaction resulting from critique and reflection on the current situation.
Robinson (1993) recommends that three criteria form the basis for such a critique and motivation for change. The first criterion inquires into the congruence of current beliefs and practices with the individual’s espoused beliefs of what is held to be desirable. It asks teachers, “Is this the kind of teacher I want to be?” For the teachers in this school, their espoused theories and the observed “theories-in-use” were reasonably congruent and unlikely to act as a lever for change. Children being inspired and motivated, and stimulating others as articulated in their beliefs is not a reason for dissatisfaction and their practices were mostly designed to promote these qualities.

The second criterion asks the teachers to reflect on the extent to which their theories fit with more accepted theories, such as those about teaching writing. This criterion can act as a motivator for change and might have been an avenue to explore if the teachers had possessed better knowledge of alternative theories of teaching writing. One possibility in this situation might have been for the facilitator to assist teachers to become more knowledgeable about alternative theories and their related practices. However, she chose as her lever for change the third criterion which presented a much more powerful option in this situation. This criterion relates to the consequences of practice. In this situation, the criterion asks teachers to consider the extent to which the consequences of their teaching practices are those desired. The way the facilitator went about the process is discussed in more detail below.

**Using consequences as the lever for change**

In this case study, there were two consequences about which the teachers cared. The first was the students’ low achievement in writing. Although they had not yet analysed their asTTle data, their previous analysis of writing samples using the Exemplars, had provided some awareness of the extent of this problem. The second consequence was the students’ limited understanding of what it was they were supposed to be learning during the writing lesson that became evident in the interviews with the researcher. When interviewed, students were all able to tell the researcher what it was they were supposed to be doing. However, when asked what they were supposed to be learning about writing, they either did not know or responded at the level of surface features only, such as punctuation, spelling, neatness and length. Similar responses were obtained to questions about success criteria and feedback. The limited student understandings of the learning aims and success criteria implicit in the lesson came as a surprise to the teachers. They had anticipated that the students would be much clearer about what it was they were supposed to be learning and the deeper aims the teachers had articulated in their planning.

By focusing on the consequences as the lever for change, critique of teacher practice was not undertaken in terms of some preferred theory of practice, but rather arose from teachers’ concerns about the consequences of their practice. The facilitator then drew a third box on the sheet of newsprint under the other two, labelled it “Consequences”, and summarised them as follows:

**Consequences**

- Teachers concerned about achievement
- All children wrote something [during the lesson]
- Didn’t know the features of a recount
- Weren’t sure whom they were writing for
- Weren’t sure how to improve their writing
- Weren’t sure how the teacher could help them improve their writing
The evidence from the student interviews created the needed dissatisfaction and provided a powerful lever from which to examine the adequacy of current theories and practices. The teachers agreed that if they wanted the students to be clearer about the features of recounts, have self-improving strategies and to become aware of audience as stated in their lesson aims, then all these attributes needed to be addressed more explicitly in their teaching practice.

They were then able to identify what it was they needed to learn and do and were able to develop a preliminary action plan involving readings to improve their pedagogical content knowledge, and seeking assistance from those with relevant expertise. Before the facilitator returned for the next visit the staff had spent a staff meeting discussing the ten characteristics of quality of teaching from the Best Evidence Synthesis (Alton-Lee, 2003) and Using Evidence for Teaching Practice (Timperley & Parr, 2004). More specific to writing, they asked the Resource Teacher of Learning and Behaviour to assist them to mark and level students’ writing samples using the Exemplars.

Feedback from the questionnaire and scenario
Prior to the next school visit, a summary of teachers’ combined responses from the questionnaire and scenario became available so the facilitator decided to look at the implications of these summaries in relation to those from the classroom observations. At the staff meeting they examined these summaries, discussed them, and with the facilitator’s assistance identified if their responses indicated a strength or a learning need and whether it should be included in their developing action plan.

By way of example, the questionnaire contained confidence scales that all began with a stem, “How confident do you feel with respect to the following statements: ‘I know how to develop the important learning objectives for a writing lesson’. Other confidence items included making the criteria for successful learning in writing clear to students and providing feedback about writing in ways that support further learning. A 1-6 scale was used with the descriptors “Definitely not confident” at one extreme (rating 1) and “Highly confident” at the other (rating 6). Staff ratings for each of these confidence items averaged 4.4 with a range from 3 to 6. After presenting these data, the facilitator asked how confident they would like to be and they all agreed they would like to be more confident, but to do so they needed to learn new skills.

The questionnaire also contained a scenario of a writing lesson in which the lesson aims were misaligned with the lesson activities. It had an additional item presenting student data from a hypothetical class for staff to interpret. None of the staff detected the misalignment in the writing lesson, and some indicated that they were unfamiliar with data like that presented for the hypothetical class so were unable to complete the exercise. The facilitator used these results to support the concerns the staff had expressed about their confidence and their desire to learn more about teaching writing and interpreting relevant evidence.

Throughout the process, the facilitator put responsibility for identifying problems and developing solutions primarily with the staff. She did not remain a silent partner, however, but guided the staff through the relevant data, assisting with interpreting their meaning and providing some options for change. With her assistance, staff listed the following ideas for change (Figure 8.3) alongside their analysis of beliefs, practices and consequences.
Figure 8.3
Ideas for Change Arising from Previous Analysis

1. Need to present models of quality writing
2. Need to investigate the features of these with the children
3. Need to give more focused feedback
4. Need to make learning intentions and success criteria more explicit
5. Need to share published work
6. Need to teach children to give focused feedback

To get started we are going to:
1. Read Chapter 4 E.L.P. on feedback approaches
2. Look at text features in First Steps
3. Read about learning intentions and success criteria in “Using Evidence in Teaching Practice”
4. Use exemplars for quality writing.

Establishing priorities
A dilemma that is difficult to resolve for any facilitator in this type of situation is whether to capitalise on the enthusiasm the teachers showed for getting started, despite the fact that they did not have sufficient pedagogical content knowledge to execute that change as well as they might, or to delay giving ideas for change until that knowledge base is established. There is no right “decontextualised” answer and both have their risks. In this situation, the principal was willing to lead the change and with the facilitator’s assistance identified appropriate resources, so it seemed unnecessary to delay the action. However, it is important to assess the likely risks associated with proposed activities and to decide on the priorities for particular learning needs. In this situation, the facilitator was confident that the staff would be able to derive adequate knowledge from structured readings to implement some new teaching practices so she decided to recommend readings as indicated in Figure 8.3 for them to undertake independently. She also believed that structured classroom observations and feedback could create powerful learning opportunities. However, she was less confident that these could be undertaken independently, so decided to focus her next visit on training the literacy leaders to observe classroom practice and provide feedback to the teachers.

Observation and feedback training
The facilitator articulated her theory underpinning the observation training in the following way:

- It was important to trust the principal and literacy leader to undertake the observations and, by doing so put the responsibility for the execution of the observations and feedback on to them;
- The facilitator’s job was to set them up for success, which meant working through the process with them and while doing so:
  - Discuss any issues that arose during the process;
  - Encourage them to look for evidence that backed up any statements they were going to say to the teacher they observed;
  - Ensure that the feedback they gave to the teachers was likely to be effective in helping them to monitor and change their practice;
  - Help both observers and teachers to keep the purpose of the observation in mind rather than get distracted by irrelevant detail.
In addition, the facilitator’s job was to:
  - Make it a purposeful experience for both observer and teacher;
  - Help the principal and facilitator to realise the importance of observations;
  - Help the school to develop a learning community through opening the classroom doors;
  - Make the observations challenging for the teachers.

Consistent with this theory, she spent the day of her next visit training the literacy leader and principal in the process of conducting observations and giving feedback in structured and focused ways. The training involved:

- Discussing the observation schedule [supplied by Learning Media], the theory underpinning it and the importance of using the schedule for improving practice;
- Discussing the evidence (key practice indicators) to guide the observations and how to record, analyse and summarise the evidence;
- Focusing the observations on the teacher’s learning goals and designing them as a support to achieve those goals;
- Encouraging the literacy leader to negotiate the focus of the observations with the teachers prior to the observation. This was done by going through the indicators with the teachers and asking them if there were additional things they would like the literacy leader to focus on during the observation;
- Giving feedback in ways that invited discussion on how to improve practice;
- Stressing that the process was not easy and not to expect to get it right first time.

During the observation training the facilitator and literacy leader observed a lesson, independently filled in the observation sheet and compared notes. In particular, the facilitator stressed the importance of focusing the observation on how to improve practice and to have sufficient evidence to back up any conclusions so that the basis for any statements about a given teacher’s practice was clear. The final step in the training involved giving feedback. The feedback processes were discussed with a suggested framework provided. The literacy leader gave feedback to the observed teachers and then discussed the teachers’ reactions with the facilitator.

In the follow-up research interviews, the literacy leader indicated that the areas of training they found most useful were the recording system and the assistance in being clear about the feedback messages they wanted to give. Subsequently, one of the literacy leaders and principal undertook five observations of each teacher and encouraged the teachers to interview their students. Through this process the teachers developed the skills to monitor the consequences of their practice and self-regulate their ongoing learning. During the first round of these interviews the teachers found that the students continued to be confused about what it was they were to learn. Consequently, the teachers decided that they needed to be still more explicit.

**Using student achievement data**

The marking of the asTTle scripts was not completed by the research team until the fourth visit by the facilitator to the school. Examination of these data confirmed that overall student achievement was well below national benchmarks. In the Year 4/5 class, the Year 4 students were scoring more than one standard deviation below the mean for all New Zealand schools, with the Year 5 students scoring only slightly better. All curriculum functions were below the mean for both groups. The worst features were structure and spelling, and the best feature was punctuation. In the Year 5/6 class, the Year 5 students scored
approximately the same as their Year 6 counterparts but were almost one standard deviation below the mean. Once again, none of the curriculum functions was comparable with the New Zealand mean, with the worst features being structure and language resources, and the best features being punctuation and spelling.

Subsequent research interviews established that the data confirmed what teachers had suspected but the qualities of the asTTle assessment gave them direction about where they needed to focus their efforts. The facilitator scheduled a follow-up visit to examine the data in more detail.

Staff reactions to the needs analysis process
Researcher interviews of the principal and two literacy leaders in early November 2004 found three very enthusiastic and well informed teachers who all reported on the increased pleasure they now derived from teaching writing. The principal described the needs analysis process as one of developing a sense of ownership through the feedback in that “... it gave staff a voice and a chance to discuss what their beliefs were and their practices, and that’s been half of it with the growth that’s happened here, and that chart and being able to discuss .... It’s just that openness, that’s probably been the biggest thing. And the fact that sometimes we’ve thought what children know is not what children know and what we think they can do, sometimes they can’t”.

One of the literacy leaders reported that previously both she and her students “loathed” writing but now loved it. She no longer had to spend most of the time motivating the students to write, and so was able to spend much more of her time explicitly teaching them how to write with focused learning intentions and success criteria. The use of these more explicit techniques had provided sufficient motivation in themselves for the students, so she did not need to spend so much of the lesson in alternative motivational activities unrelated to the writing aims. As a result the students got more time to write.

The analysis of the observations of the Year 2/3 and 4/5 teachers showed significant changes in practice and in the interview responses of their students. There was insufficient time to observe all three of the previously observed teachers and it was decided not to observe the Year 5/6 teacher who was part time and intended to retire at the end of the year.

In both classes the observed lesson aims presented to the students were consistent with those described by the teacher to the researcher. Most of the lesson interactions focused on the stated aims and in their interviews, students indicated that they understood them well, particularly in the Year 4/5 class. The only area in which the students were unsure was their understanding of the feedback given by their teacher. A brief description of the lesson is provided below.

In the Year 4/5 class the written aim was constructing the end sentence of a recount using the context of retelling about a favourite event. The teacher’s analysis of the students’ work from the previous day established that they had structured the beginning and middle parts of the recount well but many students did not have a concluding sentence. She began the lesson by revising the definition of a recount and its structural features. She did this by questioning the students and summarising their answers. For example, she asked the students at the beginning of the lesson, “Who can tell me what a recount is?” When a student began to give her details she prompted, “There’s a sentence that sums it all up”. After establishing the definition of a
recount, she then prompted the students to describe the features of recounts. She read three students’ stories that had these features, drawing attention to them as she read, and finished by illustrating the qualities of the final sentences. Those students who had finished their recounts were instructed to prepare their stories for publication. All prompts and feedback were related directly to the qualities of recounts or directions for publication. Written feedback, however, still focused primarily on spelling corrections (34 instances), ticks that did not clearly relate to any particular writing quality (20 instances) and word changes (2 instances). There were six comments related to the qualities of the recounts as described in the aims for the lesson.

The Year 2/3 teacher had broader aims. She described them to the researcher as “Learning to do a recount – structure and order of events”. She described it to the students as, “We are going to do a recount about our best birthday party”. She was less focused in her interactions and feedback than the Year 4/5 teacher in that she mostly focused on the beginning sentence in her modelling and feedback, rather than her broader aim of structuring and ordering events. She also referred to other implicit criteria such as mechanics, attributes of the person, finishing sentences and she gave non-specific praise. Written feedback had not changed from the earlier observation and focused on mechanics and motivating praise. Students’ understandings had become more focused on the deeper features of the writing task, including the content of different parts of recounts, particularly the beginning and middle. However, their understanding of the feedback (“What does your teacher tell you to work on?”) was vague with three answers indicating that they did not know and three general answers such as “doing writing”, with another three relating to mechanics.

The principal indicated that the teachers’ knowledge and beliefs about teaching writing had changed in that they realised that they had to be more specific about what it was they were to teach, and more explicit about their outcome criteria. In the teacher interviews, however, neither was able to articulate how her beliefs about writing had changed from the initial analysis. In itself, failing to recognise how more sophisticated beliefs inform practice may not matter because practices are changing and teachers reported enthusiastically about the much higher quality of students’ writing. However, if schools and teachers are to sustain change independently of facilitators, helping them to become more aware of the ways in which their beliefs about teaching writing were evolving would put them in a stronger position to drive their own change process in the future.

**Outcomes for Students**

The repeat aTTle test over the four month period showed significant gains for all year levels and these are reported in Table 8.2. The overall effect size achieved over the four month intervention was 1.03. According to Cohen’s (1977) criteria, this effect size is large. The scores of students not sitting both tests were deleted from the analysis so the descriptions of achievement are slightly different from those reported initially.
Table 8.2
Means and Standard Deviations for asTTle Scores in August and November 2004

<table>
<thead>
<tr>
<th></th>
<th>Overall mean</th>
<th>Year 4 mean</th>
<th>Year 5 mean</th>
<th>Year 5 mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
</tr>
<tr>
<td>August</td>
<td>399.6</td>
<td>348.6</td>
<td>400.77</td>
<td>422.5</td>
</tr>
<tr>
<td>(80.83)</td>
<td>(52.74)</td>
<td>(74.27)</td>
<td>(99.99)</td>
<td></td>
</tr>
<tr>
<td>November</td>
<td>476.2</td>
<td>438.8</td>
<td>471.00</td>
<td>506.3</td>
</tr>
<tr>
<td>(66.85)</td>
<td>(52.86)</td>
<td>(59.03)</td>
<td>(81.57)</td>
<td></td>
</tr>
</tbody>
</table>

The console reports for the separate curriculum functions showed that in Year 4, content, language resources and spelling dipped slightly below the mean compared with all New Zealand students but all other curriculum functions were slightly or well above the mean. In Year 5 a similar pattern emerged, with content, language resources and spelling slightly below, and the other curriculum functions at or well above the New Zealand mean. In Year 6 all curriculum functions were above the mean.

Discussion

Facilitating teacher learning in ways that achieve outcomes of improved teacher pedagogical content knowledge and student achievement, through a needs analysis process and the development of a professional community is not easy. It is tempting within a needs analysis framework to regard the initial analysis phase as one of data-collection on which to base subsequent “professional development”. This case study demonstrates an alternative approach in which feedback throughout the needs analysis process gave teachers the knowledge of a problem in need of a solution and the motivation to engage in and promote their own learning soon after the start of the project. It needs to be acknowledged that this momentum is easier to achieve in some schools than in others and the specifics of an effective approach within a particular school is inevitably contextually determined.

The eight principles outlined in the introduction to this chapter were all evident in this case study. The application of the first principle - that professional development should be driven by an analysis of the differences between goals and standards for student learning and student performance (Fullan, 1993; Howey & Collinson, 1995; Pink & Hyde, 1992) was evident through the analysis of the student achievement information and student interviews. These analyses defined what the teachers needed to learn rather than what they wanted to learn and helped to focus the outcomes of the professional development on student achievement.

The needs analysis and feedback process demonstrated the application of the second principle - that professional development should involve teaching professionals in the identification of what they need to learn and, when possible, in the development of the learning opportunity and the processes to be used (Borko & Putnam, 1995; Newmann & Wehlage, 1995). This engagement increased their motivation and commitment to learn (Hodges, 1997); empowered them to take instructional risks and assume new roles and responsibilities (Pink, 1992); increased the likelihood that what was learned was meaningful and relevant to their particular contexts and problems (Pink & Hyde, 1992). By summarising teaching practices, probing for beliefs and indicating some of the problematic consequences of their practices, the process left them feeling...
both respected and able to identify for themselves what it was they needed to do to change the problematic student outcomes.

One of the difficulties in managing professional development in this way was that the facilitator gave up much of the control about what was taught and when it was taught to the staff. In giving so much control to the teachers, the facilitator needed to have sufficient depth of content knowledge and the change process to be able to guide and facilitate learning as and when needed. A pre-determined professional development programme and sequence of activities would not have worked.

The application of the third principle - that the professional development should be integral to school operations, while avoiding the pitfalls of negative micropolitics was also evident (Feiman-Nemser, 1983; Grossman, 1992; Guskey, 1995a; Joyce & Showers, 1995; Little, 1993; Little & McLaughlin, 1993; Louis, Marks, & Kruse, 1996; Smylie, 1995). By involving nearly all the staff in the observations and feedback, they identified their common learning needs and guided the professional development designed to meet them. Through structuring the professional development in this way, the need for the fourth principle of providing learning opportunities that relate to individual needs but are primarily organized around collaborative problem solving (Fullan, 1991; Guskey, 1995a; Hargreaves, 1994; Huberman, 1995) was also evident.

