Educational Leadership Practices Survey

Report on the National Norms and Benchmarking

Report prepared for the Ministry of Education

Jacky Burgon, Hilary Ferral, Edith Hodgen and Cathy Wylie
New Zealand Council for Educational Research

AUGUST 2011

New Zealand Council for Educational Research
P O Box 3237
Wellington

Table of Contents

Tables

Figures

Appendices

Executive summary

Content of the Educational Leadership Practices Survey (ELP)

The ELP is designed to provide a robust picture of how effective a school's teachers perceive the school's educational leadership to be in those key aspects that our current evidence shows are the ones most likely to have an impact on teaching and learning.

It covers these nine different scales of school educational leadership:

- *Goal Setting*
- *Strategic Resourcing*
- *Curriculum Quality*
- *Quality of Teaching*
- *Promoting and Participating in Teacher Learning and Development (Teacher Learning)*
- *Safe and Orderly Environment*
- *Educationally Powerful Connections with Families, Whānau and Community (Teaching Learning)*
- Māori Success
- Principal Leadership.

These aspects are based on the vision for educational leadership set out in the Kiwi Leadership for Principals (KLP) framework, and dimensions for effective educational leadership practice as described in the Educational Leadership Best Evidence Synthesis.

The overall leadership scale used in the ELP is constructed from the seven scales of school leadership given in italics above. In the trial and calibration of the final ELP ready for use in the Experienced Principals Development (EPD) programme we found that each of these seven aspects had high internal reliability (alphas of 0.81 to 0.88)—the items in each scale "hung together" well, and were measuring dimensions of the same underlying construct. The seven scales also have a high level of intercorrelation, and a Rasch analysis confirms that the seven scales can be treated as different aspects of a single underlying meta-construct—overall leadership practice.

The items used to make the Māori Success scale were originally included in these seven scales (and are asked within the survey, rather than as a separate set of questions). However, we found that these items did not contribute to the internal reliability of the scales they were in, and in fact "hung together" well as a separate scale. The Principal Leadership scale was a separate set of questions from the start, and is not included in the overall school leadership scale since it focuses on the principal alone.

Together these nine scales give a rich picture of educational leadership in schools.

Focus of this report

In this report we focus on:

- reporting the patterns of distribution—norms—for each of the seven scales that make up the overall scale, as well as the Principal Leadership and Māori Success scales that form part of the ELP also, along with principals' views of the pedagogical contexts (barriers and support) for their educational leadership
- examining difference in patterns of scores related to school characteristics, and presenting scaled scores related to school characteristics (size, decile and type), to support a "schools like us" comparison
- reporting on the outcomes of a benchmarking exercise to establish some suggested benchmarks for different levels of educational leadership performance on the ELP. By benchmarking we mean comparing a school's ratings in ensuring effectiveness in aspects of educational leadership with best/effective practices (comparing a school's rating with the norms simply gives a comparison with the aggregate picture of other schools' ratings). In this report we have titled the four different levels educational leadership "invisible", "basic", "sound" and "exemplary"
- making recommendations about future administration, use, analysis and scoring for the ELP
- making suggestions for further professional development for educational leadership.

Findings

Introductory comments

The national sample we used reflected well the national distribution of school type, size and decile in New Zealand but with an oversampling of larger and secondary schools which allowed sufficient sample size to provide "schools like us" data.

The benchmarking exercise cast some very useful light on the question "How good is good enough?" The exercise results reflect good inter-rater reliability and some sound starting points for discussion with schools.

Summary of overall main patterns relating to educational leadership levels as indicated by the ELP

There was a wide range of school scores on the overall leadership scale, but using our benchmarking approach, almost all schools were functioning at least at a basic level, with many at a sound and some at "exemplary" level.

Goal Setting, a Safe and Orderly Environment and Principal Leadership were the scales that had the highest proportions of schools rating their school leadership as highly or outstandingly effective (50 percent or more). Promoting and Participating in Teacher Learning and Development (Teacher Learning and Māori Success were the scales that had the lowest proportions of schools rating their school leadership as highly or outstandingly effective.

Some key trends in each scale

Goal Setting: There appears to be more confidence about the role of leadership in relation to schools' guiding frameworks than about the embedding of the goals into ongoing use and evaluation—linking the goals back to individual teacher roles and individual student learning.

Strategic Resourcing: The schools gave highest ratings to the effectiveness of their school leadership in ensuring that the timetable reflected the school's priorities for teaching and learning, and lowest to items related to working with families and communities. In between come items related to teaching resource relevance and availability.

Curriculum Quality: School leadership was seen as most effective in ensuring the systematic monitoring of each student's progress and the existence of assessment plans to collect the information needed to monitor progress on priority learning goals, and least effective in ensuring that rigorous feedback was given to teachers about the quality of their schemes or unit plans, that all students experience challenging programmes and that all curriculum included content relevant to diverse learners.

Quality of Teaching: More than half the schools in this national sample thought that their school leadership was highly or outstandingly effective in ensuring that everyone shared responsibility for student learning, that assessment data were used to improve teaching and that those teachers with particular expertise were used in the school to help other teachers' development. The lowest rating item was students providing feedback to teachers on the effectiveness of their teaching, followed by support provided for teachers having difficulty helping students reach important academic and social goals, and challenge and support to improve teaching for teachers whose students remain disengaged.

Promoting and participating in Teacher Learning and Development: Open discussion of student results and teachers helping each other develop more effective teaching strategies, serious discussions of how to improve teaching and learning in staff meetings and analysis and use of student achievement patterns to plan professional learning priorities were the items most likely to attract highly or outstandingly effective ratings of school leadership. Schools were much less likely to give such ratings to the provision of systematic opportunities to improve teaching through observing effective colleagues at work, teachers' use of a range of evidence sources to evaluate the effectiveness of their teaching and adequate opportunities provided for teachers to discuss why they might need to change their practice.