The importance of the fifth principle - that professional development should be continuous and ongoing, involving follow-up and support for further learning (Fullan, 1993; Guskey, 1995a; Hodges, 1996; Pink & Hyde, 1992) was evident through frequent facilitator visits through this phase of the contract. Her visits, however, were not sufficiently frequent in themselves to meet the urgency of the teachers’ learning needs, so alternative support and ongoing learning opportunities were organised to keep up the learning momentum. These learning opportunities also related to the sixth principle - that the professional development incorporate evaluation of multiple sources of information on outcomes for students and processes that are involved in implementing the lessons learned (Guskey, 1995a; Joyce & Showers, 1995; Little, 1993; Tillema & Imants, 1995). Similarly, the seventh principle - that professional development should provide opportunities to engage in developing a theoretical understanding of the knowledge and skills to be learned (Eraut, 1995; Feiman-Nemser & Parker, 1992; Fullan, 1991; Joyce & Showers, 1995; McDiarmid, 1994; Tillema & Imants, 1995) was applied through these ongoing learning opportunities.

Achieving the eighth principle - that professional development should be integrated with a comprehensive change process that deals with impediments to, and facilitators of, student learning (Guskey, 1995a; Little, 1993; Smylie, 1995) could only be started within the time frame. Although this initiative was framed as one of professional development in literacy by involving the whole school, a more comprehensive change process had begun. It is only when the school-wide issues are addressed, such as developing professional learning communities within the school and monitoring the impact of practice on students, that the chances of sustaining ongoing learning are increased. Trusting teachers with the bad news as well as the good news, enskilling them to make changes while at the same time helping them to become aware of the process in which they are engaging, maximised the chances that the change would become embedded and sustained. This was not a process of professional development being delivered and “done to” the staff but rather a learning journey in which they were fully and respectfully engaged.
References


Hodges, E. (1997). Negotiating the margins: Some principles of responding to our students’ writing, some strategies for helping students read our comments. New directions for teaching and learning(69), 77-89.


Chapter 9
Facilitation Challenges

This chapter follows on from the previous chapter and should be considered in conjunction with it. While Chapter 8 described the needs analysis process in one school, this chapter reports facilitator reflections and development over the first two years of the project. The purpose of this chapter is to identify the challenges involved in facilitating teacher professional development in projects like the Literacy Professional Development Project. The chapter begins with the facilitators’ reflections on their learning at the end of the second year. The second section then traces some milestones and challenges in that learning journey at two key stages. The chapter concludes by bringing together the main themes identified by the facilitators and indicates future directions for this aspect of the research.

The data on which this chapter is based consist of audiotaped transcriptions of team conversations, interviews with teams and their leaders and the completion of a written scenario. Details of the data collected for each section are included in that section.

Reflections Back on Learning

This section highlights key issues identified by facilitators over their two-year learning journey. At the end of the second year, they were asked in their teams to describe what they had learned over the two years and the implications for future practice. The conversations of three of the four teams were audiotaped and transcribed. Some members of the fourth team did not give consent for their conversation to be recorded and so this team’s reflections are not included in the analysis.

Four themes were evident in the transcripts of the three team discussions. These themes included, in descending order of frequency, the importance of addressing leadership and change management issues, the need to have structured conversations with teachers and leaders, key issues around content knowledge and how to exit from schools at the end of the contract in ways that support ongoing learning. Each of these themes is described in turn, with representative quotes interspersed throughout in order to capture the facilitator’s intent through their actual words.

Theme One: Leadership Development and Change Management skills

One of the themes raised during these team discussions was facilitators’ increasing awareness throughout the project that leadership development was as important as teacher professional development. The facilitators’ comments referred to their realization that the teachers’ learning on which they had initially focused would not be successfully promoted or sustained if leaders were not also learning how to lead the project within their schools and manage the change process. One facilitator expressed this growing realization in the following way:

I felt reasonably secure with what I might do with kids and with teachers, but I didn’t feel so secure with the leadership, rather I didn’t actually have an expectation – an expectation to grow the leadership was a fairly gradual realization on my part... but when I’d identified that leadership actually was in the mix, it was dramatically effective.
Another group’s discussion focused on their learning about the importance of developing distributed leadership throughout a school, particularly in large schools, so that more than one literacy leader had the opportunity to develop leadership skills and literacy knowledge. One facilitator described how a more traditionally defined concept of leadership in one school had led to lack of follow-through on agreements made and to considerable teacher resistance. When the leadership became more distributed with all syndicate leaders participating as literacy leaders, she reported more collaborative problem solving took place with much higher levels of teacher engagement.

The need to develop such leadership arose from the needs analysis and the emerging focus of the project that both teaching practices and the beliefs underpinning them needed to be challenged for sustainable change. One facilitator described schools’ initial expectations of professional development in these terms, “The models they’ve had is people coming in and showing them for a day and teachers just thinking, ‘Maybe when I go back I might try that in my classroom at some point in time’”. She concluded by expressing her intent to address issues of change management much earlier with leaders in the new schools in the following way.

> What it means to lead change would be an important part of the first phase.... Sharing an understanding of change management, what that looks like, what that means.... It involves helping schools to move from ‘I’m a literacy leader, I need to be this and I need to be that and everyone will do it’. It became a knowing that you are a literacy leader and leading people to change their habits, their beliefs, the way they do things and to do that I need to have an understanding of what it might look like for you.

The approach to professional development adopted in the project also underpinned the reason for another facilitator to think about the importance of developing distributed leadership throughout the school.

> I realised that I couldn’t deliver it.... The key thing is that your job is to help others rather than take on the responsibility to change the school and doing all the work. I think when you look at that and decide to work through a strategic leadership approach you are giving the power to those leaders.... I think the key in that situation was that you are a member of the team working out a solution rather than coming in with a solution and saying ‘Well this is the way to do it’.

One facilitator was concerned that knowledge of change management was not confined to the formally appointed leaders but also wanted teachers to have the information “... so that they can actually critically reflect on that knowledge [of change].”

### Theme Two: Conversations

Facilitators had received some training in having a particular type of conversation with their schools that involved explicit contracting for the purpose and process of key conversations, the need to check mutual understandings and agreements and to be clear about their reasons for asking questions. Being explicit about reasons or the evidence on which beliefs and judgments were based is also an important feature of such conversations. The key underpinning values of these types of conversations are the development of mutual respect and trust through openness, and the promotion of mutual learning by ensuring that important information in any situation is shared by both parties. More detail on the values and structure of such conversations are in Robinson (1993) and Timperley (2001). A common theme in the facilitators’ reflections
was their difficulty in becoming more deliberate about their conversations in these ways while, at the same time, realizing that achieving greater clarity created improved professional learning conditions, greater respect and deeper levels of trust. Some facilitators talked about the need to make all conversations learning conversations. Others were more specific about the situations in which they had used them successfully. One facilitator mentioned how she had previously “fudged” her feedback to teachers in ways that resulted in their having no opportunity to understand what was problematic about their practice or to learn. Since being more explicit in her conversations, she believed she was being more effective in helping her teachers learn.

One of the things that I learned from that is being very, very clear with teachers in terms of contracting the observations and discussing exactly what the purpose is so that both sides understand. I think that’s hugely important ... Also helping them to be explicit where they are in their next step ... because teachers actually do want to know what to do, they want good quality feedback.

Another facilitator also mentioned the usefulness of learning conversations when giving feedback. She recalled an instance when she asked a series of questions about an observation without giving a reason for the questions, and how this approach had “put her [the teacher] on the back foot and she got very defensive.” But the challenge she articulated in having such conversations was, “I can’t say ‘excuse me, can I just have five minutes to work out what I think’.” She reported how working on this aspect of her practice, in terms of giving reasons and evidence had resulted in more problem-solving conversations. “Keeping that going has been great, but it has been my greatest challenge”.

Another described how she used asTTle data in a learning conversation to help an apparently competent teacher to understand why his students were not achieving as well as others in the school. After the teacher blamed the school, the test and the genre assessed, he admitted that he was not taking his groups for writing and he did not really use models because he did not know how. He also agreed to participate in the coaching offered by the facilitator, an offer he had previously declined.

Reasons for the effectiveness of using such conversations were consistent with the values underpinning them as expressed by other facilitators:

Facilitator 1: “It is clear and people know the purpose of it and they get lots of opportunities along the way to check in to see if it is kind of what they are thinking ... the whole explicit nature and deliberateness I guess makes the conversation.”
Facilitator 2: “There’s none of ‘guess what’s in my head’.
Facilitator 3: “Both sides understand what the conversation is going to be around.
Facilitator 4: “The whole idea of treating each other with absolute trust and absolute respect and learning together I think that is probably one of the biggest successes that I’ve had.”

Theme Three: Content Knowledge

Several facilitators mentioned the value of particular content knowledge that they had acquired during the contract, such as understanding asTTle writing, learning about reading comprehension and how to interpret and use student achievement data. One facilitator reported, “I think actually the biggest learning that I’ve
made is just how important that data is. The more I look at it, the more I get from it especially with a deeper analysis”.

A common theme was a need to be more aware of their own content knowledge as facilitators and the beliefs on which it was based if they were to impart that knowledge to teachers and their leaders. One facilitator described how she needed this knowledge in order to understand the teachers’ knowledge, and to work through what that meant for her facilitation practice. Her concern was that she had sometimes “found this [teachers’ knowledge] out by accident” rather than in a structured way. She intended to be more systematic in this aspect of her practice in new schools.

A second facilitator described the need to have in-depth knowledge in order to challenge teachers’ knowledge and beliefs. She described a situation in which the teacher and literacy leader in a school dismissed a child’s reading difficulties by claiming he was lazy, rather than not having requisite skills. It was only through the facilitator’s ability to accurately diagnose the child’s reading difficulties that she was able to challenge the “laziness” beliefs.

Theme Four: Exit Strategies

Given the timing of these conversations at the end of two years of working with schools, it was not surprising that a key theme involved ways to withdraw from the school at the end of the contract that left those within the school in a position to continue their learning. Several facilitators mentioned that to achieve the point where the schools had both the capacity and willingness to take control of their own learning required a different set of skills than those traditionally used in professional development situations. It required the facilitator to promote skills within the school that allowed them to continue to problem solve and to learn. As one expressed it, “... giving them [the teachers] the facility to do it for themselves or to problem solve so that they can move forward.” This involved a deliberate process of making themselves as facilitators redundant and explicitly negotiating an effective process within the schools. As one facilitator described,

I was up front about it and I said, ‘Because I’m not going to be with you beyond 2005, I am going to have to make sure that you are given the opportunity to develop practices so that this project will be the ways of working and will become sustained and automatic for you. So I won’t be giving you as much direct help. I will be working in a different way’. They thought that was cool, they really did.... I was aware that when I did that there were some huge shifts in the way that literacy leaders, principals and teachers began to respond. They actually began learning and I found them beginning to take new directions. I was absolutely delighted to see the decisions they were coming up with.... There were some frustrations because everything slowed down. People took longer to think things through, come up with solutions. But it was also quite humbling because people were coming up with ways of working and ways of embedding things in practice that were actually better for the specific school than I had thought.

Becoming redundant, however, can create mixed feelings as the following quotes from two facilitators illustrate.

Facilitator One: “The learning for me is the redundancy feeling that you sometimes have and it’s kind of exciting because you can actually go to another level and I’ve got to learn to think it’s an OK
place to be at and moving deeper into all our learning and letting schools have a go at doing what they want to do.”
Facilitator Two: “A couple of times I had to ask myself, ‘Did I actually do a day’s work?’ It was a funny feeling really.”

Summary
These quotes from the facilitators indicate that most had come to understand this project as a process of assisting schools to develop the literacy and leadership knowledge and skills to sustain their gains and continue to learn. Several noted how different the approach was from the type of professional development in which they had been involved prior to the contract. The change had not been easy, however, and the remainder of this chapter identifies some major milestones and challenges involved.

Re-tracing some Aspects of the Learning Journey
Data were collected at two points in time in the complex learning journey of the project. The first was approximately six months after it had begun and was timed to coincide with the end of the needs analysis phase. Several facilitators, seven school principals, seven literacy leaders and twenty-one teachers were interviewed to gain their perspective on the needs analysis process. Their responses are summarised in the next section. Further data were collected from a larger group of facilitators approximately eight months after this time to better understand some of the changes they had made and issues they faced. Their responses are described in the second section.

Understandings of the Needs Analysis Approach
Two aspects of the needs analysis were different from the ways in which professional development was traditionally undertaken in schools and so were unfamiliar to the facilitators. The first was the idea of undertaking a needs analysis involving collecting data on teacher content knowledge, beliefs and practices and student achievement, all of which was to be used as the basis for the subsequent work in the school. The second was that these data should be fed back to teachers and leaders in ways that allowed them to identify their learning needs and create the rationale for ongoing learning.

The research team undertook interviews at the end of this needs analysis phase. These interviews involved asking the respondents to nominate a category that described their reactions to information from the classroom observations, student achievement information and questionnaires for the staff that were undertaken as part of the needs analysis process. The response category options were “Major new insights”, “Minor new insights”, “Confirmed what I knew” and “Not useful”. As the research interviews proceeded, an additional category was added, “Can’t remember”. The responses are summarised in Table 9.1. The interviewers were surprised at the difficulty some of the school respondents had in recalling the data and how little value many placed on the process. Clearly the process was not compelling for most of the school personnel.
Table 9.1
Responses to the needs analysis in other research schools

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>Student achievement</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major new insights</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Specific / minor new insights</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Confirmed what knew</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Can’t remember</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Not useful</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Literacy leaders</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major new insights</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Some / minor new insights</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Confirmed what knew</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Confused</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Teachers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major new insights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major/minor new insights</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Some / Minor new insights</td>
<td>5</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Confirmed what knew / affirming of good practice</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Not useful</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Could not recall / absent / no feedback given</td>
<td>7</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Hated discussion</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Interviews with facilitators quickly clarified the reasons for these reactions from school personnel. The first purpose for the needs analysis, to establish the basis for ongoing work in the school, was well understood. The second purpose that schools and teachers should spend time discussing the interpretations and implications of the data in order to identify their own learning needs, had not been acted on because it was not so well understood. Rather, facilitators had understood that the data were primarily for their own use for ongoing planning and feedback had been minimal. The facilitators’ interpretation of the process was understandable because how to use the data had not been made explicit, nor had the skills related to giving feedback been specifically attended to. Although the project managers quickly responded to this need with clarification and additional training, it is important to identify that the sophisticated understandings outlined in the first part of this chapter were not achieved through a linear and smooth process.

An interesting parallel exists between schools’ traditional use of assessment information on students and facilitators’ use of assessment information about teachers. Although the use of formative assessment in classrooms was a key part of the project, the transfer of this knowledge to apply it in the same way to promote teacher learning needed to be explicitly addressed, not simply assumed.
**Developing Understandings of Facilitation**

Approximately eight months later, all facilitators were asked to respond to one of two scenarios in order to develop an understanding of their beliefs about how to resolve typical problems that arose in schools. They were asked to indicate their initial response to the situation and to provide their reasons for this response. These written responses were then discussed in teams.

The research data comprised the written responses from all facilitators, field notes taken of three team discussions following completion of the scenario and follow-up interviews with two team leaders and a facilitator.

The scenario chosen by three of the four groups read as follows:

| The literacy leaders challenged the teachers in the junior syndicate (Year 1 & 2) over the value of writing “fun” weeks they ran twice a term. The junior teachers approached the facilitator for support. They said they felt the fun weeks had great value because the children really enjoyed the writing activities they were engaged in during the week and they had an opportunity to mix with children from other classes in the junior school because each day they could select which room to go to and which writing activity to engage with. |
| Twelve of the eighteen facilitators responded in a similar way. Their responses were too long to be reported individually, but in summary, their initial response was to: |
| • Clarify the teachers’ beliefs about “fun” weeks. |
| • Clarify the evidence on which the beliefs were based and whether that evidence was focused on promoting student learning. |
| • Compare their beliefs with the need for students to have structured individualized learning opportunities in order to learn to write well. |
| • Increase teachers’ content knowledge through a variety of means. |
| A summary of the essence of their reasons for doing so comprised: |
| • It is important to clarify beliefs on which practice is based and to examine the evidence of the consequences of those beliefs for students. |
| • It is important for teachers to come to their own judgments rather than have the facilitator decide. |

Some other responses were similar but focused on just clarifying beliefs (3 responses), or developing a clearer purpose for fun weeks (1 response). One other response gave a greater focus to the facilitator’s mediation role with the primary responsibility being to resolve the differences between the teachers and literacy leader. A final response was to advise the literacy leader about the value of writing being fun while developing other strategies.

One team chose to respond to an alternative scenario. In this scenario, details were provided about how facilitator observations of writing in the junior school revealed that all classes focused on surface features of writing and that the students’ work was formulaic in nature. The teachers expressed the belief that students needed to learn the basics before they could write.
The facilitators’ reported initial responses were similar to those above. Two facilitators indicated that they would inquire into the teachers’ beliefs, probe for the evidence related to student needs, and improve teachers’ knowledge of the writing process so they could take a less mechanistic approach to their teaching. Four others mentioned two of these three elements with one other reporting a response indicating inquiry into beliefs only.

The majority of the responses to both scenarios indicated the impact of the emphasis of the project on facilitating deeper professional learning by examining the beliefs on which practice is based, making decisions on the basis of the evidence in relation to student needs or the impact of practice on students, and promoting student learning through improved knowledge on which to base more effective teaching. However, when facilitators were asked to articulate exactly what they would say in these situations in the follow-up discussions, it became apparent that most restricted their input to asking questions. They were careful not to give their own opinions directly. The likely consequences of this style of conversation were discussed, and the three teams agreed for their team discussions to be part of the research.

Although the questions, as facilitators formulated them, were probing of teacher and literacy leader beliefs, at the same time they were highly leading in the sense that the facilitator’s reasons for asking them and suggestions for how to improve decision-making were embedded within the questions. For example, if a facilitator asked teachers for the evidence of the impact of fun weeks on students, then it was obvious that the facilitator believed that evidence was important for making decisions. Alternative conversations were modelled for the facilitators, in which they were encouraged to be clear what they considered to be important, to give their reasons for their beliefs and to check for agreement or disagreement from the teachers and literacy leaders. In this way, the same principles of uncovering teacher beliefs, clarifying evidence related to impact on students and giving the school-based personnel responsibility for decisions were evident, but facilitators were able to be more explicit about why they thought these aspects were important and gave others a chance to debate the options. In this way greater trust and mutual respect could also be achieved because the agendas were open rather than buried in questions. In addition, the facilitator’s expertise could be used more explicitly while not providing “the answer”.