Ensuring a Safe and Orderly Environment: Most schools gave their school leadership highly or outstandingly effective ratings for ensuring they had positive environments for learning in which

student learning is the central focus, and that the school was a positive environment for everyone, irrespective of culture. The gathering and use of student views in relation to school safety and culture were the two items on which the school leadership was least likely to be rated as highly or outstandingly effective.

Ensuring Educationally Powerful Connections with Family, Whānau and Community: Two-thirds of the schools thought their leadership was highly or outstandingly effective in ensuring that staff were responsive to families' views about their child's learning needs. However, only about a third of schools thought that their school leadership was highly or outstandingly effective in ensuring that parents understand the achievement levels of their children in relation to national benchmarks.

Māori Success: Schools were markedly less sanguine in this area. Schools were most likely to rate their school leadership as highly or outstandingly effective in relation to having clear school-wide targets for the academic achievement of Māori students (just over one-third of schools in the national sample), and least likely to rate them so for ensuring that there were professional development opportunities that enabled teachers to develop the knowledge and skills needed to provide quality teaching to Māori learners (only around 15 percent of schools).

Principal Leadership: The top items in this scale were mostly related to the principal promoting the values of the school, having integrity and gaining and showing others respect; it also included making tough decisions when necessary. Identifying and resolving conflict quickly and fairly was the item with the lowest proportion of schools rating their principal as showing highly or outstandingly effective leadership.

Characteristics related to differences between school scores for educational leadership practices

High-scoring schools on the ELP scale are most likely to be primary schools, small schools, high-decile schools, with a Māori roll of less than 15 percent. These differences in school characteristics suggest that teacher views of school leadership effectiveness are likely to be lower where the school organisation is more complex—as it is in secondary and larger schools; or where the challenges of the student population are greater—as they are in deciles 1–2 schools, and in secondary schools.

Relatively higher scores on the Māori Success scale were likely to occur in small schools, lower decile schools and those with over 30 percent Māori enrolment.

Principal Leadership ratings showed less difference related to school characteristics than the other ELP scales. However, secondary and special school principal effectiveness had lower ratings, as did principals of larger and/or lower decile schools.

Multilevel modelling showed that some variables do seem to account for much of the difference between schools in their overall school leadership scores. After accounting for these variables,

only around 10 percent of the schools in the sample showed distinctly different scores (either very low or very high).

The variables that the multilevel modelling found to be associated with differences in school perceptions of the quality of educational leadership practice included contextual factors—particularly school decile, school type and, to a lesser extent, the support for pedagogical leadership (and barriers to its exercise). The modelling also provides some indicators that the school leadership practices covered in the ELP have positive links with teacher morale, good workplace practices and judgements of principal quality.

When looking at the pedagogical context for educational leadership, while supports seem quite similar across different school characteristics, there are some differences when looking at barriers. Capacity and student issues were more likely to occur for principals at deciles 1–2 schools. Secondary principals and principals of larger schools were more likely than others to experience staff management as an erosion of their time for pedagogical leadership, and secondary principals were somewhat less likely to think their workload was manageable or sustainable.

Implications of ELP patterns in relation to the need for focused professional development and support for school leadership

The ELP is based on existing research that shows associations between most of the practices asked about on the ELP and student achievement. Our benchmarking process conducted in this study provides some preliminary information on "how good is good enough" and suggests most schools have a "basic" or "sound" level of educational leadership. This overall national picture provides grounds for both confidence, and some challenges. The current levels of educational leadership practices do suggest that it is important to support further development.

Should professional development and support for educational leadership be targeted?

The analysis shows that the length of principal experience, either in total or at their current school, is not associated with school scores for the effectiveness of either the principal leadership, or the school leadership as a whole. However, the clear associations with school size, decile and type with the ELP scores, and in relation to principal perceptions of support or barriers to their pedagogical leadership, bring up the very real questions of factors beyond individual school control. They also pose real policy issues. Given this real constraint, if there is any need for prioritisation, deciles 1–2 schools and secondary and composite schools stand out.

There appears to be most scope for further development in relation to *Promoting and participating in Teacher Learning and Development* scale, and Māori Success; and in terms of practices related to feedback on performance and effectiveness, providing timely challenge and support to both teachers and students, including student voice, and supporting parent

understanding of student achievement. Of particular concern are the very low ratings around availability of good professional development in relation to Māori achieving success as Māori. Recent changes in the delivery of Ministry of Education-funded professional learning and development contracts have had this as a focus area; over time the ELP might be one way of monitoring outcome changes in this area, as well as enriching needs analyses at school level.

Implications for the ongoing development and use of the ELP

The ELP does provide a useful way of gauging and describing school leadership practices that are linked to teaching and learning. The original intention was to provide a tool with a range of items, including aspirational items that allowed "stretch", and the capacity to show change over time. The national sample results and the benchmarking results both show that the ELP has this capacity.

We cannot tell from the ELP levels alone whether they are high enough to make a real difference to student engagement and performance, or whether there is a minimal level that is necessary to ensure a given level of student engagement and performance. The benchmarking exercise in this report does help us in this latter regard. However, to further investigate the benchmarking results and the relationship of the ELP results to student achievement trends/changes, we would need to link patterns in ELP scores over time to patterns in student engagement and learning, and to the benchmarking results we have outlined in this report.

ELP development

Evidence-based tools need to stay current with the evidence. It is anticipated that over time there might be a need for a revision of the tool to reflect the most recent evidence. One example raised in this report is in the area of effective leadership in the secondary school environment as it related to teacher learning and development. Some further investigation may be warranted in this area. We would suggest a revision of the ELP within the next 5 years to ensure the tool remains reflective of current evidence-based practice.

As already noted, there is still a need to look at the relation between ELP changes and changes in student achievement over time to ensure the ELP has strong concurrent validity.

Sharing the results of this national sample and benchmarking outcomes with sector groups is an important step in continuing to collect data to confirm or refine the benchmarking exercise as well as an opportunity to provide current and potential users of the ELP with in-depth information about the tool and the associated information in this report. This in-depth knowledge has the potential to add to ELP interpretation and implementation of subsequent change at school/cluster level, and trigger opportunities for validity studies (in relation to link with ELP scores and student achievement levels).

Marking services and database analysis

There is currently a single marking service for the ELP which mean all results can be readily aggregated. The availability of a single database with ELP scores provides useful information to national sector groups and to the Ministry of Education about trends over time, and trends in response to major policy initiatives (e.g., professional learning and development contract changes). In determining next steps with marking services the ready access possible with use of a single ELP database needs to be considered; all available data are easily aggregated with fewer transactional costs required.

Recommendations

Given the conclusions and comments above, our recommendations are as follows.

Professional development and learning

- The lower ratings for educational leadership ensuring processes are in place in relation to Māori Success suggest that consideration should be given to ensuring that educational leadership can access professional development opportunities that enable teachers to develop the knowledge and skills necessary to provide quality teaching for Māori learners.
- The additional context challenges and lower ratings of effectiveness on a range of items in the ELP for secondary and composite school educational leadership suggest targeted support for educational leadership at secondary and composite level.
- Lower ratings in the area of teacher learning and development again suggest targeted support for educational leadership in this area.
- Consideration of support for principals and educational leadership generally in the area of identifying and resolving conflict quickly and fairly.
- Consideration of targeted support for lower decile schools in the area of educational leadership in a way that reflects their current perceptions of supports and barriers in a low-decile context, in relation to pedagogical leadership.
- Consideration of targeted support for educational leaders to have systems and routines in place to ensure the high-level planning is actually reflected in day-to-day implementation of systems and processes within schools.
- Opportunities for professional development facilitators in the area of Leadership and Assessment to hear about and respond to the national sample results, and to the benchmarking exercise described in this report.

Where next with the ELP?

- Promulgation of use in a range of contexts (e.g., through Ministry funded Leadership and Assessment contracts, and First Time Principal contracts).

- Retention of a single scoring system to maintain a single database capable of reporting on trends over time at a national level so that there is a mechanism for monitoring trends and major policy initiatives in this area.
- Consideration of use in conjunction with student achievement outcomes to ensure the ELP is a valid tool in relation to predictors of improved student outcomes (perhaps by matching NCEA patterns of achievement with ELP score patterns, using the “schools like us” models and benchmarking models provided in this report, and at primary level by perhaps including in the National Monitoring cycles in some way).
- Exploration of the ELP’s specific validity in relation to secondary school educational leadership.
- Ensuring the ELP continues to be reflective of current evidence by undertaking a revision within the next 5 years.

1. Introduction

Background to the development of the Educational Leadership Practices Survey (ELP)

Why an ELP for New Zealand?

The original impetus for the development of the ELP survey came from a shared realisation of the Ministry of Education, sector group representatives (New Zealand Principals' Federation [NZPF], New Zealand Educational Institute [NZEI], Post Primary Teachers' Association [PPTA], Secondary Principals' Association in New Zealand [SPANZ] and New Zealand School Trustees Association [NZSTA]) and leadership professional development providers, who came together in a Ministry of Education-convened leadership external policy group,[1] that New Zealand needed a practical tool for principals to support their educational leadership work and continued professional growth. Such a tool could also serve as an ongoing formative tool to monitor the effectiveness of the Ministry of Education's professional learning plan at the system level. With NZCER capturing all the data from these surveys electronically, in a secure "data bank", an opportunity was available to use the schools' data at an aggregate level to provide ongoing trend data at a national level for Ministry and sector use.

In the New Zealand self-managing schools context, principals have both more opportunity to lead their schools, and more challenge from their administrative responsibilities and the varied and high nature of their workload (Ministry of Education, 2007; Robinson, Hohepa, & Lloyd, 2009; Wylie, 2010). The survey was developed to reflect the vision for educational leadership set out in the *Kiwi Leadership for Principals* (Ministry of Education, 2008) and the dimensions for effective educational leadership practices described in the *Educational Leadership Best Evidence Synthesis (BES)* (Robinson et al., 2009). It does not include all aspects of school leadership, such as some of the administrative roles. It focuses on those that the existing evidence bases showed were most clearly associated with valued student outcomes.

There were four criteria used to develop ELP survey items. The items needed to:

- be relevant to the leadership of teaching and learning
- express desirable leadership practices (and so include aspirational items that can show change over time)
- link to valued student outcomes

[1] This leadership external policy group met regularly between 2007 and 2009.

- be likely to show a range of scores (i.e., able to differentiate different levels of educational leadership and again, therefore, be able to show change over time).

The New Zealand student context for educational leadership has been an important consideration in the development of the survey items. Robinson et al. (2009) outline the key New Zealand student contextual challenges in their BES:

- In a range of international studies, the mean test scores of New Zealand 15-year-olds are generally high in reading, science and mathematics. However, the same data also reveal a wide disparity between low and high achievement. The system is underperforming for some of our most rapidly growing young populations—Māori and Pasifika students. Educational leadership must prepare **all** students for the future. Items that allow exploration of educational leadership for diverse learners were therefore important.
- A marked improvement in education provision to support Māori achieving success as Māori is required, hence the inclusion of items specifically related to Māori success.
- There is a need to strengthen valued social outcomes—the same international studies indicate that some of our students do not feel safe at school. This survey includes an aspect, therefore, on a safe and orderly environment.