**Interview Responses**

The team discussions were followed by two interviews with team leaders and one with a facilitator. In these interviews, the interviewees were asked how this project was different from their previous experiences and the main challenges that were associated with those differences. All expressed in some way the sentiments in this quote from an experienced team leader “It has been the steepest learning curve of my life”. The differences from other projects in relation to the work in schools that the interviewees included in their responses follow:

- The focus on improved student achievement (beyond just passing a test) as an expected outcome;
- The project is supporting changes in practice across the whole school including leaders;
- The development of professional learning communities within schools is focused on promoting professional learning in ways that impact on students;
- Systems are being put in place in schools to ensure changes are sustained and ongoing;
- The change from providing teachers with new knowledge and inspiration and assuming they were able to put it into practice, to more explicitly addressing the translation of that knowledge into sustained classroom practice.
These changes in schools required differences in facilitation practice. Differences in this project from others that specifically mentioned facilitation included:

- An emphasis on the role of facilitation and the ongoing training and readings that have challenged previous assumptions and facilitation practice.
- The key challenges for facilitators have shifted from imparting content knowledge to taking a group of teachers on a learning journey;
- The use of data in the broad sense to make a range of decisions e.g. examining facilitation processes to analyse what was leading to teacher resistance in order to have better levels of engagement, rather than giving advice about how to overcome resistance.

The specific challenges mentioned in these interviews included:

- Making links and learning explicit so teachers are able to understand more clearly what is being talked about;
- Accepting that just because you have taught something, it has not necessarily been learned;
- The “jump” from being a classroom teacher to a facilitator of teachers’ learning is huge and there is insufficient time to cover facilitation skills in depth;
- Facilitators need more time in schools to achieve the depth of change for both teachers and leaders;
- Taking the risk to deepen the talk within the facilitation teams, which has meant being public about what is being learnt together with the associated insecurities and personal challenge this process inevitably brings.

**Conclusions**

In the reflections on completion of two years on the project, facilitators showed a sophisticated understanding of how to facilitate teacher learning in ongoing and sustainable ways. The previous two sections of this chapter demonstrated that getting to this point involved “steep learning curves” for most. A distinguishing feature of the Literacy Professional Development project is that it has been framed as a learning project for everyone. Over time what has needed to be learned and how it is learned has been identified and addressed. As with any new enterprise, it was not clear at the beginning what learning might be needed or what the journey might look like. This chapter has documented some of the process in order to identify the associated challenges.

A limitation of the data presented in this chapter is that they consist primarily of facilitator reports about their beliefs and understandings in relation to their facilitation practice. As one facilitator noted in the final section, in any professional learning situation translating understandings and knowledge into sustained practice cannot be assumed. For this reason, the research on facilitation practice over the next two years will use a different data-base in that it will involve collecting evidence related to actual practice and its impact on teachers. In this way, facilitators, leaders and researchers will continue to develop further understandings of the reality of what is involved in developing effective facilitation of teacher learning and so contribute to a very limited research knowledge base about how to do so.
References


Chapter 10

Professional Learning Communities

In any change situation, teachers need opportunities for ongoing discussion of the change process in order to make sense of it and to receive support in implementing new practice (Spillane, Reiser & Reimer, 2002). A considerable body of research has supported the idea that in-school professional learning communities can be a powerful vehicle for providing these opportunities (e.g. Louis, Kruss, & Marks, 1996; Louis, & Leithwood, 1998; Louis, & Marks, 1998). However, in-school professional communities can also lead to the entrenchment of previous practice rather than to its improvement unless particular conditions are evident (Little, 1999; Timperley & Robinson, 1998).

Our analysis of the professional learning communities in the research schools focused on the presence of those conditions that have been demonstrated to have an impact on professional learning and practice in ways that lead to improved student achievement and are consistent with the goals of the professional development contract. The content of what is discussed in the professional communities needs either to deepen understanding of existing professional knowledge or to challenge the adequacy of existing knowledge when current beliefs and practices are based on unhelpful assumptions (Timperley & Robinson, 2001). In either case, both teaching practices and the beliefs underpinning them need to be the target of discussion because a recurring theme in the research literature is that simply providing useful practical advice does not change practice in ways that are sustainable.

One of the potential strengths of school-based professional communities in change situations is that they can provide ongoing opportunities for teachers to question and discuss the efficacy of new practices and for beliefs to evolve as the new practices are tried. Beliefs and practices typically evolve together, rather than in a linear manner. For this process to result in change that impacts positively on student achievement, however, it is essential that some external reference to the teaching practice itself is used to judge its efficacy. Given that the purpose of teaching is to promote student learning, evidence that it is doing just that must form a touchstone for the professional learning community to allow it to judge whether efforts to change practice are having the desired outcomes. Relying on perceptions of the efficacy of changes in practice has proved an inadequate basis for judging whether changed practice is having the desired impact (Timperley & Wiseman, 2003).

The most effective professional communities, therefore, have a constant focus on the links between teaching and learning. It may take the form of reflection, with the members of the community asking how effective their teaching has been on promoting student learning. Alternatively, evidence of current knowledge and skills may form the basis of future planning. What is most important is the constant focus on the practice/outcomes links, therefore, particular attention was paid in the analysis of the professional learning communities to the ways in which evidence of student achievement was used to assist teachers to reflect on the efficacy of past teaching practice and to plan future teaching.
The research team considered that syndicate meetings were the most likely forum in which these conditions could be created and so were used as the basis for observation and interviews. Syndicate meetings typically are made up of relatively small groups of teachers who teach similar year levels and have opportunities to discuss common issues of practice and their impact on students.

Such communities are unlikely to be effective, however, unless leaders accept their pivotal role in such meetings as one of promoting professional learning through making these links (Spillane et al., 2004; Timperley, 2005). For this reason, the participants were also asked about what constituted effective leadership.

**Method**

Two main data collection methods were used for this aspect of the research. The first was ratings of perceptions about the usefulness of meetings and the professional learning climate within the school. The second was observations of meetings with follow-up interviews about the participants’ perceptions.

**Participants’ Perceptions**

Perceptions of participants in the research schools about the usefulness of meetings for teachers to improve their teaching was obtained from questionnaires that were completed by all staff at the beginning of the contract and at the end of the first and second years. Two items of relevance to this chapter read:

- Please indicate the extent to which you agree or disagree with the following statement:
  
  Meetings at this school really help me to teach those students I find most difficult to teach.

- Please indicate your perceptions of the climate in your school by circling the appropriate number and giving reasons for your rating”.

  “The climate in this school really focuses on teachers as learners.”

*Please rate 1-6: disagree - agree*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>Agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Observations and Interviews**

Baseline data were obtained from six Northland and Auckland schools (one was not included) through audiotape and observation of a syndicate meeting, follow-up interviews of the participants and questionnaire responses. The meeting observations all took place early in Phase One of the professional development. In all Auckland / Northland schools (including the one not observed initially), follow-up observations took place towards the end of year. Additional observations were undertaken in three of the continuing schools at the end of the second year. This chapter compares the processes evident in the initial and follow-up observations and notes trends in student achievement over the year for each school.

In order to maximise the possibility that the meeting would meet the criteria for an evidence-based professional learning community, literacy leaders were asked to select the syndicate meeting for observation on the basis that either student achievement data (related to the professional development) or literacy programmes were to be discussed. Any reference materials, such as programme plan of achievement data,
used at the meeting were copied for further analysis. The methodology had serious limitations, however, in that so few meetings were observed. It is possible that the activities associated with improving student achievement within professional learning communities were occurring at other times.

The coding schedule developed by Timperley and Wiseman (2003) for coding Year One syndicate meetings was used as the basis for the coding of these meetings. Five new codes were added to the original schedule to allow more relevant coding of meetings that were typically focused on teaching older students than in the earlier research. The coding system was designed to discriminate between the attributes of professional learning communities that were more or less likely to result in raised student achievement as outlined briefly in the introduction to this chapter.

Each transcript was coded line-by-line then the number of lines per code calculated. The percentage of meeting time assigned to one code was determined by calculating the percentage of lines assigned that code. Any codes that were used for less than 5% of the meeting time were not included in the analysis for reasons of clarity.

In order to illustrate the type of conversation assigned a particular code, a randomly selected quote for each code from the initial meeting transcripts is presented below. Most quotes are from literacy leaders because of their pivotal role in developing professional learning communities. Three additional codes from the meetings analysed at the end of the second year were added, so the illustrative quotes have been selected from these later meetings.

**Analysis of Data**

**Activity – analysing Running Records:**

*Literacy leader: We really need to do the MSV analysis here. I’ve brought mine and I would just like to go over that. I’ve put an MSV on every line that there’s an error and if there are two errors on a line I’ve put MSV x2.... If it looks visually correct, 50% of the word has to be correct, so it might be the ending and the beginning, you would circle, up until that part of the word, which one’s they’ve used. It’s a really helpful tool and then what I’ve done at the bottom of mine is I’ve put ‘out of 11 errors they used 4 meaning cues, 5 structural cues and 6 visual cues’ so you can get a general picture of what meaning cues they’re using.*

**Identifying Groups / Target Students**

**Activity – identifying groups from STAR data**

*Literacy leader: You have to identify the critical and highlight them. What’s the best thing to do is choose a colour for your criticals and use that colour all the time. Then, after that, so I’ve actually written the criticals. My criticals are green. The typical range goes from 9-10, the maximum score possible is actually 12. Anything above 10-11 or 12 – is obviously children who are ok.*

**Descriptions of Resources**

**Activity – providing an overview of the term programme**

*Literacy leader: On Thursday I’m having a book seller come in and she’s got a lot of low level high interest novels. The covers look interesting. They might be great stories but the kids won’t pick them up. So these covers look really great – so I’ll order some of those. We’ve also got, just as an
example, an English resource- these ones here are for low level generally. So they have got nice little chapters and they have got pictures. There’s different titles in this series.

Descriptions of Teaching Programmes
Activity – describing how to undertake group rotation

   Literacy leader: It has got to be part of your group rotation so that I spend a lot of time, particularly with my low group last year, on going through the activities together. On your rotation one group is doing independent novel study work but it’s not every day – like once a week. Other times are instructional time. It is really really important. I know it is really tempting to go ‘off you go’ and they do it themselves. I’ll copy some stuff and add it to your folder about instructional time. Journals are the best place to go and get your stories from.

Organizational (teaching related)
Activity – assessment overview for term

   Teacher: I’m just wondering, do we do that because we’ve got integrated units...
   Literacy leader: I’m sure we did it. We looked at the topics that we’d done Term One and Term Two and we basically sat down, we have three sheets in front of us and said, OK, Planet Earth will be on, yes.
   Teacher: OK, I’ll just put that in there.

Organizational (for assessment)
Activity – assessment overview for term

   Literacy leader: Just feel, if there’s anything you want further explained, just jump in. So where are we at? Have you put down the STAR test?
   Teacher: STAR test, week two, next Wednesday.

Explanations for Low Test Scores
Activity – looking at PAT scores

   Another thing I looked at with the word recognition part is some of the words that my children circled I am sure there was confusion. Has that been factored in the test? It might explain some of their answers.

Positive Evaluations of Previous Year’s Programme
Activity - Describing modelling in reading / writing programme

   That was so powerful for me last year. I just found the kids they all came up because they could see that I’m capable of writing stuff and I gave them enough to want to write as well. They went back and they read stuff all the time.

Self-identified Teaching Problems
Activity – looking at the term overview

   Teacher 1: Why do you think Samuel’s off-task? He says he can’t do this.
   Teacher 2: You know, Carl is like that ....
   Teacher 3: I’ve got children, especially ones that come from other schools who do this.
   Teacher 1: Teletext sort of thing ... For addition of 20 plus three
   Team leader: Oh, so they’re starting on one, counting from ones, so they can’t hold the 20 in their head.
Teacher 1: Yes, and they don’t ...
Team leader: Oh my god.

The next three quotes have been selected from the meetings at the end of Year Two because these codes were not used in the initial meetings. All came from a meeting at one school where the teachers were discussing asTTle writing results.

**Implications of Data for Teaching**

*Literacy leader: Is there any sort of trend [in the results] where you think you targeted your teaching?*

*Teacher 1: I started off teaching structure and then the content and the ideas fitted quite well to that. I didn’t really focus on the audience awareness probably as much as I should. I carried on with the structure and the content and looking at the language features, using powerful language and that sort of thing. It is hard to leave one behind to focus on the other.*

*Literacy leader: Did you find that J.?*

*Teacher: I found that language resources were one that was kind of left behind a bit, but that could have been my own doing.*

*Literacy leader: Lessons that we can learn for next year.*

*Teacher: Better content knowledge and resources…teacher knowledge of a particular genre.*

**Discussing Target Students’ Progress**

*Teacher: I had some really good gains. I’m looking at V’s first piece of writing where she wrote, one, two, three sentences to her last piece here which is so logical, so clear, elaborated. It was just like she has got it and she has. She has moved 2A to 3A. B. moved a little bit. A. went down.*

*Literacy leader: Have you got examples of his?*

*Teacher: [read students’ work] .... [Conversation continues with teacher and literacy leader analysing what student was able to do and was not able to do.]*

**Reflections on Why Programme was Successful**

*Literacy leader: Any key reasons why you think it was successful?*

*Teacher: I think giving the kids a better idea about audience because it really affected the reasons [for their arguments] and also knowing more about the topics so they could actually writing on them... I also think for me it’s doing lots of planning first with those that do struggle.*

In order to understand the participants’ view of the meetings they were interviewed as soon after the observed meeting as was practical – typically either the same day or the day following the observed meeting. In the initial interviews, the literacy leader was interviewed alone, and the teachers were interviewed individually in all schools except School Four where they were interviewed as a group in all cases. In follow-up interviews the teachers were interviewed as a group in all cases. The interview questions used in this analysis asked the literacy leaders to identify the purpose of the meeting, to rate the extent to which the purpose was met and the qualities they considered to be important for leadership of such meetings.
Results
In this section, participants’ perceptions from the questionnaires are presented first, with the analysis of meeting activities following.

Participants’ Perceptions
The extent to which staff agreed or disagreed with the statement that school meetings really helped them to teach those students they found most difficult to teach changed significantly over the two years. The average ratings on the 1-6 scale for the three time samplings are presented in Table 10.1. The trends in the ratings on this table indicate that participants viewed meetings as increasingly helpful with respect to teaching such students ($t(87) = 9.48, p<.01$). Below mid-point ratings at the beginning of their involvement in the contract, changed to above mid-point ratings by the end. Principals and literacy leaders, in particular, showed large changes in ratings between the beginning and end of the first year. This trend continued throughout the second year.

Table 10.1
Average ratings for all participants of the extent to which meetings helped staff to teach students they found most difficult to teach.

<table>
<thead>
<tr>
<th></th>
<th>Prior to PD</th>
<th>End Year 1</th>
<th>End Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Range</td>
<td>Average</td>
</tr>
<tr>
<td>Principals</td>
<td>2.6</td>
<td>2-3</td>
<td>4.8</td>
</tr>
<tr>
<td>Literacy leaders</td>
<td>2.7</td>
<td>1-4</td>
<td>4.6</td>
</tr>
<tr>
<td>Teachers</td>
<td>2.6</td>
<td>1-5</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Average ratings of the second item of relevance on the questionnaire also indicated that the participants viewed the professional learning climate in the school to have improved. On the six-point scale asking whether the climate in the school focused on teachers as learners, the average ratings of the 114 teachers who rated this item on both occasions increased from 4.6 to 5.1. This increase was significant ($t(113) = 4.93, p<.01$).

Analysis of Meeting Activities
An overall comparison is provided for the frequency of various meeting activities for participating schools during the initial and two end-of-year meetings (Table 10.2). The analysis includes the number of meetings in which the identified activities were evident and the average percentage of time spent in the activity for those meetings in which the activity took place.
### Table 10.2
**Average percentage of time spent in different activities in the observed meetings**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Initial observation (N=6)</th>
<th>End of Year 1 observations (N=7)</th>
<th>End of Year 2 observations (N=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N. meetings Av. % of time</td>
<td>N. meetings Av % of time</td>
<td>N. meetings Av % of time</td>
</tr>
<tr>
<td>Neutral descriptions of data</td>
<td>4 34</td>
<td>4 28</td>
<td></td>
</tr>
<tr>
<td>Analysis of data / moderation</td>
<td>4 28</td>
<td>5 35</td>
<td>20</td>
</tr>
<tr>
<td>Identifying groups / target students</td>
<td>4 31</td>
<td>2 12</td>
<td>1</td>
</tr>
<tr>
<td>Descriptions of resources</td>
<td>2 35</td>
<td>1 9</td>
<td>56</td>
</tr>
<tr>
<td>Descriptions of teaching programmes</td>
<td>2 28</td>
<td>2 33</td>
<td>1</td>
</tr>
<tr>
<td>Organizational (teaching related)</td>
<td>3 9</td>
<td>1 9</td>
<td></td>
</tr>
<tr>
<td>Organizational (for assessment)</td>
<td>2 32</td>
<td>3 22</td>
<td></td>
</tr>
<tr>
<td>Explanations for low test scores</td>
<td>2 12</td>
<td>2 13</td>
<td></td>
</tr>
<tr>
<td>Possible evaluations of students (checklists)</td>
<td>1 20</td>
<td>1 7</td>
<td></td>
</tr>
<tr>
<td>Self-identified teaching problems</td>
<td>1 7</td>
<td>1 11</td>
<td>89</td>
</tr>
<tr>
<td>Positive evaluations of previous year’s programme</td>
<td>1 6</td>
<td>3 17</td>
<td>1 7</td>
</tr>
<tr>
<td>Problems with data or test</td>
<td></td>
<td>4 11</td>
<td>1 15</td>
</tr>
<tr>
<td>Implications of data for teaching</td>
<td></td>
<td></td>
<td>1 19</td>
</tr>
<tr>
<td>Discussion target students’ progress</td>
<td></td>
<td></td>
<td>2 20</td>
</tr>
<tr>
<td>Reflections on why programme successful</td>
<td></td>
<td></td>
<td>1 20</td>
</tr>
</tbody>
</table>

**Changes in Meeting Activities**

Most of the changes evident between the first and second meetings during the first year could be explained by the timing of the meetings. Overall, less time was spent putting students in groups and more time was spent analysing the data in the second meeting. Much less time was spent describing resources at the end of the year.