Development and use of the ELP to date

The development of the ELP is one component in establishing practical links between the Kiwi Leadership for Principals framework, the powerful Leadership BES findings and educational leadership practice in schools. Two of the most respected architects of system change in education, Ben Levin and Michael Fullan, noted in their foreword to the *School Leadership and Educational Outcomes BES* that:

> Having such a high-quality review is an important accomplishment, but it is not enough ... Although well-written and full of useful information, it is likely that relatively few practitioners or policy makers will read the report in its entirety ... We also know that written documents, while important in generating knowledge from research, are insufficient to create changes in behaviour. The 'mobilisation' of this knowledge and its transformation into changed and better schools, will require much effort beyond the publication of this important report. (Levin & Fullan, as cited in Robinson et al., 2009, p. 14)

In April 2009 Viviane Robinson and Cathy Wylie developed a draft survey tool during a short consultancy contract with the Ministry of Education, with commentary and feedback from the rest of the Ministry of Education leadership external policy group. This led to a paper version of the draft ELP that was trialled by NZCER with a cross-section of 37 schools in the last week of term 3, 2009. The results of this trial were analysed by NZCER and discussed with Viviane Robinson, and then with the Ministry of Education leadership external policy group, leading to the development of the survey, scales and reporting methods that are used in this report.

We wish to thank Andrew Porter (University of Pennsylvania), Joseph Murphy (Vanderbilt University), Ellen Goldring (Vanderbilt University) and Stephen N. Elliott (Arizona State University), the authors of the *Vanderbilt Assessment of Leadership in Education*, for use of portions of the Vanderbilt Assessment of Leadership in Education (VAL-ED) structure and VAL-ED item stems.

The first use of the ELP survey was in relation to the Experienced Principals Development (EPD) programme, which ran from August–September 2009 to the end of 2010. In September 2009 the Ministry of Education contracted NZCER to administer the ELP and provide schools and the EPD providers with reports, all electronically. The ELP was used as part of the needs analysis for the principals taking part in the EPD, and to provide a baseline picture of school leadership at the start of the EPD (Wylie & Hodgen, 2010). EPD schools were also to undertake the ELP again at the end of 2010, to provide a picture of any shifts in their leadership practices over the course of their EPD work. These shifts are reported in NZCER's report on the aggregate pattern of shifts in educational leadership after principals completed the EPD programme (Wylie, Brewerton, & Hodgen, 2011).

The ELP was well received by the principals taking part in the EPD, and the professional development providers who worked with them. It is being used in the First Time Principals' programme, and been made available for individual and school cluster use from late 2010, by a range of schools and clusters. NZCER provides electronic administration and reporting for a small fee ($200 per school), with school results able to contribute to the national NZCER database that safeguards school identity while allowing national patterns to be tracked over time and reported periodically. It is envisaged that the ELP will also be further developed in the light of new evidence and need.

The electronic scoring, provision of customised reports back to schools and capacity to collate data and trends over time and within one database is an example of a "smart tool"[2] that can provide feedback both to schools and the education system as a whole over time. Robinson et al. (2009) identify the use of "smart tools" as an important influence in effective educational leadership and enhanced student outcomes.

In 2010, a variation to the EPD Ministry/NZCER contract provided the opportunity for additional ELP administration and inclusion of ELP results from other projects that had used the NZCER electronic administration of ELP to create a national sample. This sample could provide insight into educational leadership practices across the compulsory education sector in New Zealand and enable tracking over time of the national profile of educational leadership, by establishing 2010 baseline profiles that could be compared with future national samples (e.g., in 2015). In addition,

[2] Smart tools are *thinking* and *learning* tools that support transfer of knowledge. The underpinning idea is drawn from complexity theory, specifically the idea that complex systems learn and adapt in response to feedback generated by the activities of the system at work. Smart tools generate feedback that allows their users to: inquire into their own and others' practice; lead change; analyse and respond to data sets; think smarter and therefore act more effectively; and evaluate, monitor and assess.

it could be used to establish some national benchmarks that schools could use for their own formative analysis of their educational leadership practices. This report describes this national sample and benchmarking.

Description of the ELP survey contents

The ELP comprises the following scales to assess educational leadership:

- *Goal Setting*
- *Strategic Resourcing*
- *Curriculum Quality*
- *Quality of Teaching*
- *Promoting and Participating in Teacher Learning and Development (Teacher Learning)*
- *Safe and Orderly Environment*
- *Educationally Powerful Connections with Families, Whānau and Community (Teaching Learning)*
- Māori Success
- Principal Leadership.

The ELP consists of 66 items that ask teachers and principals to rate the effectiveness of a school's educational leadership in ensuring a particular practice associated with valued outcomes for students (related to the dimensions above). In addition, both teachers and principals are asked to rate the effectiveness of the principal (16 items) in relation to a range of items the literature reports as important.

Principals and teachers are asked the same questions for each of these nine aspects, and to also give their sources of evidence for their judgement (personal observations, school documentation or other sources). Principals also answer a set of questions related to the school context for pedagogical leadership (18 items). Teachers answer a short set of questions related to their morale and workload. Both also answer questions about their experience and demographic characteristics (7 items). These questions on school context, morale and workload, work experience and demographic characteristics, along with information on school characteristics (type, size and socioeconomic decile) are reported descriptively, and also used in analysis of the nine scales of school leadership in our reports on the ELP.

The overall leadership scale used in the ELP is constructed from the seven scales of school leadership given in italics above. This enables the survey to be used to provide an overall school leadership level for each school. In the trial and calibration of the final ELP ready for use in the EPD we found that each of these seven aspects had high internal reliability (alphas of 0.81 to 0.88)—the items in each scale "hung together" well, and were measuring dimensions of the same underlying construct. The seven scales also have a high level of intercorrelation, and a Rasch analysis confirms that the seven scales can be treated as different aspects of a single underlying meta-construct—overall leadership practice.