Although more time was spent analysing and discussing the data in the later meetings, it appears that the majority of teachers and literacy leaders did not make the link between the results and their implications
related to teaching practice during the meetings. There was little discussion about which teaching practices or strategies to employ to improve the achievement of the students, in particular those in the target groups. The illustrative quotes in the method section are indicative of this issue. There was also little discussion about how to improve the teaching practices of the teachers whose classes were not succeeding. When the data indicated that some classes were not achieving as well as others, reasons such as more ESOL students (or a different type of ESOL), or more difficult students, were given to explain this discrepancy, or it was ignored.

Meetings observations were undertaken at the end of the second year for only three schools so few claims about changes over the time can be made. The pattern of activity in two of these schools was very similar to the patterns of activities evident during the first year. Resources for the rest of the year were described or self-identified teaching problems posed for teachers to discuss. In the first school, most of the time involved a teacher, whose turn it was to identify a teaching problem receiving suggestions from others about how to help students decode single vowel sounds in multi-syllabic words. No reference was made to data to determine the extent students were experiencing problems with this particular skill or how they would check if the strategies had been successful.

In the second school most of the time was spent describing resources. However, some differences from the first year were beginning to emerge to the extent that the usefulness of these resources for target students was a constant reference point. Another difference was that 15% of the meeting involved discussion of some very general implications for teaching of the achievement data that had been presented to them the previous day by the school’s facilitator. The time of the year (December) may have accounted for the general nature of the implications discussed. For example, after noting that one student was still struggling, the literacy leader noted, “We need to look at kids like him very carefully next year.”

The meeting in the third school, from which the illustrative quotes in the method section are taken, spent much more time focused on evidence-informed reflections and implications of the data for their teaching. Strong elements of a professional learning community were evident in this school.

Although it is difficult to make many judgements about progress in the development of professional learning communities from the small number of observations, it appears to be an area that needs close attention by facilitators. What happens in schools when facilitators are not present is an important element if professional learning is to be promoted between visits and after the contract has finished.

Participants’ reflections on the meetings
In the follow-up interviews, participants were asked about the purpose of the meeting and to rate whether that purpose had been met on a 1-6 scale with 1 representing “Not met at all” and 6 representing “Fully met”. Purposes varied as would be expected, and overall, when ratings were provided they were relatively high. Criteria for assigning ratings, however, often did not relate to the original purpose. Rather, the reasons for ratings typically focused on whether the leader believed the teachers were positive or the process was collaborative. Figure 10.1 describes the literacy leaders’ main purpose, ratings of the extent to which the purpose was met and reasons for satisfaction. Unfortunately, ratings or reasons for them were not always provided.
Figure 10.1
Purpose of individual meetings, rating\(^1\) of extent to which purpose met and reasons for satisfaction with meeting.

<table>
<thead>
<tr>
<th>Initial meetings</th>
<th>Rating</th>
<th>Reasons for rating</th>
<th>End of first year meeting</th>
<th>Rating</th>
<th>Reasons for rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide an assessment overview for term so all were on “the same playing field”.</td>
<td>6</td>
<td>Not asked</td>
<td>Analyse data and check whether target groups had improved</td>
<td>6</td>
<td>The positive way the discussion happened</td>
</tr>
<tr>
<td>Collating the children in the target group</td>
<td>6</td>
<td>Happy they looked closely at their data</td>
<td>Summarise the data from the teachers’ point of view</td>
<td>6</td>
<td>An acceptance that any contribution [from teachers] is valued.</td>
</tr>
<tr>
<td>Make teachers aware of what “critical” and “at risk” mean so they can do something about it.</td>
<td>Not asked but no teaching implications discussed</td>
<td>Look at trends to see if detected a problem with comprehension</td>
<td>4</td>
<td>Some teachers did not have data for meeting</td>
<td></td>
</tr>
<tr>
<td>To get consistency with running records</td>
<td>Not asked</td>
<td></td>
<td>To see whether “we’d made a difference” particular with vocabulary</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>To get planning out to people and get them on board with it.</td>
<td>6</td>
<td></td>
<td>To talk about specific children and ensure teachers understood purpose of STAR</td>
<td>4</td>
<td>Provided opportunity for teachers to talk, but could have been more talk.</td>
</tr>
<tr>
<td>No initial meeting analysed</td>
<td></td>
<td></td>
<td>Compare results with those from the University</td>
<td>5</td>
<td>Teachers think they are better teachers of writing.</td>
</tr>
</tbody>
</table>

\(^1\) Note: A 1-6 scale was used with 1 representing “Not met at all” and 6 representing “Fully met”.
Leadership Qualities
After the initial meeting, literacy leaders were also asked about the leadership qualities that were important to them in such meetings. Understandably, their responses were similar to their criteria for judging meetings as successful, and focused very much on clarity of message, organisation and process, rather than any reference to improving knowledge or promoting professional learning. The leadership qualities described included:

- Making sure everything was covered and everyone could do what was expected and that the leader listened;
- To be well prepared, clear in the direction they were going and ensure good pacing, timing and receptiveness to responses;
- Being clear about what I’m doing and getting the message through;
- Listening, encouraging dialogue or encouraging open communication so that people don’t feel a failure sort of thing;
- I want them to own it [unit plans], and I want them to feel confident and excited about teaching it.

Unfortunately, these questions about leadership were not asked in subsequent phases of the research, however, these responses are included in this report as a possible issue that needs to be addressed specifically if literacy leaders are to see themselves as pivotal in promoting teacher learning.

Discussion and Conclusions
Participants’ ratings of the extent to which meetings were helping them to teach their target group children and that the school had developed a more focused professional learning climate, showed a significant increase. Indeed, initial ratings of the usefulness of meetings with regard to these kinds of purposes were surprisingly low. Clearly, meetings were not perceived to be useful at these times.

Comparisons between the first, second and third meetings using the observations and interviews were difficult to make because of the small number of schools overall, school attrition by the third meeting and irregularity of the follow-up interviews. However, the meetings appeared to be focused on organisational and resourcing issues, with data descriptions rather than use, predominating. While it is important to address organisational issues because good management is fundamental to a well-run school, only one of the analysed meetings exhibited the qualities of professional learning communities that are associated with improving student outcomes. These meetings focus on the teaching–learning–outcomes links. Participation and valuing of all teachers’ contributions was given greater weight in most meetings than focused analysis of the teaching–learning relationship.

It may be important for facilitators to check whether schools continue with activities focused on professional learning when they are not present. Intermittent visits from facilitators cannot be considered by the schools as a sufficient support for professional learning, because teachers need multiple opportunities to learn and try new aspects of practice. In addition, dependence on facilitators’ presence is not likely to promote sustainability once the contract has finished. The above analysis indicates two key areas may need to be addressed for this ongoing learning to occur. The first is that leaders have a vision of themselves as leaders
of professional learning. The second is that the teaching implications of data, both in terms of reflection and planning, are addressed explicitly.

References


Chapter 11

An Analysis of Professional Learning Episodes

The purpose of this chapter is to refine our understanding of the qualities of professional development episodes that promote teaching learning. It is not intended to promote one kind of activity as more effective than another, because a key principle of learning is that repeated opportunities over a number of different kinds of learning episodes are needed to develop deep understanding. This level of understanding rarely comes from one-off events or repeated exposures to only one type of activity.

A second reason for not advocating a particular kind of learning activity is that teachers are not a homogeneous group and are unlikely to find the same opportunities equally effective. Different exposure to prior learning and sophistication of understandings will lead to the same learning “episode” being interpreted differently by participating individuals. What this analysis is intended to do is to consider how to maximize the effectiveness of particular kinds of learning opportunities.

A dilemma arose when deciding how to structure this aspect of the research. In a professional development situation, learning is not “event specific” but rather occurs over a period of repeated opportunities. In the early stages of this project, the approach of interviewing teachers after participating in a series of events was trialled but abandoned because the teachers’ responses were too general to identify the more or less powerful episodes or their key features. Memories typically became intermixed. It was found that interviewing after a single event allowed teachers to be much more specific about what they had or had not learned during that time and what needed to happen next for them to be successful in putting what was learned into practice. To some extent, therefore, the methodological approach had serious limitations in terms of determining teacher learning over an extended period of time and is not intended to portray either the totality or intensity of the professional learning that occurred throughout the Literacy Professional Development Project. On the other hand, this analysis uncovered some very useful information on how to improve the facilitation of teacher learning.

The Framework Guiding the Analysis

The framework used for the analysis of the episodes was guided by theories of learning. They focused on the content to be deepened and the skills refined, together with the learning processes.

The Content to be Deepened and Skills to be Extended

The content of the professional development is assumed to provide teachers with the resources to deepen professional understandings and extend relevant skills. It was assumed that the participating teachers already had a wide range of professional knowledge and skills, because this project is concerned with promoting the learning of practicing teachers who, at a minimum were provisionally registered. Many were highly experienced.
The understandings and skills used in the analysis were identified from a systematic synthesis of studies for the Best Evidence Synthesis on Professional Learning and Development (Timperley, Wilson, Barrar & Fung, 2007) that were also supported by the theoretical literature on how teachers learn. They fell into three broad but not exclusive groups. The first group, the largest of the three, included understandings that could be used to inform practice, or the knowledge to teach. Specific content within this group included engagement with the fundamentals and interrelationships of teaching, including curriculum, pedagogical and assessment knowledge, and in some cases, an awareness of standards of teaching practices in particular curriculum areas. Also included in this group was an understanding of theoretical frameworks and tools within which the above content could be situated. Other categories in this group included knowledge of students, their expected developmental progressions in a particular curriculum area and how they learn. Teachers’ social constructions of students and how that influenced their teaching practice were also included.

Understandings are of little use to teachers facing a class of students each day, unless they can be linked to practice. It is not surprising, therefore, that a second group of studies made reference to providing resources directly related to an individual teacher’s practice. These references included an analysis of teachers’ own practice. Other studies also included an analysis of how that practice impacted on their diverse student learners. Both involved a vision of new possibilities for practice and its impact.

The third group included the development of methods of inquiry that challenged teachers’ taken for granted assumptions about the effectiveness of their practice and provided opportunities for them to learn the skills of inquiry into how to improve in an ongoing way. These studies focused more systematically on building the meta-cognitive processes of self-regulated learning and were considered particularly important because they potentially allow teachers to take control of their learning in systematic and evidence-based ways.

**Learning Processes**

When identifying the learning processes, we assumed that adult learning is fundamentally similar to that of students (Donovan, Bransford, & Pellegrino, 1999). In making this assumption, it is not intended to discount the obvious differences between adult and student learning situations, such as the richer life experiences on which adults draw, the learning contexts in which they occur and the greater demand adults place on the relevance of learning in order to engage. We identified four learning processes that were important in explaining particular outcomes. The first process involved cueing and retrieving prior knowledge with the outcome of consolidating it, or alternatively examining it for its adequacy. Cueing prior knowledge was unlikely to substantively change practice, but if it resulted in examination of its adequacy then this could form the basis for change. This process is sometimes considered equivalent to engaging teachers’ personal theories underpinning their practice (Robinson, 1993) so that any new practices can be considered in light of their congruence with existing practice.

The second process involved becoming aware of new information or skills and integrating them into current values and beliefs systems with the outcome that new knowledge is adopted or adapted. This process could involve a very superficial level of acquisition of new knowledge or much deeper learning. Many professional development situations are based on the idea that teachers needed to learn “new things” rather than to examine how these new things related to existing knowledge. The problem of over-assimilation (Hammerness et al., 2005) where teachers believe they are enacting
new practice but it only represents this practice in superficial ways, however, is well documented (Firestone, 2004; Spillane, 2000).

The third process involved creating dissonance with current beliefs, attitudes or knowledge of effective practice. Some authors advocate that such dissonance is a requirement of substantive change because teachers have to unlearn much of what they believe, know and know how to do (Ball, 1988) but it also runs the risk of rejection of key messages as teachers dismiss new possibilities as impossible in their situation (Coburn, 2001). Resolution of dissonance may, therefore, involve reconstructing alternative current values, beliefs and knowledge being reconstructed in ways consistent with the change messages, or rejecting the messages altogether.

The final process was based on the literature of co- and self-regulation (Butler & Winne, 1995) and examined the extent to which the professional learning opportunities promoted processes of inquiry into the adequacy and improvement of teaching practice. Although there are many theoretical approaches to self-regulation, we adopted a position consistent with that of Butler and Winne in which self-regulated learners are those who, “… judge performance relative to goals, generate internal feedback about amounts and rates of progress towards goals, and adjust further action based on that feedback” (p. 258). It is, in their view, a deliberate, judgmental, adaptive process.

Reactions/Actions of Diverse Teacher Learners

The above processes can result in a wide range of reactions that are portrayed in Figure 11.1. The reactions used to categorize teachers’ responses in the learning episodes ranged from ignoring or rejecting new theories and practices, to actively engaging with new ideas and applying them to new learning situations and/or enhancing self- and others’-regulated learning. Those reactions least likely to result in changed practice are at the top of Figure 11.1 with more desirable reactions appearing towards to bottom of the figure.

Figure 11.1
Reactions/Actions of Diverse Teacher Learners

<table>
<thead>
<tr>
<th>Reactions/Actions of Diverse Teacher Learners/communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment &amp; interpretation of the relevance, usefulness, cost / benefit result in one or more of the following:</td>
</tr>
<tr>
<td>• reject / ignore new theory and practice and continue with prior practice</td>
</tr>
<tr>
<td>• would like to implement new practice but have insufficient skills</td>
</tr>
<tr>
<td>• believe enacting new practice but continue with prior practice</td>
</tr>
<tr>
<td>• select parts of new theory and practice and adapt to current practice</td>
</tr>
<tr>
<td>• implement as required</td>
</tr>
<tr>
<td>• actively engage with, own &amp; apply new theory &amp; practice and change practice substantively</td>
</tr>
<tr>
<td>• enhanced regulation of own and others’ learning</td>
</tr>
</tbody>
</table>
Method

The learning processes and qualities of Figure 11.1 were itemised on a spreadsheet and 18 episodes of teacher learning were analysed for each of the qualities. The first 11 episodes were analysed separately by one of the principal researchers and a research assistant. High levels of agreement were achieved between the two coders with the small number of disagreements resolved before the final coding was entered. The final seven episodes were analysed by the principal researcher only.

The analysed episodes were constructed events facilitated to promote teacher learning. Some episodes were facilitated by a project facilitator; others were facilitated by literacy leaders in the schools. All these personnel have been identified as “facilitators” because the purpose of the analysis is not to focus on any particular group of people in particular positions, but rather to identify which kinds of episodes appeared to promote teacher learning. In all, six different facilitators and literacy leaders were involved, all of whom are considered highly competent.

Episodes were nominated by the participating schools as suitable for analysis of teacher learning. They were not pre-selected by the researchers. The main episodes involved the facilitator observing practice and receiving feedback, and meetings or workshops designed to promote teacher learning. Two episodes involved what is commonly referred to as “modelling”, that is, they involved the teacher in observing the facilitator and / or other teachers teach. Table 11.1 identifies the number of different types of episodes included in the analysis.

Table 11.1
Types of events analysis

<table>
<thead>
<tr>
<th>Type of event</th>
<th>No. of episodes</th>
<th>No. of teachers interviewed</th>
<th>No. of facilitators (including literacy leaders)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation &amp; feedback</td>
<td>9</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Syndicate meeting</td>
<td>4</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Whole staff meetings</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Observing teaching of own class</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>and others’ classes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Twelve different teachers participated in the episodes. Some participated in several episodes that are identified in Table 11.2. They came from five different schools and taught a range of year levels from Year 0/1 to Year 7/8. Most were in the second year of the project, although three were in the first year. They were evenly split between having reading and writing as their project focus.
### Table 11.2
**Participating teachers, year of project and curriculum focus**

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Type of incidents in which engaged (episode reference in brackets)</th>
<th>Year of project</th>
<th>Curriculum focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>• Three syndicate meetings (1,4,6) • Observation and feedback (2) • Whole staff meeting (5)</td>
<td>Year 2</td>
<td>Writing</td>
</tr>
<tr>
<td>B.</td>
<td>• Three syndicate meetings (1,4,6) • Observation and feedback (3) • Whole staff meeting (5)</td>
<td>Year 2</td>
<td>Writing</td>
</tr>
<tr>
<td>C.</td>
<td>• Three syndicate meetings (1,4,6) • Whole staff meeting (5)</td>
<td>Year 2</td>
<td>Writing</td>
</tr>
<tr>
<td>D.</td>
<td>• Syndicate meeting (7) • Observation and feedback (8)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>E.</td>
<td>• Syndicate meeting (7) • Observation and feedback (9)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>F.</td>
<td>• Syndicate meeting (7)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>G.</td>
<td>• Syndicate meeting (7)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>F.</td>
<td>• Observation and feedback (10)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>G.</td>
<td>• Observation and feedback (11)</td>
<td>Year 2</td>
<td>Reading</td>
</tr>
<tr>
<td>H.</td>
<td>• Whole staff meeting (12) • Observation and feedback (13) • Facilitator modelling (14) • Observing another teacher teaching (15)</td>
<td>Year 1</td>
<td>Writing</td>
</tr>
<tr>
<td>I.</td>
<td>• Observation and feedback (16) • Whole staff meeting (17)</td>
<td>Year 1</td>
<td>Writing</td>
</tr>
<tr>
<td>J.</td>
<td>• Observation and feedback (18) • Whole staff meeting (17)</td>
<td>Year 1</td>
<td>Writing</td>
</tr>
</tbody>
</table>

All observation and feedback episodes were audio-taped and transcribed. Field notes were taken of most of the workshops / meetings. The meeting involving teachers I and J were audio-taped. The data source for each episode is identified in Figure 11.4. The episode numbering is ordered according to the sequence of events within a particular school. Therefore, the different types of episodes are intermixed.
Table 11.2
Data source for episode

<table>
<thead>
<tr>
<th>Episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
</tbody>
</table>

The limitations related to generalizing from the analysis of individual episodes are mentioned above. Other limitations that need to be noted are that the facilitation practice described cannot be considered representative of the style or impact of any particular facilitator over time or of all facilitators or literacy leaders involved in the project. In addition, these episodes are too brief to relate them to improvements in student achievement. What the approach is designed to identify are the qualities of teacher learning episodes that are differentially effective in promoting teacher learning from the teachers’ perspectives. To this end, criteria for effectiveness were established against which to judge quality. The first was the engagement of teachers existing theories and the co-construction of meaning of any critique of those theories and the practice on which they were based. The second was that the foundations for ongoing improvement were laid through the development of self-regulation, that is, teachers had goals to work towards, had ways to monitor progress towards those goals and an indication of sufficient understanding of how to change practice to make
the necessary adjustments. The third criterion was the teachers’ reactions to the episode, in particular whether they indicated that they had changed or intended to change practice. This criterion was used as an independent check on whether the processes on which the analysis was focused related to teacher reactions.

Results
The analyses of three main types of episodes are presented in the results section. Ten of these episodes included what is generally referred to as observations and feedback (with and without student interviews). Seven of the episodes involved meetings and workshops. Two of the episodes involved a teacher observing modelled practice. Each type of episode is analysed separately, with overall conclusions presented at the end of the chapter.

Observations and Feedback
Nine observation and feedback episodes were analysed. An overview of the episodes that included student interviews are summarized in Table 11.3 and those without student interviews are summarized in Table 11.4. Appendix D has a more detailed summary of both types of episodes. In both tables, qualities related to the first criterion of principled content, co-construction of meanings and future actions are noted in columns 2 and 3 under engaged with T. theories / co-construction (Column 2) and content/activities (column 3). The second criterion of self-regulatory processes involving goals, monitoring progress and understanding how to change in relation to those goals are addressed in column 4, under self-regulatory learning (SRL) processes. Teachers’ learning processes, their actions and reactions are described in column 5.
<table>
<thead>
<tr>
<th>Ep</th>
<th>Engaged with T. theories / co-construction</th>
<th>Content / activities</th>
<th>SRL 7 Processes</th>
<th>Learning processes, reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
</table>
| 2   | Limited engagement; limited co-construction | • Used check list of effective practices  
• Students’ responses - need reasons for activities  
• Limited mention of principles | Goals non-specific. No evidence of other SRL processes | T. gained understanding of issue of students not understanding reasons for activities. Indicated changed practice. No indication of enhanced SRL. | Most salient learning - student interviews but T. wanted to observe someone implementing particular aspects of practice. |
| 3   | Limited theory engagement, limited co-construction | • Used checklist of effective practices  
• Summarised student responses – no evaluative comment  
• Limited mention of principles | Goals non-specific. No evidence of other SRL processes | T. could not see how to implement advice. | T. more positive than F. about lesson. T. did not accept validity of student interviews. Prior negative T. reaction led F. to be non-specific. |
| 13  | Extensive theory engagement but not in sufficient depth. Co-construction attempted but T. difficult to engage. | • Variety of teaching strategies to promote confidence  
• Principles explicit | T. not asked about goals. No evidence of other SRL processes | No change in practice. T. believed children learn and become interested when ready. Did not know where to start. | Combination of problematic beliefs and limited skills meant could not follow advice. |
| 16  | In-depth theory engagement and extensive co-construction | • Understanding of impact of teaching on student understanding  
• Pedagogical content knowledge  
• Theoretical frameworks | Goals non-specific. Evidence of other SRL processes promoted. | Created dissonance with existing beliefs about student understanding. Highly motivated to change practice. Created ways for self-regulation of own learning | Student interviews created motivation to learn and change. Co-construction led to good understanding. |
| 18  | In depth theory engagement and extensive co-construction | • Understanding of impact of teaching on student understanding  
• Pedagogical content knowledge  
• Theoretical frameworks | Goals non-specific. Evidence of other SRL processes promoted. | Created dissonance with existing beliefs about student understanding. Highly motivated to change practice. Created ways to self-regulate own learning | Student interviews created motivation to learn and change. Co-construction led to good understanding. |

6 Episode analysed  
7 Self regulated learning
Table 11.4
Summary of Observation and Feedback Episodes without Student Interviews

<table>
<thead>
<tr>
<th>E.</th>
<th>Engaged with T. theories / co-construction</th>
<th>Content / activities</th>
<th>SRL Processes</th>
<th>Learning processes, reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>No evidence of this</td>
<td>Brief feedback on lesson. Teaching strategies provided but unrelated to observed lesson or teacher requests. Reasons given but not in principled framework.</td>
<td>Goals not shared. No evidence of other SRL processes.</td>
<td>T. stated did not get knowledge or strategies looking for.</td>
<td>F. appeared to have pre-determined skills and knowledge wanted to cover rather than observed lesson.</td>
</tr>
<tr>
<td>9</td>
<td>No evidence of this</td>
<td>Limited feedback on lesson Teaching strategies provided – unrelated to observed lesson. Reasons given but not in principled framework.</td>
<td>Goals not shared. No evidence of other SRL processes.</td>
<td>T. indicated had gained new knowledge but areas not mentioned in feedback. Has no idea of next steps.</td>
<td>T. indicated wants to learn – but found experience did not promote this.</td>
</tr>
<tr>
<td>10</td>
<td>Limited engagement no evidence of co-construction</td>
<td>Pedagogical and behaviour management advice embedded in questions.</td>
<td>Goals non-specific. No evidence of other SRL processes.</td>
<td>T. rejected implied criticism. Rarely teaches reading so cannot see when would implement advice.</td>
<td>F. and literacy leaders’ serious concerns about her management and pedagogy not clearly conveyed or understood</td>
</tr>
</tbody>
</table>

Analysis of observation and feedback as a vehicle for teacher learning

These eight episodes of observation and feedback showed a variety of learning processes and teacher reactions. Those episodes that met the criteria of providing opportunities for theory engagement and developing self-regulatory processes were accompanied by teachers reporting either that they had changed their practice or that they intended to change. These episodes had different qualities from those that did not meet the criteria.

Substantive learning appeared to be the case in episodes 16 and 18. The catalyst for the learning was an understanding in general that student achievement at the school was problematic and specific information from the student interviews, that they did not understand the intended learning focus for the lesson or how to be successful in achieving it. Co-construction of the meaning of these student interviews, in terms of the implications for practice, created sufficient dissonance for the teachers in terms of their current perspectives on the effectiveness of their teaching, that they were highly motivated to change their practice. The learning was at the level of principle and self-regulation. How the specifics of practice were to be addressed in subsequent sessions was negotiated during the feedback session.
Sufficient learning to change practice was also evident in episode 2 when a checklist was used to indicate desired practice and students’ responses to facilitator questions formed the basis of the feedback. Particularly salient for the teacher was the feedback that students did not understand the reasons for engaging in particular activities. This teacher indicated that she understood the need to provide such reasons. She also indicated, however, that in order to put all the facilitator’s suggestions into practice, she needed to observe someone implementing particular aspects of practice.

In all other episodes, the teachers involved indicated that the observations and feedback in themselves were insufficient to change practice or that they had no intention to change. The teacher in episode 9 was reasonably positive in the sense that she wanted to improve her practice, but when asked if she knew what to do following the feedback she replied, “No, I have no idea really at all”.

A problem mentioned by two teachers was that they needed to observe modelled practice in order to understand more fully what the changes advocated by the facilitator looked like (episodes 8, 11). As the teacher in episode 11 explained,

“She [the facilitator] highlighted a couple of points where I could look at improving … more to do with asking questions … and how to take it a step further. She is going to do some demonstration lessons so I can see more about what she was talking about…. I mean I was interviewed two weeks ago and it’s week seven now and I don’t know actually when I’m due to observe her…. I mean this whole term will go by before I have really got a lot of meat out of it. I would make it my priority in the classroom…. All I know at this stage is I am due to wait for her to come back and me to observe her.”

In other episodes the problem appeared to rest with the teachers disagreeing with the facilitator’s implied negative judgments of their teaching and the different views of effectiveness (episodes 3, 10), or the kind of knowledge they valued was not provided (episode 8). Another teacher in episode 13 indicated that she had insufficient skills to use the suggestions offered. In her interview with the researcher, she explained, “I just don’t know where to start really”.

Further analysis of the episodes provide an exemplification of the conditions associated with more and less effective practice in relation to the activity involved in undertaking observations and providing feedback that lead to apparently differential opportunities for teachers to learn. The indications obtained from the evidence related to these episodes are well supported by theories of learning and are described below.

A touchstone against which to judge effectiveness of practice

The first of the qualities that distinguished effective episodes was that improvements in practice were referenced against an external criterion, beyond stated preferences by the facilitator or the teacher for particular practices. Student interviews provided this reference in the two most effective episodes (16 and 18). Student reports of superficial interpretation of, or confusion about, the learning aims of the lessons, what success looked like, and confused and superficial understandings about what their teachers told them to work on in their writing appeared to be a powerful way to engage teachers in promoting their own learning. During the session, further professional development was planned so that the teachers had a clear idea of what kind of support they would receive and what they could do to help themselves in the meantime. These student responses not only served as a motivator to change practice, but they also served as the criterion against
which to judge the success of improvement. In this way, the conditions for ongoing self-regulation of learning were established (Butler & Winne, 1995).

This same external reference was used for the next most effective learning episode (2). The teacher’s response in this episode, however, indicated that she did not have the skills to change practice in the ways needed to change the students’ understanding. In her interview with the researcher, she expressed the desire to see how the verbal advice could be implemented in practice. The follow-up may have occurred but was not indicated during the feedback.

The problem of skills to make the needed changes was even more evident in Episode 13. The teacher in this episode wanted her young students to be confident to write independently, but the students indicated in their interviews that they had few strategies to do so. Their main reported strategies when they could not spell a word, or ran out of ideas to write about were to “Ask the teacher” or “Don’t know”. The teacher was unable to capitalize on these reported difficulties, however, because she had insufficient skills to do so.

Episode 3 indicated another condition for student interviews to be the catalyst for change. In this episode the teacher did not accept the validity of the student interview data. She indicated in her follow-up interviews that she believed the facilitator had asked the wrong students about their understandings. Compounding the problem were the serious concerns about this teacher’s practice held by the facilitator and literacy leader and the teacher’s previous negative reactions to feedback. This prior history had led the facilitator to be non-specific and use a descriptive summary of the students’ responses, rather than put them in an evaluative frame. As a result, the meaning and implications of the students’ comments were not discussed. Such differences of opinion between teachers and facilitators, especially when they remain unexpressed, are difficult to resolve and how they might be approached is described in the conclusion to this section.

Sharing learning goals
Little can be concluded about the effectiveness of sharing learning goals in relation to these episodes, because in none of them was the teacher clear about specific learning goals underpinning the observations beyond a generalized “Snapshot of where I am in my teaching” or “To give us feedback on our teaching methods and suggest ways of improving it”. While obtaining such a snapshot may be appropriate at times, and three teachers were in the first year at the school, it could be expected for other teachers in the second year of the contract that learning goals would be more specific than these generalized statements if specific goals of teacher learning had underpinned the interactions.

Sharing learning goals is not intended to imply a unilateral decision on the part of the facilitator or the teacher but, rather, a process of negotiation using the understandings and expertise of both. While such negotiation may have occurred prior to the observation, the teachers’ responses to interviewer questions about such goals indicated that they were not sufficiently salient to remember them. Professional development goals are not always shared explicitly, with providers’ goals for improvements in practice, not necessarily shared with teachers. The absence of shared learning goals is likely to lead to a lack of focus of such learning episodes, leaving teachers, like their students (see chapter 5) unable to monitor whether they have learned worthwhile skills and knowledge. If we are to develop the notion of professional development being about promoting ongoing learning, then learning goals need to be more explicit.
Engaging and bypassing teachers’ personal theories

Deep engagement of teachers’ personal theories about teaching students and the conditions that promote their learning occurred in only two episodes (16 & 18). In a subsequent interview, the facilitator indicated that ascertaining these beliefs was very difficult. Given the importance of beliefs in determining practice, this difficulty, together with the absence of attempts to do so in other episodes raises some concern.

In a third episode (13), the facilitator attempted engagement but was able to achieve this in a superficial way only. As noted above, the teacher wanted the children to have the confidence to write, but the student interviews indicated that the students did not have sufficient strategies to do this. While theories around specific teaching strategies were engaged with during the feedback, the deeper theory that needed to be challenged was that the teacher did not believe she had an active role in teaching those strategies. As she described in her interview,

Teacher: … children develop when they are ready… it will just come to them… It seems to take quite a few boys a whole year before they are really interested and before they start.

Interviewer: And you were waiting for them to become interested?

Teacher: Yes, I suppose I was.

The pervasiveness and undesirable consequences of “readiness” theories in New Zealand education are well described by McNaughton, Phillips and MacDonald (2000). These theories can lead to those students who have had the necessary prior experience to read and write being taught to build on their previous knowledge, while others who have not had this prior experience continue to be disadvantaged while their teachers wait for them to be ready.

Considerable challenges are associated with theory engagement and the response of this teacher indicates the need to keep probing to find out what level is needed to engage. It is unlikely to happen in a single episode and subsequent facilitator interactions did achieve this outcome.

In other episodes (2, 3, and 10), facilitators also attempted to ascertain the teachers’ beliefs underpinning particular practices, but the teachers’ descriptions of their beliefs were brief and superficial and were not further engaged in or discussed with them. After a brief question / answer sequence, the conversation typically moved to working through the checklist without further checking the teachers’ understandings. It is apparent that engaging teachers’ beliefs is a difficult process. Yet, it is widely acknowledged that practice is based on such beliefs and without achieving a substantive shift, any new practice is so strongly influenced by old beliefs, that change is more superficial than real (Darling-Hammond, Bransford, & LePage, 2005).

Joint re-construction of practice

Theory engagement, co-construction of meaning and implications for practice, and jointly re-constructing practice appear to be closely linked, because either they all occurred together, or were absent from a particular episode. In those episodes seen to be most effective by the teachers (16 and 18), the meaning of the student responses, how they related to current practice and the implications for future practice were co-constructed by teachers and facilitator. In the feedback, the facilitator systematically provided descriptions of
teaching practice, checked the accuracy of these descriptions with the teachers, provided evaluative criteria against which the effectiveness of practice could be assessed, and probed for the beliefs on which their practice was based. Through this kind of engagement the facilitator was able to ascertain whether important meanings were shared with the teachers. At the end of the session, the facilitator then worked with the teachers to understand what needed to change and established priorities for future work.

It is of concern that similar examples of co-construction were not evident in the other episodes. Feedback tended to take the form of questions on the part of the facilitator, to which the teachers responded without further engagement with the answers, or the facilitator suggested new strategies that the teachers should/could use. Rarely was the usefulness of this advice systematically checked or developed jointly with the teacher. Often these questioning and telling styles of feedback occurred in the same session.

Knowledge and skills developed through observation and feedback

Most professional development descriptions focus on the knowledge and skills offered to teachers or on the activities in which they engage. Given that the activity is “observation and feedback” this section focuses on the content of the learning.

The content discussed in these observation and feedback episodes typically involved descriptions of teaching strategies, rather than specific content knowledge. This emphasis probably reflects the nature of the learning episode because these situations are based on the enactment of classroom strategies and practices. However, if the underpinning knowledge and principles have been provided in other situations, such as workshops, how this earlier knowledge relates to practice can be assisted by reference back to it in these more practice-based situations. Rarely was reference made to previously delivered content in these analysed episodes. There may be reasons specific to the episodes for this absence. In two episodes (16 and 18) these formed the first contact with the school. In all other episodes, except episode 13, however, the contract was in its second year, so it could be expected that other, more formal “content-focused” workshops would have taken place. A reasonable assumption could be that the content of the checklist used in episodes 2 and 3 had been discussed with the teachers prior to the observed episode but no specific reference was made to earlier coverage during the recorded feedback.

None of the teachers cited disagreement about the worth of the curriculum or pedagogical content advocated as the reason for not acting on the feedback, so a close analysis has not been undertaken of this aspect of the interactions. Rather, the focus of the analysis is on what could have been expected from this type of learning episode in terms of new understandings and what resulted.

Understanding own practice in relation to a standard of practice

A reasonable expectation of engaging in observations and feedback could be that teachers develop an understanding of the effectiveness of their practice against some ideas of desirable standards of practice or student outcomes. The development of this understanding, however, was inconsistent in these episodes for a variety of reasons. In episodes 16 and 18, the most effective episodes, information from the students’ responses during the interviews led the teachers to understand the areas of their practice that were problematic and to want to know more about effective strategies to develop improved student understanding.
In episode 11, the feedback mostly confirmed teacher practice, although no criteria or explicit standards were used to justify the positive evaluation. Some specific suggestions were made for improvement but like other teachers, this teacher needed to see what the changes looked like through observing others.

The use of a checklist of preferred practice and responses from students in episodes 2 and 3 provided a framework for understanding current practice against some ideas of desirable practice. The teacher in episode 2 indicated that the exercise had been useful and had attempted to change her practice although she, like the teacher in episode 11, indicated that it would be useful to see such changes in action.

The use of the checklist was not as useful to the teacher in episode 3. One reason, as discussed above, was her rejection of the validity of the students’ interview responses. Another reason may have been the vague and non-specific descriptions provided by the facilitator during the feedback even though she and the literacy leader had serious concerns about this teacher’s practice. As described above, a history of negative reactions on the part of the teacher had led the facilitator to act in this way.

Disagreements between the teacher in episode 10 and the facilitator about appropriate standards of practice also underpinned problems with a decision not to act on the advice given. The facilitator and literacy leader had serious concerns about this teacher’s behaviour management and teaching practices. Specific concerns about noise levels were embedded in questions; advice was given, but not checked to see whether the advice was useful. In the follow-up interview, this teacher expressed the belief that the noise level indicated high levels of student engagement and that her teaching practices were effective, even though she was aware that they were not those advocated by the facilitator and literacy leader.

Reasons for not understanding how a particular teacher’s practice compared with a standard occurred for a different reason in episodes 8 and 9. In the feedback, the facilitator mentioned the observed lesson only briefly; rather she appeared to follow a pre-determined focus using pre-prepared resources. The teacher in episode 8 indicated that he did not find the suggestions useful because he wanted specific guidance about how to teach and observations of others of how to enact it. The other teacher believed that the advice was useful but had not received a written summary, so was unable to act on it. She requested a transcript of the feedback from the researchers to refresh her memory. At the time of the interview, her recall of the suggestions offered was unrelated to the recorded conversation with the facilitator.