The items used to make the Māori Success scale were originally included in these seven scales (and are asked within the survey, rather than as a separate set of questions). But we found that these items did not contribute to the internal reliability of the scales they were in, and in fact "hung together" well as a separate scale. The Principal Leadership scale was a separate set of questions from the start, and is not included in the overall school leadership scale since it focuses on the principal alone.

The ELP therefore comprises seven scales contributing to an overall leadership scale, a separate Māori Success scale and a Principal Leadership scale. Each is measured and reported in a different way and contributes to the overall picture of educational leadership both at a school level and at a system level. (See Figure 1 and also Appendix 1 for examples of the information provided to schools.)

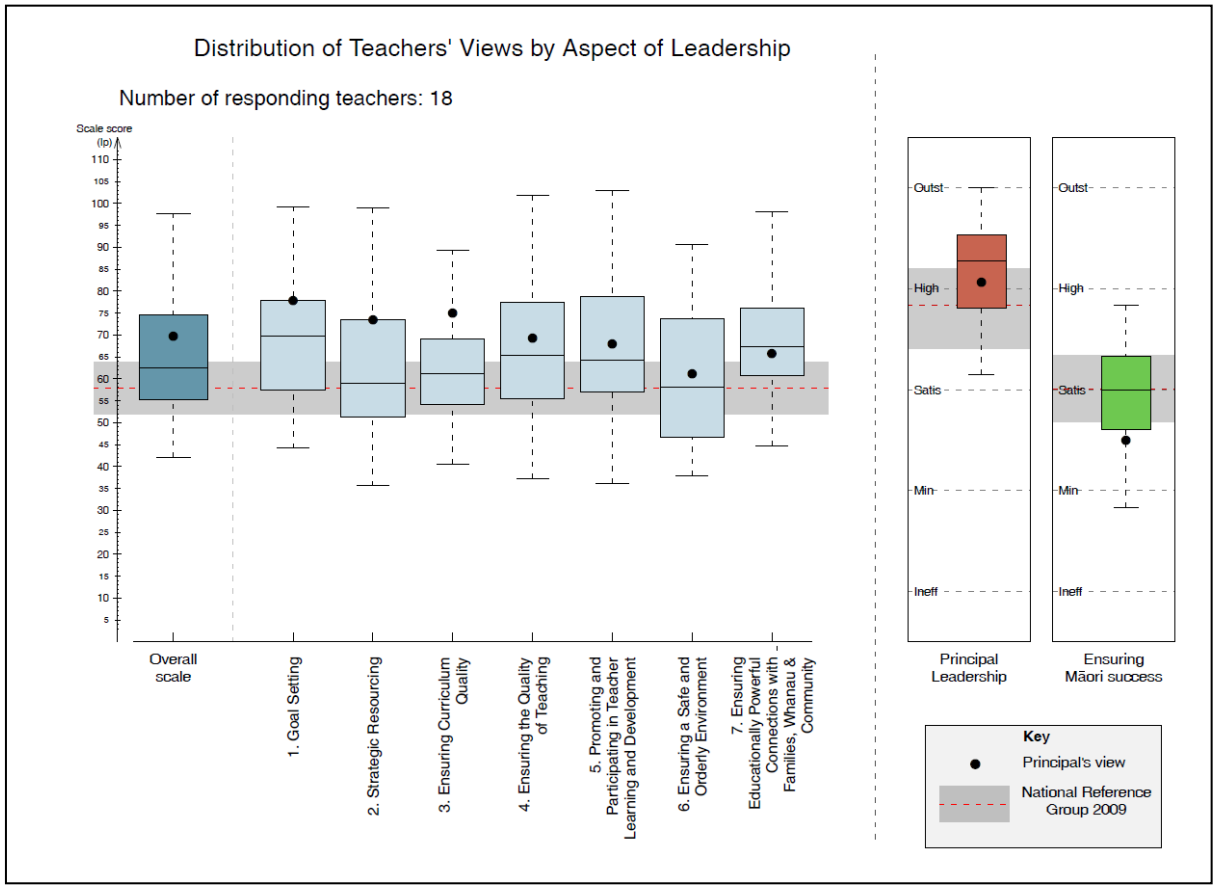
As noted above, key evidence sources for the items were the Best Evidence Synthesis on *School Leadership and Student Outcomes* **(**Robinson et al., 2009), along with the Kiwi Leadership for Principals framework materials. The evidence base for each of the scales precedes the description of schools' ratings for effectiveness on those aspects. We also acknowledge the prior work of Viviane Robinson and Cathy Wylie and the current doctoral work of Linda Bendickson that also contributed to the thinking behind the items.

We also acknowledge a licensing arrangement with Vanderbilt University and credit to the VAL-ED authors—Andy Porter, Ellen Goldring, Joseph Murphy and Stephen N. Elliott—for use of portions of the VAL-ED assessment structure and selected item stems.

ELP reporting

Individual school reports give principals their own and the average teacher response at that school. This enables principals to compare their responses with their teachers', as well as see the picture for their school as a whole. Figure 1 is an example of the high-level reports provided for schools. The range of teacher responses is indicated by box plots for each of the nine scales, and the fuller range of reports to schools is contained in Appendix 1.