Enhanced self-regulation of teachers’ learning
Another distinguishing quality associated with the more effective practice was the promotion of self-regulation. Improvements in practice are almost inevitably restricted to the particular skills and knowledge offered in the immediate teaching episode, if teachers do not have a clear goal to achieve and the skills to monitor progress towards that goal and the support to reach it.

In episodes 16 and 18 interviews with the teachers established that they had clear goals (improved student understanding of the intended learning), strategies for learning the necessary knowledge and skills for achieving those goals (identified readings, engagement with appropriate observation and feedback systems) and ways to monitor their progress towards them (ongoing interviews with students). The goal of improved student understanding of the learning intentions underpinning particular lessons and the skills to observe one
another and interview students, allowed these teachers to engage in ongoing improvement. Although not all of this was accomplished in the feedback session, the framework was established with the teachers so they had a clear idea how to construct their own learning opportunities. Indeed, they continued to do so following the session.

Conditions for promoting self-regulation were not particularly evident in the other episodes. The discussions were focused on the here and now of the lesson or a preferred practice constructed by the facilitator. Links to prior established goals or future goals and how their achievement might be monitored was not evident.

Conclusions: Observation and feedback as a way to promote teacher learning
Observation of individual teachers and providing feedback is an expensive and time-consuming process. It is important, therefore, that the usefulness of the process in promoting teacher learning is maximized. This analysis indicates that from the perspective of the majority of the teachers, the process did not realize its potential. It is possible that the activity was undertaken primarily for the facilitator to gain an understanding of where to focus other professional development activities, such as staff and syndicate workshops, and that the feedback to teachers was secondary to this aim. If this was the case, the question needs to be raised about the benefits to the teachers. If teachers are to co-construct their learning opportunities, have their theories about students and effective teaching strategies deepened and challenged, then they need to understand their strengths and what is problematic and a means to improve. When teachers have a vision of the principles, practices and outcomes that they are trying to achieve, observation and feedback should become an opportunity to learn about progress towards that vision, rather than being the recipient of someone else’s judgment against criteria they may or may not understand and the value of which may not be shared.

There was no evidence in the analysis of meetings and workshops (see next section) that explicit links were made between workshop activities, teacher’s individual learning goals and the feedback related to the observations. Giving feedback in ways that facilitate learning is a difficult skill, and one that needs to be further developed if the opportunities for teacher learning are to be optimised. One particularly difficult situation arises when teachers disagree with the judgments of facilitators about the effectiveness of practice. For this reason, a short section on how to resolve such differences is presented below.

Resolving differences of opinion
In two of the episodes (3 and 10) a difference in opinion about the quality of the lesson was evident between the facilitator and teacher in both episodes with the data from the student interviews contested by the teacher in episode 3. A similar situation arose in episode 8 when the teacher did not particularly value feedback from observations. His preference was structured input and the opportunity to observe others. In these three episodes, these differences contributed to the lack of change to practice reported by the teachers concerned. In the first two episodes, serious concerns were held by the facilitator and literacy leader, about the quality of teaching practice. The option of glossing over these differences and relying on input alone, therefore, was potentially detrimental to students.

One way to approach these differences is to surface them in a way that does not presume the correctness of either position and then establish a mutually agreed process against which to test each person’s assumptions. To do this, the assumptions need to be made explicit, rather than buried in questions. For example, in episode
when the teacher believed that the wrong students had been interviewed, it could have been arranged to ask different students similar questions to ascertain their understandings in a subsequent lesson. Criteria for student selection need to be established to avoid the bias of interviewing only the top-achieving students. The process of engaging in such a process may well shift the teacher’s practice in desirable ways to ensure student understanding prior to their interviews in this subsequent lesson and not be a “test” of the original lesson. In a professional learning situation, these changes would be desirable, so this should be of little concern. Similarly, the teacher in episode 8 may need to be challenged about his preferred learning activities that avoid any analysis of his own practice. The situation would need to be negotiated whereby he had the input he believed he needed, together with some way of checking whether that input was having the desired impact on his practice and the outcomes for his students.

**Syndicate and Whole Staff Meetings / Workshops**

This section analyses four syndicate meetings (episodes 1, 4, 6, 7) and three staff meetings (episodes 5, 12, 17) designed to promote professional learning. The key features are summarized in Table 11.5.
### Table 11.5
Summary of Meeting / Workshop Episodes

<table>
<thead>
<tr>
<th>E.</th>
<th>Engaged with T. theories / co-construction</th>
<th>Content provided / engaged</th>
<th>SRL Processes</th>
<th>Learning processes, reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Engaged and challenged at times</td>
<td>Using asTTLe data as formative assessment.</td>
<td>Goals: 2 T’s clear, 1 unclear. SRL promoted for the two.</td>
<td>2 T’s - new skills in data interpretation and linking to teaching. 1 T. – nothing new. Both actively engaged but T.2 did not see herself as a learner.</td>
<td>1 T. preference is to observe others rather than discussing student outcomes.</td>
</tr>
<tr>
<td>4</td>
<td>Limited</td>
<td>Using asTTLe data in one genre to inform teaching in another.</td>
<td>Goals initially unclear but sorted. SRL promoted.</td>
<td></td>
<td>Understandings finally coming together for T.2 after 18 months.</td>
</tr>
<tr>
<td>5</td>
<td>Limited</td>
<td>Various theoretical frameworks, linking formative assessment to teaching, building pedagogical content knowledge</td>
<td>Goals not shared. No indication of SRL.</td>
<td>Mixed and relatively negative. Cosmetic changes only.</td>
<td>Agenda needed to be renegotiated. T’s indicated needed to observe before could implement.</td>
</tr>
<tr>
<td>6</td>
<td>No evidence of this</td>
<td>Using asTTLe data for formative assessment.</td>
<td>Goals: clear statement by F. at meeting but T’s all unclear. Potential for self-regulated learning but not realized.</td>
<td>All T’s indicated some kind of resentment because of timing of episode.</td>
<td>Same activity as episode 1 but very different reaction – timing important.</td>
</tr>
<tr>
<td>7</td>
<td>3 T’s indicated engagement. 1 T believed bypassed.</td>
<td>Reading / writing links</td>
<td>Goals explicit and shared. SRL processes not explicitly addressed.</td>
<td>3 T’s indicated would change practice. 1 T. opposing viewpoint that did not change.</td>
<td>Opposing viewpoint T acknowledged resistance – wants to watch modelled lessons.</td>
</tr>
<tr>
<td>12</td>
<td>No evidence of this</td>
<td>Formative assessment in writing</td>
<td>Goals not shared. SRL for this teacher not promoted because did not engage.</td>
<td>Interviewed T. did not believe applied to her. Others were engaged.</td>
<td>Some T’s unable to learn from discussing issues about other’s practice.</td>
</tr>
<tr>
<td>17</td>
<td>Yes</td>
<td>asTTLe levels and implications for teaching</td>
<td>Goals shared. T.s had improved knowledge but not specific to SRL.</td>
<td>Both T’s found session useful for deepening understanding of features and purposes</td>
<td>Analysis not reflecting on quality of teaching. Did deepen understanding of asTTLe indicators.</td>
</tr>
</tbody>
</table>

Analysis of structured meetings / workshops as a vehicle for teacher learning
From the teachers’ perspective, it appears that some of these meetings were more effective in promoting their learning than others. Some similarities in the quality of effective episodes were evident in the analysis of observations and feedback episodes above. The same criteria were used to judge the effectiveness of these episodes, that is, they promoted opportunities for teachers to learn principled knowledge through engagement.
with teacher theories, the foundations were provided for self-regulated learning and teachers indicated that they had changed or intended to change their practice.

The teacher interviews established that two episodes (episodes 4 and 17) were effective for all teachers who participated. In episode 1, two of the three teachers engaged, but one did not. Episode 6 with similar content and structure and the same participants, on the other hand, was not effective for any of them. Episode 7 proved highly engaging for three of the four teachers, but one teacher was resistant, confused and believed that if she enacted the ideas promoted, “... in two years time I will be told that was all wrong and I’ll have to do it a different way again”. Episode 5 resulted in one teacher being highly positive but two of the others resented participating, believing they had learnt little. In episode 12 the teacher on whom we focused was disengaged but her other colleagues were highly engaged (see episodes 16 and 18 – observation and feedback).

This differential engagement of the teachers in the episodes underpins the importance of acknowledging the diversity of learning needs within any group of teachers and underlines the importance of being responsive to specific professional learning needs at any given time. When teaching students, responsiveness to diversity is central to success, and the same principle applies when promoting professional learning.

Shared learning goals
The differential teacher responses beg the question about what it might mean to be responsive to teachers’ learning needs. Part of responsiveness involves sharing learning goals. As indicated above, this sharing does not mean that facilitators unilaterally define the goals and tell teachers what the goals of the session comprise, nor that teachers define them independently of facilitator knowledge and expertise, rather it means negotiating learning goals throughout the episode and jointly deciding whether they are worth pursuing. Goals need to be those valued by both facilitator and teachers for them to play a role in success. An analysis of the learning episodes appeared to show some relationship between having shared goals and the level of teachers’ engagement, although this relationship was complex.

In one of the episodes that engaged all teachers (episode 4), there was some initial confusion about the learning goals between the teachers and the facilitator, but this confusion was discussed and both facilitator and teachers agreed that the goals for the session were “to use data from persuasive writing to plan a different genre”. Both interviewed teachers found this to be useful and worthwhile. In the second episode in which all teachers were engaged (episode 17), the learning goals came from the staff themselves. They decided they needed to deepen their understanding of the asTTle indicators in order to teach a new writing purpose more effectively. This decision was counter to the facilitator’s earlier advice but she welcomed their decision because, “They are theorising themselves why they want to do something”.

In episode 1 the two teachers who were engaged were clear about the goals “to link the data with next teaching steps” but the third teacher who was much less engaged was less clear and stated a strong preference for alternative kinds of support, such as being given detailed units of work, observing particular aspects of teaching in practice and “having a go”, provided she had the content knowledge. A similar pattern was evident in episode 7. The three engaged teachers and the facilitator shared the worth of the goal of understanding the “relationship between reading and writing”. The fourth teacher understood this goal but
rejected it as valid by reporting in her interview that she still taught reading, writing and word study separately. “So I’ve come from a completely opposing viewpoint to the complete and utter opposite and now I’m really confused about what I should be doing.”

In episode 6 in which none of the teachers engaged, the facilitator stated the learning goals clearly at the beginning of the session but, in the interviews, all the teachers rejected the relevance of these for that particular stage of the teaching cycle. Identifying target student progress in week 3 of a new unit of work was considered too soon to have assessed, taught and judged progress. This episode highlighted the importance of negotiating the relevance of goals at particular times because a similar activity earlier in the year had led to very high levels of engagement by two of the three teachers (episode 1). It cannot be assumed that a given activity at one point in time is equally effective at a later point in time.

A similar picture was evident in episode 5. The facilitator believed that teachers had requested the session as a follow-up from a previous session, but two of the three teachers believed it was largely a waste of time. In this case the problematic assumption was that identified learning needs at one point in time had remained static between sessions. Checking rather than assuming that goals decided earlier are still relevant is crucial to the process of ongoing negotiation.

Episode 12 involved a disengaged teacher; she was unclear about the learning goals and did not see the relevance of discussing teaching practice in the senior school to practice with her more junior students.

Engagement of personal theories
A second area considered in the differential engagement of the teachers was whether their theories were bypassed or engaged. It was difficult to assess the level of engagement because few of these episodes were audiotaped so the analysis relied on teacher and facilitator interviews. It appeared that compatibility of personal theories with those offered by the facilitator was a more important predictor than whether those theories were engaged. Those who were engaged typically provided accounts of the sessions that indicated high levels of compatibility between their personal theories with those offered during the session. Those who rejected the usefulness of the learning opportunities, did so either on the basis that the kind of support was not useful to them, that is they had different theories of how to promote professional learning (episode 1 and 6), held different theories about appropriate student pedagogy (episode 7, 8), or they simply did not see the relevance of what was offered (episode 12).

Differing theories about how to promote professional learning provide a considerable challenge. One theme evident in the observation and feedback analysis repeated itself in this analysis. Several teachers indicated that they needed to see what was proposed in the professional learning situation being taught in action (teachers A, C, F). These teachers were different from those reporting this need to see modelled practice in the earlier observation and feedback episode analysis, indicating that seeing something in practice may be an important complement to principles and descriptions in a variety of situations.

A second theme that was evident in the meetings / workshop analysis was that some teachers preferred to receive materials and instruction from the facilitator, rather than have their own practice or student outcomes analysed (Teachers B & E). Shifting the focus from practice and its outcomes to inputs on the part of a
facilitator essentially leaves their own teaching practice and its consequences unexamined. If professional development is to have a positive impact on student achievement, privatizing practice is likely to be counter-productive to achieving this goal. These teachers need to be challenged in ways that they can see the benefits of shifting from a purely input orientation to learning from examining the effectiveness of their own practice – a personally more risky exercise.

Another issue that was mentioned incidentally by one teacher (Teacher C) was that she did not see herself as a learner. When asked what she had learned, she replied, “I don’t know about learnt”, then continued to list a variety of knowledge and skills that was new to her, including identifying target groups, looking for language resources, transferring skills from different genre, testing children before teaching and increased content knowledge. Promoting professional learning can be difficult when teachers do not perceive themselves as learners.

Content and Activities
No particular content or professional learning activity appeared to be any better received than others in terms of promoting teacher engagement in the meeting / workshop format. This is not intended to mean that all content and activities are equally worthwhile, because it is well established that some content is much more likely to promote student learning than others (Alton-Lee, 2003) and some activities are more likely to promote self-regulated learning. Rather, the intended meaning of this statement is that teachers responded both positively and negatively to the relevance of particular content and the usefulness of particular activities within this meeting / workshop format. A better predictor of teacher engagement than particular content or activities was in the areas mentioned above; that the learning goals were shared and the teachers’ personal theories of learning (in relation to themselves and their students) were compatible with what was offered and how it was offered.

The request for observations of modelled practice indicates that teachers sometimes could not see how to translate “the talk” in a workshop or meeting into their classroom practice, has already been mentioned. Another theme that was evident for the three teachers in episodes 1 and 4 was their independent mention of the benefit of repeated opportunities to learn, to practice and to reflect. The teacher who had reacted negatively to a number of professional learning opportunities over the previous 18 months of the professional development, found episode 4 particularly useful. She described it as,

“I felt things are now falling into place easier .... To me, I struggled with what makes a level 2 argument and once I understood that, then it was easy enough to teach it and get the kids to learn the parts .... I think actually having talked about the marking key for asTTle understanding that then it leads to teaching because you know if they are at 2b this is what they should be doing in 2a, what to teach and that sort of thing. I found that a lot more useful now I can understand it .... Once you know what you are doing then you can actually teach the kids. That’s the big thing.”

Given this perspective, some of her earlier negativity could be better understood.

Conclusions from analysis of meetings / workshop as a vehicle for teacher learning
This analysis of the qualities of teacher learning episodes in the meeting / workshop format indicates that what is often promoted as effective or ineffective in promoting professional learning, may not be focusing on
the important dimensions. Effectiveness is often defined in terms of activities, such as analyzing student work or achievement, ensuring teachers engage in activities rather than listen passively, or have real cases to analyze. These activities may assist with teacher engagement, but they do not appear to support learning, as reported on by the teachers in this study. What was more predictive of engagement was the compatibility of learning goals between facilitator and teachers, and beliefs about appropriate learning activities.

It is well accepted that student learning is promoted through a variety of activities, but what is becoming increasingly clear from the research evidence, is that knowing what it is they are to learn, being supported to learn it through a variety of activities and having the skills to judge progress are central to developing deep and self-regulated learning. Only occasionally were these conditions provided for the participating teachers, yet if we were to provide similar conditions for teachers, then it is likely that their learning will also be enhanced. In the case of teachers it is appropriate to negotiate rather than provide learning goals and appropriate activities because, as adult learners, they are less likely than students to engage for reasons of compliance. Negotiation does not mean that teachers dictate the learning agenda, because as learners they need to be guided by those more expert and an avoidance of analysis of practice or student outcomes is likely to be to the detriment of students. On the other hand, it also means that those with more expertise do not dictate the learning agenda independently of teachers’ beliefs about appropriate pedagogies and professional learning opportunities.

Observing Others as a Way to Promote Teacher Learning

A number of teachers indicated in their interviews, related both to observations / feedback and meetings / workshops, that they needed to see the teaching practices described or discussed in action. Part of the reason for the frequency of these requests may be the limited co-construction of the meaning of suggested new or revised practices between facilitators and teachers. The instruction tended to be one-way, from facilitators to teachers. In the most successful episodes of observation and feedback (episodes 16 and 18) where meanings and implications were co-constructed throughout, none of the teachers requested modelling, nor was it offered, and teachers were successful in changing their practice.

From a learning perspective, if modelling is one of several opportunities to learn new knowledge and skills, it may well complement other professional learning activities for many teachers. Teachers need to know the principles underpinning the modelled practices, understand explicitly what particular aspect of practice was being modelled, the reason for demonstrating those aspects and the co-construction of meaning following the episode. In the only two episodes involving modelling (14 & 15) that were analyzed for this chapter, modelling was differentially effective. In episode 14, all the above conditions were present, except that the teacher did not fully understand the reason for particular practices being modelled. Her theories about students and teaching at that time were not sufficiently sophisticated to benefit from the demonstration. Episode 15 which involved her watching other teachers teaching children of the same age, as in her class, was highly successful. A description of these two episodes follows. They involved the same teacher who participated in episodes 12 and 13. She did not engage with either of these other professional learning opportunities. Her disengagement with the first occurred because she did not see the relevance of discussing strategies for older students when she was teaching Years 0/1. In the second episode the feedback did not fit with her strongly held “readiness” beliefs about student development that involved waiting until they were ready to learn to write and she did not have the skills to implement the suggested changes. In the follow-up interviews with the facilitator, she acknowledged that some teachers do not learn through discussion alone,
“I realize when I talk to her, it doesn’t sink in… so I’m trying not to talk too much to her now. I’m just trying to show her…. She was much better when I modelled things for her and I gave her an observation sheet and told her to take note of everything that she’d seen and to try and analyse things a bit…. I’m taking her to see another new entrant teacher who’s brand new who is also struggling and I’m also taking her to see an experienced teacher and I’ve got them each to talk about two different things with her.”