Figure 1 **Example of high level-report provided to a school principal**



Focus of this report

In this report we focus on:

- reporting the patterns of distribution—norms—for each of the seven scales that make up the overall scale, as well as the Principal Leadership and Māori Success scales that form part of the ELP also, along with principals' views of the pedagogical contexts (barriers and support) for their educational leadership. Norms provide a basis for comparing the individual school with a national group of schools. That is, we provide a picture of how schools rate effectiveness in ensuring various aspects of educational leadership in relation to other schools
- examining difference in patterns of scores related to school characteristics, and presenting scaled scores related to school characteristics (size, decile and type). The 2010 ELP contract with the Ministry of Education required a "schools like us" section to provide a sense of different leadership patterns across school type, size and decile
- reporting on the outcomes of a benchmarking exercise to establish some suggested benchmarks for different levels of educational leadership performance on the ELP. By benchmarking we mean comparing a school's ratings in ensuring effectiveness in aspects of

educational leadership with best/effective practices (comparing a school's rating with the norms simply gives a comparison with the aggregate picture of other schools' ratings)
- making recommendations about future administration, use, analysis and scoring for the ELP
- making suggestions for further professional development for educational leadership.

Types of analyses and the measurement metrics used in this report

To fulfil the various purposes of this report, a range of different metrics and analyses are used, as described next.

Distribution patterns: The national norms

We present the range of scores for each aspect of the ELP first for **schools**; that is, the score per school using teacher scores only. To provide frequency descriptions (e.g., the proportion of schools that give high scores for a particular item on a scale) we have used the average teacher score per school.

After we look at the patterns of educational leadership using schools as our unit of analysis, we report the data from **teachers** and **principals**, aggregating teachers as individuals rather than within schools. This allows us to see proportions of teachers working with various levels of school educational leadership as well as get a sense of the distribution of scores for principals.

Given the formative nature of the ELP, we also present some data that allow school leadership teams to look in detail at the items that appear to differ in relation to school contexts of type, size and decile.

We also present scale score averages for each of the seven educational leadership scales that make up the overall scale according to school type, size and decile, to allow a "schools like us" comparison (using the Rasch scaling described below). While the Māori Success and Principal Leadership scales do not have Rasch scales, we have undertaken analyses that still allow a "schools like us" comparison of these two scales.

Reporting school scores

In baseline report on the ELP (Wylie & Hodgen, 2010) we considered using teacher and principal ratings together, but decided against this since a simple aggregation would mix two different perspectives, and potentially compress the between-school differences. And, consideration of weighting the principal's perspective raised the question of what weight to give the principal's ratings—to which there is no clear-cut answer, meaning that different weightings could yield different pictures. What we have done in Section 2 is to report both school average and the principal/teacher ratings. Our main unit of analysis for this overall picture is the individual school. For this, we have used teacher ratings only. To provide frequency descriptions (e.g., the

proportion who score highly for a particular scale item), we have used the average score for each school. In Section 2 we have also reported teachers' individual scores, without aggregating them per school, since it is useful to see patterns for both the schools and for teachers as a cross-school group (one can think of the frequency data here as showing what proportion of teachers work in schools with a particular level of educational leadership in the scale of interest).

Calculating school scores

To take a closer look at the items that make up each scale, we used the mean teachers' score for each school. Since means compress the range of views in a school, and often do not provide whole numbers (e.g., the mean of the five teachers who score a particular item 5, 3, 4, 4 and 5 is 4.2), we could not simply translate each school's mean into the existing whole-number 5-point scales. We plotted the distribution of school scores on a range of items for different scales, and used this distribution to suggest these cut-off points to assign school average scores to an equivalent whole-number 5-point rating. So, for the nine scales outlined above, we have converted school averages of 1–1.75 to a 1; 1.751–2.9 to a 2; 2.91–3.5 to a 3; 3.51–4.25 to a 4; and 4.251–5 to a 5. We have then been able to place each school on a 1–5 scale, comparable to the scale we used in the survey, where 1 = not at all effective level, and 5 = outstandingly effective level. The cut-points we have decided to use are not completely arbitrary—they were not plucked from the air—but slightly different cut-points could be used based on the same distribution that would give a slightly different distribution of schools in each level of school leadership.

Principal teacher comparisons

In Section 2 we use straight percentages of principals and teachers reporting the different levels of ensuring effectiveness for each of the scales, and for pedagogical context and teacher work experience and morale.

Rasch scaling

We have used Rasch scaling to develop the overall educational leadership scale and its seven different scales (the aspects described in italics on page 4).

In Rasch scaling, item calibration is an initial step, in order to create a scale. In calibrating the ELP items, the smaller the proportion of highly positive responses (ratings of 4 or 5), the higher the difficulty of an item and hence the higher the item's scale location. Once item locations are scaled, the school/principal/teacher responses are measured on the scale. As a result, we are able to place schools on a single scale for each scale, or for the overall scale, even though there are multiple items within each scale. This is the unit of analysis used in Table 2 (Section 2), in most of the analyses presented according to school characteristics in Section 3 and in the comparison with benchmarking in Section 5.

Multilevel modelling

In the multilevel models we have undertaken to gauge the factors that might account for differences between school scores on each of the scales (school size, type, decile), we have been able to use each teacher's individual ratings ("nested" or "clustered" within the school).

Benchmarking approach

In addition to normative information we have worked to establish what in fact constitutes effective leadership with respect to the ELP. A distribution of scores gives a picture of educational leadership across the country but it cannot on its own report on "how good is good enough". Ideally, one would establish benchmarks by linking ELP data with other data on school functioning and effectiveness, such as student outcomes and engagement. What we were able to do was a benchmarking exercise, using a panel of leadership experts to provide their perspectives on which ELP items are critical for educational leadership to function effectively (fully described in Section 5).

Sampling approach to collect a national sample

ELP survey data in this national sample include two general sets of data: 2010–11 data and 2009 data when we could not get 2010 data.

Group 1 data comprised approximately two-thirds of the sample—schools that completed the survey in 2010–11. Some of these schools had done the survey at the conclusion of the EPD programme, and some of these schools were newcomers, added to the sample to ensure the national sample was not underrepresented in particular areas such as school type and decile. There were challenges to data collection for this sample, given the time of the year it began, and the events in Canterbury, hence the wide time period for data collection.

Group 2 data accounted for one-third of the sample and were from 2009 ELP survey administration, where those schools had not completed the survey in 2010 (at the end of the EPD programme).

Characteristics of the ELP national sample

The final total sample provides a good balance of schools across deciles when compared with the national distribution of schools. The sample is slightly weighted toward larger schools and secondary schools; this is a better skew than being weighted to smaller schools, because it is important to have a sufficient sample of more complex contexts of educational leadership (and more students are served by these schools) (Wylie & Hodgen, 2010) (see Table 1).

Additionally, the oversampling in these areas allows us to have sufficient schools in each of the groups to provide scaled scores for the "schools like us" part of this report. Table 2 provides the distribution of the national ELP sample across school type, size and decile.

We further checked the representativeness by cross-checking length of experience as a principal in the ELP national sample with the NZCER 2010 National Survey principal data. The sample for the ELP contains principals with more experience than those samples in the primary national survey in 2010 and secondary survey in 2009. The most notable difference is that the ELP sample contains less principals with less than 3 years' experience as a principal (see Table 1), as we did not include First Time Principal professional development programme participants in the sample.

Table 1 **Experience as a principal: ELP sample compared to NZCER primary and secondary national survey samples**

Length of time	Primary survey 2010 %	Secondary survey 2009 %	ELP sample (%)
Under 3 years	19	24	6
3–5 years	22	14	12
6–10 years	16	35	31
11–15 years	11	16	22
15+ years	33	10	28

One caution about the national sample for the ELP is that, while it reflects quite well the range of school characteristics across New Zealand, it is **not** a randomised national sample. As noted previously, it is a sample of EPD participants, augmented by data from additional clusters of schools.

Table 2 **Profile of schools in the national ELP sample**

	MOE national total for state and state-integrated schools (2009) (n=2,486)	ELP schools %	(n=369)
School type			
Contributing primary	32	31	131
Intermediate	5	11	45
Full primary	43	32	136
Secondary	13	17	73
Special schools	2	3	12
Composite	4	6	24
School size (U grade)			
U1 & U2	26	13	56
U3 & U4	35	33	140
U5 & U6	29	36	153
U7+	10	17	73
Location			
Urban	29	20	335
Rural	71	79	86
Socioeconomic decile			
1–2	21	19	82
3–4	20	22	92
5–6	20	20	84
7–8	19	18	75
9–10	19	21	87
Proportion of Māori students on roll			
Less than 8%	14	16	68
8–14%	21	22	94
15–30%	30	32	136
31%+	34	29	121

Note: Percentages may not add to 100 because of rounding.

2. National norms on the ELP

This chapter reports the norms, or distributions, for the ELP overall scale, the seven separate scales that make up that scale, the Māori Success scale and the Principal Leadership scale. It also reports the picture provided by principals of the context for the school's pedagogical leadership, and teacher reports of their morale and workload. As noted in the introductory section, we look at three sets of ratings:

- schools (the combined aggregate score for staff at a school—across schools **excluding principals**)
- principals
- teachers' scores (not aggregated by school).

Aside from the table of means and standard deviations for the scales in Table 4 below, the descriptive metric used in this section is the percentage giving their school leadership a highly or outstandingly effective rating in ensuring that some aspect of school practice happened. For the sections on teacher work experience/morale and the pedagogical context for educational leadership, the descriptor is the percentage agreeing or strongly agreeing with items.

Overall means on the scales and overall scale

Generally, the means and standard deviations across the seven scales that make up the overall ELP are relatively similar, due to the intercorrelation between the seven scales. On average, principals are a little more conservative in their ratings than schools. In three scales, principals' scores are a little higher than staff's. These scales are Strategic Resourcing, Ensuring a Safe and Orderly Environment and Māori Success. The Ensuring a Safe and Orderly Environment scale had the highest mean for both schools and principals.

Teachers' mean scores also tended to be lower than principals' mean scores (aside from Māori Success and Principal Leadership). Table 4 below provides the mean scores and standard deviations across scales for school, principal and teacher ratings. The means and standard deviations in Table 4 are derived from the Rasch scaling described in Section 1 for the seven scales that feed into the overall scale, and on ratings of 1–5 on a Likert scale for the Māori Success and the Principal Leadership scales. The standard deviations indicate the spread of ratings around the mean, and these show that principal ratings were more varied than schools', but less varied than teachers' ratings.

Scales most highly rated (highly or outstandingly effective) were Principal Leadership (69 percent), Ensuring a Safe and Orderly Environment (58 percent) and Goal Setting (55 percent).

The scale least frequently highly rated was Māori Success, where just 17 percent of schools rated this as highly or outstandingly effective (see Table 3 below).