In the two subsequent episodes (14 and 15), the transcripts revealed very little initiation from the teacher concerned. The facilitator questioned the teachers and posed questions for the teacher to take from her modelling of practice. As a result of these activities, together with examining student outcomes, this teacher finally made changes to her practice. When asked what led to the change, she indicated a number of influences that finally came together for her. The most salient for her was observing other teachers. She described the experience in the following way:

“... visiting the other schools and just observing how the other teachers teach writing. I think I was having low expectations of my children and I’ve got high expectations now.... I just thought, ok they are five they can only write one sentence type .... After observing at the other school and they were doing narratives and I thought “My gosh, my goodness me”, and I thought “My kids could do that.... Yes well here I was thinking I was doing the right thing.”

She also found it useful to observe the strategies teachers used to stimulate the children with their writing, the resources used and ways to give a purpose for writing.

This teacher had also observed the facilitator prior to this experience (episode 14), but this episode was not as powerful a learning experience at this time. The limited learning probably arose for three reasons. She indicated in her interview with the researcher that she was unsure of the reason for the modelling and her expectations of her students and herself at that time were still low. She also mentioned a theme evident in the other sections of this analysis, “I just don’t know where to start. I think ‘Where do I start?’ This teacher needed to have a vision of possibilities for students like hers and how they could be taught before she could engage in developing new strategies for teaching.

Another influence on her teaching that she referred to in her interview with the researcher, but the episode itself was not recorded, was working through the results of the Observation Survey (Clay, 1993). She described the changes like this:

I suppose the alarm bells weren’t ringing with the six year nets because I thought it will just come to them when they are ready.... [The facilitator] gave me a wake up call and sort of showed me, ‘Look these kids need to know this’, and I thought, ‘OK’ and here I was just cruising along, just waiting for the children to be ready.

She concluded her interview by indicating how much her job satisfaction had improved since she had changed her practice.

**Conclusions**

These teacher learning episodes are analyzed mostly as a snapshot in time. Although they were continued over several episodes for four teachers, for others the research was typically a one-off event in a series of
professional learning experiences. By undertaking this detailed analysis of specific episodes, it is not intended to imply that one-off learning events are effective. In general, they are not. Several teachers mentioned that the value of the Literacy Professional Development Project was its ongoing nature. As one teacher described, “I think having [the facilitator] come in keeps you going. Sometimes you have a meeting and after a couple of weeks it drifts away because school is a busy place, and there’s always other things happening. But if you know [the facilitator’s] coming, you think, ‘Oh I must do this or I must keep up’.”

Another contrasted the project with the type of professional development she had experienced in the past. Well the thing is you go on a course, you sit down, you are wowed, you take these great notes, you get back and no one else in the school has done that. Two days later you think you should do it and it’s gone, it’s not as clear as it was before and you forget to do it, you get busy. This way is better where you can actually try it and get feedback from the person who gave you that idea, not just throw the idea out because it didn’t work.

Even within the obvious methodological limitations of this analysis of teacher learning episodes, some important principles can be drawn from the distinguishing features of more and less successful episodes from the teachers’ perspectives.

The first relates to the overall principle of negotiation and co-construction as ways to promote engagement and learning. Whether it is the purpose of a particular activity, the form of the activity, the learning the activity is supposed to promote or the meaning of particular data, each aspect needs to be negotiated and co-constructed with participants. The number of teachers who reported disengagement from particular episodes for a variety of reasons indicates how important it is to undertake the construction of these activities together. This principle does not mean that teachers should control the agenda because the facilitators’ expertise is a crucial ingredient to the process. Teachers may wish to avoid having their practice deconstructed, or the outcomes for their students examined, but this should not be an option if it is potentially detrimental to their students’ learning. On the other hand, this principle does not mean that facilitators should control the agenda because teachers have insights into their learning needs and the circumstances under which they learn best. Ongoing checking is an essential part of the process.

Associated with this principle is that of engaging teachers’ existing theories of students and how to teach them effectively. This also is a process of co-construction. Few episodes showed substantive engagement at this level yet it is important for two reasons. The first relates to the activation of prior learning so that new learning can build upon prior learning, as outlined in the introduction. The second relates to the high possibility that substantive change involves challenging existing theories and creating dissonance with particular positions. In the absence of theory examination, new practices are typically overlaid on previous practice, with superficial, rather than substantive change evident (e.g. Darling-Hammond et al., 2005). Darling-Hammond et al. have referred to this problem as one of “over-assimilation”.

Another associated principle is that of promoting self-regulation. When teachers have participated in setting personal goals, understand how to monitor their progress towards them and have the support to make appropriate changes, they are more likely to engage at a deeper level. A second reason for promoting self-regulation is that all professional development projects come to an end and unless teachers have the skills and systems in place to continue to monitor their own improvement in teaching and outcomes for students, further learning will remain dependent on the presence of an external person.
References


Appendix A

A Guided Reading Scenario

The following is a description of a teacher’s instruction during a guided reading session with a group. The teacher’s lesson objective is to raise the students’ comprehension through close reading and, in this lesson she aims to teach the students how to summarise information by using the strategy of making a visual summary. She has selected a school journal text about a lighthouse in a very dangerous area for sailors—French Pass in the South Island.

The teacher begins by telling the students the title of the story and asking them what they know about lighthouses. She asks the students to look at the pictures in the text and use them to help make some predictions about what might happen. She talks about the lonely but sometimes exciting life of lighthouse keepers and relates it to a book they have previously read about shipwrecks.

After reading the first two action-packed paragraphs, the students take turns to tell their partner what has happened so far. The teacher then asks the students to read the next five paragraphs to find out what happens next. After a brief discussion to help the students’ understanding of the text, they then read the rest independently and complete the follow-up activity that involved sketching a lighthouse and its likely location.

At the end of the lesson, the teacher summarises the session by asking the students what they have learned about lighthouses. They recalled that lighthouses are often built in dangerous places because they are there to warn ships. The teacher was impressed with their understanding of the text and responded, “Great, you seem to have understood it well.”

Rate the effectiveness of the following aspects of the lesson using the 1 – 6 scale. Give reasons for your rating.

<table>
<thead>
<tr>
<th>highly ineffective</th>
<th>moderately ineffective</th>
<th>slightly ineffective</th>
<th>slightly effective</th>
<th>moderately effective</th>
<th>highly effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

6. The aim for the lesson (paragraph 1) Rating:__________
Reasons for rating: ______________________________________________________
________________________________________________________________________
________________________________________________________________________

7. Arranging activities with partners to support the teacher’s aim for the lesson
(paragraph 3) Rating: _________________
Reasons for rating: ______________________________________________________
________________________________________________________________________
8. Feedback to the learners (paragraph 4)  
Rating: ____________

Reasons for rating: ________________________________________________
_____________________________________________________________________
_____________________________________________________________________

9. In the context of this lesson, identify two effective teacher-directed moves:

A. _________________________________________________________________
_____________________________________________________________________

B. _________________________________________________________________
_____________________________________________________________________

10. Give reasons for each of your choices:

A. _________________________________________________________________
_____________________________________________________________________

B. _________________________________________________________________
_____________________________________________________________________

11. In the context of this lesson, identify two ineffective teacher-directed moves:

C. _________________________________________________________________
_____________________________________________________________________

D. _________________________________________________________________
_____________________________________________________________________

12. Give reason for each of your choices:

C. _________________________________________________________________
_____________________________________________________________________

D. _________________________________________________________________
_____________________________________________________________________

A Writing Scenario

The following is a brief description of a teacher’s instruction while modelling writing for the whole class (a Year 4 class) where the writing task for the children was to write a recount of what they did on a recent trip or holiday. The purpose was to write to tell other people about what happened on the trip and the aim was to produce writing with appropriate structure that will be of interest to the reader.

Miss A modelled the writing process by composing a description of a beach she had been to on holiday. She asked the children to picture a place they had been to recently that stood out in their minds. Several children contributed by telling about their trips to fun parks and beaches. The teacher then emphasized the need for specific vocabulary to convey a visual picture to the reader.

She got the children to draw a picture of the place they visited and label features to help them with vocabulary they might use. Then they swapped pictures with a partner and asked each other questions about the drawing. Then the children wrote their recount.

Rate the effectiveness of the following aspects of the lesson using the 1 – 6 scale. Give reasons for your rating.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>highly ineffective</td>
<td>moderately ineffective</td>
<td>slightly ineffective</td>
<td>slightly effective</td>
<td>moderately effective</td>
<td>highly effective</td>
</tr>
</tbody>
</table>

6. The aim for the lesson (paragraph 1)  Rating: __________
Reasons for rating: ________________________________________________________________
________________________________________________________________________

7. Introducing the writing task (paragraph 2)  Rating: _______________
Reasons for rating: ________________________________________________________________
________________________________________________________________________

8. Arranging the paired activities to support the writing (paragraph 3) Rating: ____________
Reasons for rating: ________________________________________________________________
________________________________________________________________________

Here is the writing produced by one student. Read what Keri produced as a recount and give your feedback to the student.
Keri’s writing

Hi I am at home planning my next trip to syned and Brisbaned for Christmas and New years day I am going to stay in syned for a week and I am going to stay in Brisburnd for a week with my mum’s flat mate. When I went over to syned I met rua his a dog of chris’s. Chris is one of marys flat mate now last time when I went there I had to count his money. There is bazz she’s another one of marys flat mate and nan she has 3 children one is around 14 years old the seoncod is 2 years old and the youngs child is nine mothes old. Then I went to Brisbured. When I got to syned I am going to go to all this fantsey parks and wet and wild and I am going to stay in a hotal.

9. Write what you consider would be the two key aspects of feedback you would give to this child that will help to improve his/her writing.

Feedback 1- _____________________________________________________________
________________________________________________________________________

Feedback 2- _____________________________________________________________
________________________________________________________________________
Appendix B

Part A

Assessing student achievement: The case of Riverdale School and use of evidence

The staff at Riverdale School decided that they wanted to improve their students’ reading levels. They thought comprehension was a problem because when they did the running records most of the students could read the words accurately at their age levels but the teachers felt that they didn’t really seem to know what the passages were about.

At the next staff meeting they met and brainstormed all the different ways they could help improve the students’ comprehension. Some of the teachers knew that the neighbouring school had great success with a tape-assisted reading programme [like Rainbow Reading]. The teachers in that school had told them how much more the students enjoyed reading. The teachers agreed to try the programme in their classes and to review it at the end of the term.

They found the process really supportive because the first 10 minutes of each syndicate meeting was devoted to discussing implementation issues, such as organizing the listening corner and finding the time to test the students to establish if they should move up a book level. Other teachers described how they managed these kinds of problems.

At the end of the term, the teachers reported that their work with the programme appeared to be bringing about significant gains. The teachers completed their usual end of term running records and they noticed that the results showed nearly all the students were reading more fluently and confidently compared with the previous term. They decided to continue with the programme the following term.

16. Please rate how effective you think Riverdale School seemed to be in their approach to the following (and give reasons for your rating):

<table>
<thead>
<tr>
<th>Rating</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>highly ineffective</td>
</tr>
<tr>
<td>2</td>
<td>moderately ineffective</td>
</tr>
<tr>
<td>3</td>
<td>slightly ineffective</td>
</tr>
<tr>
<td>4</td>
<td>slightly effective</td>
</tr>
<tr>
<td>5</td>
<td>moderately effective</td>
</tr>
<tr>
<td>6</td>
<td>highly effective</td>
</tr>
</tbody>
</table>

A. The way they identified the students’ needs

B. The basis for their decision to adopt the tape-assisted reading programme

C. Identifying issues at the syndicate meeting

D. The use of the running records in deciding to continue with the tape-assisted reading programme
Part B:

Helping Tracey make sense of some data
17. Tracey is a first year teacher in your Year 6 syndicate. She brought her reading test results to you because she did not know what they mean or how she should use them. The school’s emphasis is on comprehension, so she particularly wanted to know about this aspect of reading.

Reading Test Results for Tracey’s Room.

<table>
<thead>
<tr>
<th>First name</th>
<th>Raw Scores for Comprehension of Paragraphs (20)</th>
<th>Raw Scores for Comprehension of Sentences (10)</th>
<th>Raw Scores for Word Recognition (10)</th>
<th>Total Raw Score (40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Sarah</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Colin</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>Angus</td>
<td>18</td>
<td>10</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>Stacey</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>Olivia</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Sione</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>Thomas</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Mali</td>
<td>19</td>
<td>9</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>Simon</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Witi</td>
<td>14</td>
<td>7</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Julia</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Christopher</td>
<td>12</td>
<td>8</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Moana</td>
<td>14</td>
<td>7</td>
<td>8</td>
<td>29</td>
</tr>
<tr>
<td>Rachael</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>Johnathan</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Junior</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Lauren</td>
<td>14</td>
<td>8</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Andrew</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Mere</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Av. For class</td>
<td>10.45 out of 20</td>
<td>8.9 out of 10</td>
<td>7.05 out of 10</td>
<td>25.7 out of 40</td>
</tr>
</tbody>
</table>

**Note:**
- A critical score is a low score that is a cause for concern.
- Children who score 11 and below on the total score are considered to be non-readers.

**Raw Scores Comprehension of Paragraphs**
- class mean = 10.45  NZ mean = 13.9
- class range = 3-19  typical range = 11-17
- critical score = 7

**Raw Scores Comprehension of Sentences**
- class mean = 8.9  NZ mean = 8.6
- class range = 5-10  typical range = 8-10
- critical score = 7

**Raw Scores Word Recognition**
- class mean = 7.05  NZ mean = 7.9
- range = 3-10  typical range = 8-9
- critical score = 5

**Raw Scores Total Score**
- class mean = 25.7  NZ mean = 30.9
- class range = 17-38  typical range = 26-37
- critical score = 23
Questions:

A) What would you tell Tracey about the meaning of the sentence and paragraph comprehension results for Room 10?

B) The syndicate leader has asked each teacher to summarise how well his/her class is reading. What main points would you suggest that Tracey include in this statement using the results from her class?

C) Rate the usefulness for Tracey in using results like these for her teaching.

Please circle one

<table>
<thead>
<tr>
<th>definitely not useful</th>
<th>not really useful</th>
<th>slightly useful</th>
<th>moderately useful</th>
<th>mostly useful</th>
<th>definitely useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Reasons for rating: _____________________________________________________________

____________________________________________________________________________

____________________________________________________________________________
## Appendix C

### Timetable of visits and in-school activities

Chronology of visits, main activities and people involved

<table>
<thead>
<tr>
<th>Date</th>
<th>Main activities of facilitator and staff</th>
<th>Who involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 July</td>
<td>Initial interview, 3 observations and research questionnaires / scenario completed</td>
<td>Facilitator and researcher</td>
</tr>
<tr>
<td>9 / 10 August</td>
<td>Cluster meeting – outlined phases of programme, effective assessment, data collection tools (STAR, asTTle), described the needs analysis phase, analysed samples of writing using exemplars.</td>
<td>Team leader and facilitator - attended by literacy leaders (not principal)</td>
</tr>
<tr>
<td>24 August</td>
<td>Staff meeting: feedback on observations through analyses of practices, consequences and beliefs (see separate description).</td>
<td>Facilitator - attended by literacy leaders and staff (not principal)</td>
</tr>
<tr>
<td>Between visits</td>
<td>Discussed the Best Evidence Synthesis (Alton-Lee, 2003) and 10 characteristics of quality teaching and Shifting the Focus (Timperley, 2003). Developing positive school culture for writing. RTLit showed how to level writing samples.</td>
<td>Whole staff</td>
</tr>
<tr>
<td>2 September</td>
<td>Worked with literacy leader to observe staff and give feedback (see separate description). Followed by a staff meeting: Feedback on questionnaire results, linked results to the observation data and wrote the “action steps” on the analyses completed 24 August. Advised on readings to start improving practice in areas identified. Wrote whose responsibility to do readings had how were going to be fed back to staff.</td>
<td>Facilitator (Not principal)</td>
</tr>
<tr>
<td>Between visits</td>
<td>Principal and literacy leader undertook 30 minute classroom observations and interviewed students. Found that students were not as articulate about what they were supposed to be learning and how they would know they are successful. Encouraged teachers to interview students. Observation and feedback questions comprised: “What were the learning intentions and how did the children know these?” “What is a recount and how do the children know how to place these in their stories?” Prior knowledge was discussed but the learning intentions were not stated, “Did the children know what they were?” “What opportunities did you give to students to talk about quality writing?” Discussed how they would report specific learning intentions to parents and how they could help. Also discussed writing levels using exemplars.</td>
<td>Staff</td>
</tr>
<tr>
<td>Date</td>
<td>Activity</td>
<td>Facilitator(s)</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------</td>
</tr>
</tbody>
</table>
| 18 October | Staff meeting: Analysis of asTTle data planned but not ready. As a result of above class observations, facilitator asked to present a workshop on learning intentions and success criteria. Approached the task by refining staff understanding of the needs analysis and what data had led them to prioritise this aspect of their practice. Readings presented (in jigsaw) with three guiding questions posed to help teachers reflect on the readings and their presentation:  
  • Why is it necessary to make the learning explicit?  
  • How would you write and present LI and SC?  
  • What might it mean for your own practice?  
  Recommended activities and learning between visits:  
  1. Learn about the features of recounts from recommended readings prior to developing LIs;  
  2. Read chapter 4 of ELP about feedback and different approaches that could be used. | Facilitator & all staff – including principal |
|            | Undertook two activities in staff meetings as described above.  
  Reading recovery teacher explained how writing undertaken in reading recovery.  
  Discussed curriculum plans and vision in charter for writing goals. Discussed specific goals in English for teachers to focus on and to include in variance report. | Staff                               |
| 27 October | Updated principal on all data because principal had not been present at previous meetings. Discussed individual teachers and their practice. Established goals for the school and where should be heading next.  
  Staff meeting: Examined asTTle and exemplar data | Facilitator                         |
|            | Discussed key competencies for curriculum delivery from Ministry – decided wanted specific goals for English and Maths. Decided focus next year will be on deeper features of writing.  
  Performance management data - individual staff filled in individual goals and how they were achieving them. | Staff                               |
| 9 November | Interviewed principal, literacy leaders and staff – responses formed the basis of this case study write-up | Facilitator and researcher           |
| 11 November| Observed x2 teachers to check on progress. Arranged for follow-up asTTle data.                                                                                                                       | Facilitator and researcher           |
Appendix D
Analysis of observation and feedback (including student interviews) as opportunities to learn

Typically took the form of a teacher being observed, sometimes with criteria shared prior to the observation, followed by feedback.

<table>
<thead>
<tr>
<th></th>
<th>Prior conditions</th>
<th>Learning goals explicit &amp; shared</th>
<th>Engaged teachers’ theories / co-construction</th>
<th>Content covered</th>
<th>Any other activities apart from Ob &amp; Fdbk</th>
<th>Learning processes, reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Professional community well established within syndicate</td>
<td>At a general level: “To provide me with feedback so I can improve my teaching”</td>
<td>Began by asking questions and reasons for doing things, but theories not engaged. Time constraints led to “telling”. Limited co-construction</td>
<td>• Used checklist of effective pedagogical techniques with reasons. Linked to how to teach the content but content itself not engaged in depth. • Checklist used to explain that students will understand better if they are given reasons because interviews established that reasons for class activity not given. • Checklist based on a theoretical framework but principles not mentioned to teacher in this episode. • Checklist used to comment on teacher practice. • Limited skills for self-regulation.</td>
<td>F. provided possible student activities (but limited co-construction)</td>
<td>T. indicated gained an understanding of problem related to students not understanding reasons for activities. Indicated had clear idea of how to change and has changed practice. No indication of enhanced self-regulated learning (SRL).</td>
<td>Most salient learning occurred when facilitator described students’ response. Teacher indicated that what she really wanted was to observe someone implementing particular aspects of practice</td>
</tr>
<tr>
<td>3</td>
<td>Professional community well established within syndicate</td>
<td>At a general level: “to give us feedback on our teaching methods and suggest ways of improving it”</td>
<td>Began by asking questions and reasons for doing things, but theories not engaged and limited co-construction. Time constraints led to “telling”.</td>
<td>• Used checklist of effective pedagogical techniques with reasons. Linked to how to teach the content but content itself not engaged in depth.</td>
<td>• Analysed current practice but did not co-construct new practice.</td>
<td>Some cueing and consolidating of prior knowledge / understandings. T. indicated could not see how to implement advice. “I took it on board, but didn’t change anything as a result”.</td>
<td>F. had concerns about lesson. Teacher thought the lesson was good. F. was unclear in feedback about what was problematic and embedded suggestions in praise. Previous experience had led to negativity on the part of T. T. did not learn from students’ responses - not clear in feedback if F. found them problematic and T. indicated F. had asked wrong students. Feedback generally vague and descriptive.</td>
</tr>
<tr>
<td>13</td>
<td>Participation but limited engagement in a professional community</td>
<td>Not asked</td>
<td>Engaged T’s theories at superficial level – key issue T. wanted students to be confident to write but students did not have strategies for doing so. Not engaged to extent that T. believed students would learn when ready. Co-constructed meaning and misconceptions.</td>
<td>Pedagogical content knowledge - variety of strategies for assisting early learning in writing. Checked understanding. Follow-up in 2 weeks proposed as problem-solving.</td>
<td>Ideas for equipment such as whiteboards.</td>
<td>Did not change practice (follow-up observation from F). T. indicated “… children develop when they are ready … I thought it will just come to them” Together with “I just don’t know where to start really”</td>
<td>T. not sufficiently skilled to benefit from feedback – combination of beliefs and skills. Later interviews indicated did not understand her role as a teacher and did not understand F. suggestions. F. realized problem and changed to modelling practice, “I realize when I talk to her it doesn’t sink in”</td>
</tr>
<tr>
<td>Prior conditions</td>
<td>Learning goals explicit &amp; shared</td>
<td>Engaged or bypassed teachers’ theories</td>
<td>Learning resources provided</td>
<td>Any other activities apart from Ob &amp; Fdbk</td>
<td>Learning processes</td>
<td>Reactions / actions</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------</td>
<td>----------------------------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| Limited engagement in a professional community | No – beyond getting a snapshot of how teachers were teaching | Closely engaged T. theory through analysis of practices and beliefs underpinning them | • Pedagogical content knowledge.  
• An understanding of impact of practice on students  
• Theoretical frameworks into which specific practices were located  
• Knowledge and next steps co-constructed.  
• How might inquiere into impact of practice (self-regulation) | Readings related to areas of concern | Created dissonance with beliefs about what students understood from instruction. Created the motivation to engage and change practice. | Actively engaged with readings. Changed practice substantially. Began to interview own students to ensure learning purpose underpinning instruction clearly understood. Had clearly understood next steps. | Students’ interview responses created motivation to learn new skills and knowledge. Co-construction process ensured teachers understood the purpose of future activities. |
|   | Limited engagement in a professional community | No — beyond getting a snapshot of how teachers were teaching | Closely engaged T. theory through analysis of practices and beliefs underpinning them | Pedagogical content knowledge. | An understanding of impact of practice on students | Theoretical frameworks into which specific practices were located | Knowledge and next steps co-constructed. | How might inquire into impact of practice (self-regulation) | Readings related to areas of concern | Created dissonance with beliefs about what students understood from instruction. Created the motivation to engage and change practice. Actively engaged with readings. Changed practice substantially. Began to interview own students to ensure learning purpose underpinning instruction clearly understood. Had clearly understood next steps. Not as skilled as T. 16 to implement practice | Students’ interview responses created motivation to learn new skills and knowledge. Co-construction process ensured teachers understood the purpose of future activities. |
## Observation and feedback without student interviews

<table>
<thead>
<tr>
<th>Prior conditions</th>
<th>Learning goals explicit &amp; bypassed teachers’ theories</th>
<th>Engaged or bypassed teachers’ theories</th>
<th>Learning resources provided</th>
<th>Any other activities apart from O &amp; F</th>
<th>Learning processes Reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>No well established professional community</td>
<td>No engagement with teacher theories or co-constructin</td>
<td>• Feedback on lesson brief, non-specific and positive.</td>
<td>• Gave strategies for teaching vocabulary but unrelated to lesson observed or teacher questions. Appeared to be a prior decision by facilitator.</td>
<td>Difficult to determine. T. indicated liked to be told what to do and how to do it. T. indicated:</td>
<td>F. appeared to be following a pre-determined agenda unrelated to observation.</td>
</tr>
<tr>
<td></td>
<td>Learning goals did not appear to have been explicit or shared prior to observation</td>
<td></td>
<td>• No theoretical frameworks offered (in feedback) and teacher indicated did not like theory (in interview).</td>
<td>• Teacher’s request for specific information met with non-specific responses.</td>
<td>• not given the prescription he was looking for.</td>
<td>Liked practical suggestions – but not theory.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• No skills for self-regulation</td>
<td>• F. recommended readings (not related to lesson observed or T’s questions)</td>
<td>• would prefer to observe others teaching (for organization and delivery) than receive feedback on own practice. Does not find talking useful.</td>
<td>Issue – how to balance self-identified needs with other-identified needs. Neither engaged during feedback.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• F. recommended readings (not related to lesson observed or T’s questions)</td>
<td>• wanted (but did not get) a framework for curriculum progressions</td>
<td></td>
</tr>
<tr>
<td>No well established professional community</td>
<td>Learning goals did not appear to have been explicit or shared prior to observation</td>
<td>No engagement with teacher theories or co-construction</td>
<td>No feedback on lesson.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Rationale given for practical suggestions but in terms of a coherent theory – disconnected from each other</td>
<td>• Provided some pedagogical content knowledge in terms of teaching ESOL students. Explicit suggestions about how to teach but unrelated to observed lesson and disconnected from each other.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No skills for self-regulation</td>
<td>• Discussion and examples of how to record students’ responses for monitoring purposes (unrelated to above)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Reasons always given for suggestions but not presented in terms of a coherent framework.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teacher indicated had gained new knowledge but areas referred to not mentioned in feedback. T. indicated had no idea of where to next: “No. I have no idea really at all. T. says wants to be given clear ideas to work on.

Problems may have arisen because feedback occurred 3 weeks after observation. Teacher asked for interview of feedback transcript because thought had useful ideas and had not received anything in writing from F.
<table>
<thead>
<tr>
<th>Prior conditions</th>
<th>Learning goals explicit &amp; shared</th>
<th>Engaged or bypassed teachers’ theories</th>
<th>Learning resources provided</th>
<th>Any other activities apart from O &amp; F</th>
<th>Learning processes Reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established through project over previous year</td>
<td>Not explicit beyond, “This is a snapshot of where you are in your teaching”.</td>
<td>Teacher often defended what doing in response to implied criticism. Theories not really engaged.</td>
<td>F. gave pedagogical advice embedded in questions.</td>
<td>Some specific suggestions about questioning, and reducing noise levels and introducing learning intentions at beginning of lessons.</td>
<td>Difficult to discern. T’s interview responses indicated had learned little. Little reason to be optimistic about future action. Insufficient facilitator guidance or teacher knowledge to put into practice. Rarely has reading lessons (week 7 and only 4-5 lessons).</td>
<td>Feedback embedded in questions followed by advice with no checking advice was useful. F. and literacy leaders serious concerns about her management and pedagogy not conveyed or understood.</td>
</tr>
</tbody>
</table>
| No. | Established through project over previous year | No. T. “They told me they would be coming to observe me and not to worry, they wanted to see where I was at”. | Bypassed – questions followed by advice. Limited co-construction | • Told on the right track – needs to share success criteria and work on paragraphs.  
• No or skills for self-regulation | • T. asked for book to read to class – F. offered to help.  
• F. indicated needed whiteboard to display LIs and SC. | Not clear – possibly becoming aware of some new knowledge to limited extent. Confirmed current practice. T. indicates needs modelling and more discussion to change practice. | • T. waiting for F. to model practice (has indicated she would). |
## Analysis of Syndicate and whole staff meetings / workshops

Episodes 1, 4, 6, 7 - syndicate meetings

Episodes 5, 12 – whole staff meetings

<table>
<thead>
<tr>
<th>Focus of meeting</th>
<th>Prior conditions</th>
<th>Learning goals explicit &amp; shared</th>
<th>Engaged or bypassed teachers’ theories</th>
<th>Learning resources provided</th>
<th>Learning processes / reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Examine shifts in student achievement and identifying next teaching steps</td>
<td>Had engaged in activity prior to this – gradually becoming more focused</td>
<td>T’s A &amp; B clear about goals – to link the data with next teaching steps, T. C unclear – prefers other kinds of support.</td>
<td>Engaged – particularly T’s A &amp; B and challenged at times.</td>
<td>How to use asTTle data to identify what students know and what needs to be (T. 3 pleased with progress so found exercise irrelevant.) Principle of using data formatively embedded in how to do it. Responsibility for promoting student learning firmly placed with teacher (digressions challenged) References also made to importance of feedback. Self-regulation promoted by using asTTle to understand progress and whether progress was adequate.</td>
<td>Developing new skills in interpreting data and linking to teaching. Tchrs 1 &amp; 2 reported how intended to practice as result of meeting. T. 3 believed progress adequate so no need to change. Helped her to focus on students who need attention.</td>
<td>Differential success with teachers. T.2 liked visual representation and discussion. Both T. 1 &amp; 2 indicated the need for multiple opportunities to learn in different ways and this episode helped with that process – needed to be combined with observation of other teachers (T.1) and discussion (T.2). T.3 wants detailed units of work, observing others and having a go herself. Belief that own students showing good progress not borne out by results. Her stated preferences do not put her in the spotlight.</td>
</tr>
<tr>
<td>4</td>
<td>Examining shifts in student achievement in one genre and working out ways to use this information for teaching a second genre</td>
<td>Episode 1 and a number of other similar ones</td>
<td>Some confusion. T’s initially thought were planning for next term but after some discussion during meeting had goal closer to that of F. – “using data from persuasive writing to plan a different genre”.</td>
<td>To a limited extent – T’s described what doing in their classrooms and why.</td>
<td>Using data to inform subsequent teaching. Pedagogical content knowledge for writing. Principles related to teaching one genre transferred to another. Limited focus on self-regulation in that focus was on using data for future planning.</td>
<td>Both actively engaged and described major new learnings. T.2 Used previous knowledge of one genre to inform teaching in another. T.3 In response to question – what did you learn during this session, “I don’t know about learnt” then went onto to describe all things had learnt and how could apply them including: identifying target groups, looking for language resources, transferring skills from different genre, testing children before teaching, increased content knowledge.</td>
</tr>
<tr>
<td>5</td>
<td>Building content knowledge of types of formative assessment and conferencing</td>
<td>Episodes 1-4 and previous year’s work on contract</td>
<td>F. believed had requested session as follow-up to earlier session. Teachers indicated did not need it.</td>
<td>Some engagement during role play sessions, but large amount of content knowledge meant F. mostly telling.</td>
<td>Linking formative assessment with teaching and conferencing, building pedagogical content knowledge using Bloom’s taxonomy. Touched on a variety of theoretical frameworks including Bloom’s taxonomy, Vygotskian ZPD. Variety of activities (whole day session) – own writing prior to session, listening to F., role-playing student – teacher conferences. Meeting structured around F. handouts</td>
<td>Episode designed to be built around prior knowledge from previous PD. T. 1 “waste of time but made her more aware of importance of children working together and with teacher to share their ideas.”. From report appears to be implementing it at a low level and introducing cosmetic change as a result of session. T.2 “extension of stuff already looked at” – would like to be able to use it but doesn’t think can until sees it in action. (compare with previous reactions of T.1 &amp; 2). T.3 confirmed what already knew but positive about it – reports is developing new questioning skills (contrast with previous reactions of T.3).</td>
</tr>
<tr>
<td>6</td>
<td><strong>Examining shifts in student achievement and identifying next teaching steps</strong></td>
<td><strong>Episodes 1 &amp; 4</strong></td>
<td><strong>Clear statement at beginning by F. — identifying progress of target students, problem, solutions, but all teachers unclear of specific purpose because stated purpose did not relate to time frame of term — had insufficient time to teach and assess.</strong></td>
<td><strong>T. interviews indicated theories bypassed.</strong></td>
<td><strong>Formative assessment (but teachers did not have any data to work with). Relationship between examining student progress and teaching an underpinning assumption throughout meeting. Teachers asked to raise problems and suggestions offered by others. Designed to enhance self-regulation but failed to achieve this.</strong></td>
<td><strong>All T’s indicated some kind of resentment. T. 1 — spending too much time on writing. T’s 2 &amp; 3 had insufficient time to teach and assess prior to meeting (compare with episode 1).</strong></td>
</tr>
<tr>
<td>7.</td>
<td>Examining the relationship between reading and writing</td>
<td>Not sure</td>
<td>Clear statement by F. at beginning of meeting, “Relationship between reading and writing”. Teachers all aware of learning goals.</td>
<td>T. 1, 2, &amp; 4 indicate engagement. T.3 interviews indicated episode bypassed personal theories.</td>
<td>Pedagogical knowledge of text analysis, close reading, critical thinking, language exploration, information processing, reading / writing links. Provided with principles of complementary nature of reading and writing. Activities structured around f. handouts on key content areas.</td>
<td>T’s 1, 2, 4 identified new information and skills and identified how have and would continue to change practice – a powerful session. T.3 some dissonance created but rejected it – still teachers writing, reading and word study separately. “So I’ve come from a completely opposing viewpoint to the complete and utter opposite and now I’m really confused about what I should be doing. So for me it’s why rush and listen to this because in 2 years time I will be told that was all wrong and I’ll have to do it a different way again … It’s very hard for old habits to die”.</td>
</tr>
<tr>
<td>12</td>
<td>Feedback from observations of staff from senior school (see episodes 17 &amp; 18). T. interviewed Year 0/1.</td>
<td>First staff meeting of contract</td>
<td>Goals understood by teachers from senior school but Year 0/1 teacher did not see them as relevant to her level.</td>
<td>Bypassed person theories of what it means to teach Year 0/1. “No, I definitely didn’t feel part of it”</td>
<td>Formative assessment in writing</td>
<td>Senior teachers found session powerful in changing practice. Year 0/1 teacher indicated did not learn anything and did not apply to her.</td>
</tr>
<tr>
<td>17</td>
<td>Reviewing student progress and planning next steps</td>
<td>9 months into contract including episodes 12, 13, 16, 18</td>
<td>Meeting requested by literacy leaders and teachers. Goals contested by new member of staff.</td>
<td>Theories engaged including contesting F. theory</td>
<td>asTTle levels, implications for teaching</td>
<td>Both teachers interviewed found session useful for deepening understanding of features and functions of writing. Indicated would change practice.</td>
</tr>
</tbody>
</table>
### Teacher observing Others

<table>
<thead>
<tr>
<th>Prior conditions</th>
<th>Learning goals explicit &amp; shared</th>
<th>Engaged teachers’ theories / co-construction</th>
<th>Content covered</th>
<th>Learning processes, reactions / actions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 9 months into contract. Episodes 12 &amp; 13.</td>
<td>Explicit on part of F. but T. had undisclosed competing theories and unsure of reason for modelling.</td>
<td>F. attempted to do so at practice level T. very difficult to engage. Deeper competing “readiness” theories and consequent role as a T. not engaged.</td>
<td>Strategies for engaging early writers, vocabulary development and differentiated instruction for students.</td>
<td>Limited learning because “readiness” theories not challenged. Some minor shifts in practice but indicated “didn’t know where to start”.</td>
<td>T. been provided with multiple opportunities to learn but had competing theories and insufficient skills to benefit substantively. Modelling marginally better than telling in that some practice shifts evident according to F.</td>
</tr>
<tr>
<td>15 10 months into contract. Episodes 13 and 14.</td>
<td>Shared and finally understood by T.</td>
<td>Indirectly. F. attempting to engage level of theory underpinning practices. Observing other T.s allowed T. to see what Yr level of students capable of and the implications for her role.</td>
<td>The development of early writing strategies. Differentiated instruction.</td>
<td>Dissonance created between T’s expectations of students and her own role and what was observed in other T’s classrooms.</td>
<td>T. could not benefit substantively from analysis of own practice until had a better understanding of developmental progressions in writing and her role own, more active role as a T.</td>
</tr>
</tbody>
</table>