Table 3 **Percentage of schools rating school leadership as highly or outstandingly effective in ensuring practices across the nine scales**

	Goal Setting	Strategic Resourcing	Curriculum Quality	Quality of Teaching	Teacher Learning	Ensuring a Safe and Orderly Environment	Ensuring Educationally Powerful Connections with Families, Whānau and Community	Māori Success	Principal Leadership
Percentage of schools	55	42	46	42	37	58	45	17	69

Table 4 **Mean scores and standard deviations across scales for school, principal and teacher ratings**

	Goal Setting	Strategic Resourcing	Curriculum Quality	Quality of Teaching	Teacher Learning	Ensuring a Safe and Orderly Environment	Ensuring Educationally Powerful Connections with Families, Whānau and Community	Māori Success	Principal Leadership
School average item score	59.7	59.5	59.5	58.6	59.1	59.7	59.1	3.1	3.8
School standard deviation	10.8	10.4	10.8	10.2	11.3	10.8	10.2	.5	.6
Principal average item score	*57.7*	*60.5*	*57.2*	*58.4*	*58.3*	*61.7*	*56.2*	*3.8*	*3.1*
Principal standard deviation	*13.3*	*12.1*	*13.9*	*12.4*	*12.6*	*11.5*	*13.1*	*.7*	*.4*
Teacher average item score	56.9	56.6	56.9	56.8	56.9	57.1	57.2	3.7	3.0
Teacher standard deviation	16.5	15.8	16.3	15.8	16.5	16.8	16.7	.9	.8

The scales in detail

Goal Setting scale

Evidence base

Robinson et al. (2009) describe this dimension as:

> … the exercise of leadership though the setting and communicating of goals for teacher and student learning. (p. 40)

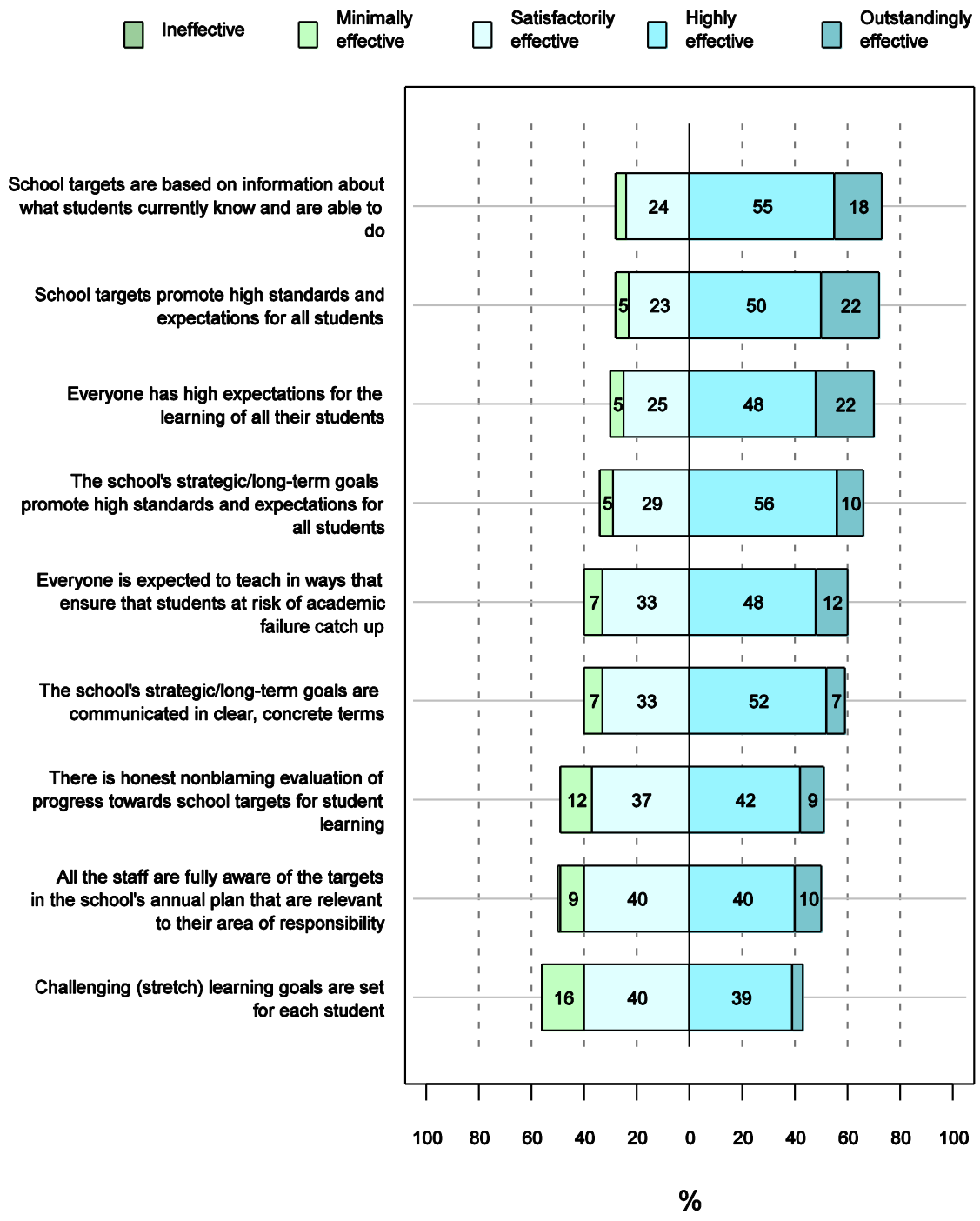
Leaders need to be sure the importance of goals is understood, that the goals are clear and that staff are committed to the goals:

> There is evidence that the content of goals may be as important as the process of goal setting: leaders need to know *what* goals to set as well as how to set them. In high performing schools, there was a stronger emphasis on academic goals, though this was not incompatible with a further emphasis on social goals. (Robinson et al., 2009, p. 41)

School views

On average, 55 percent of schools rated their school leadership on the Goal Setting scale items as highly or outstandingly effective. This ranged from 73 percent of schools rating their school leadership in *ensuring that the school targets are based on information about what students currently know and are able to do* as highly or outstandingly effective, to 43 percent of schools rating their school leadership as highly or outstandingly effective in respect to *ensuring that challenging (stretch) learning goals are set for each student* (see Figure 2). There appears to be more confidence about the role of leadership in relation to schools' broader guiding frameworks than about the embedding of the goals into ongoing use and evaluation for teachers' work and in relation to every student.

Figure 2 **Goal Setting scale: School ratings**

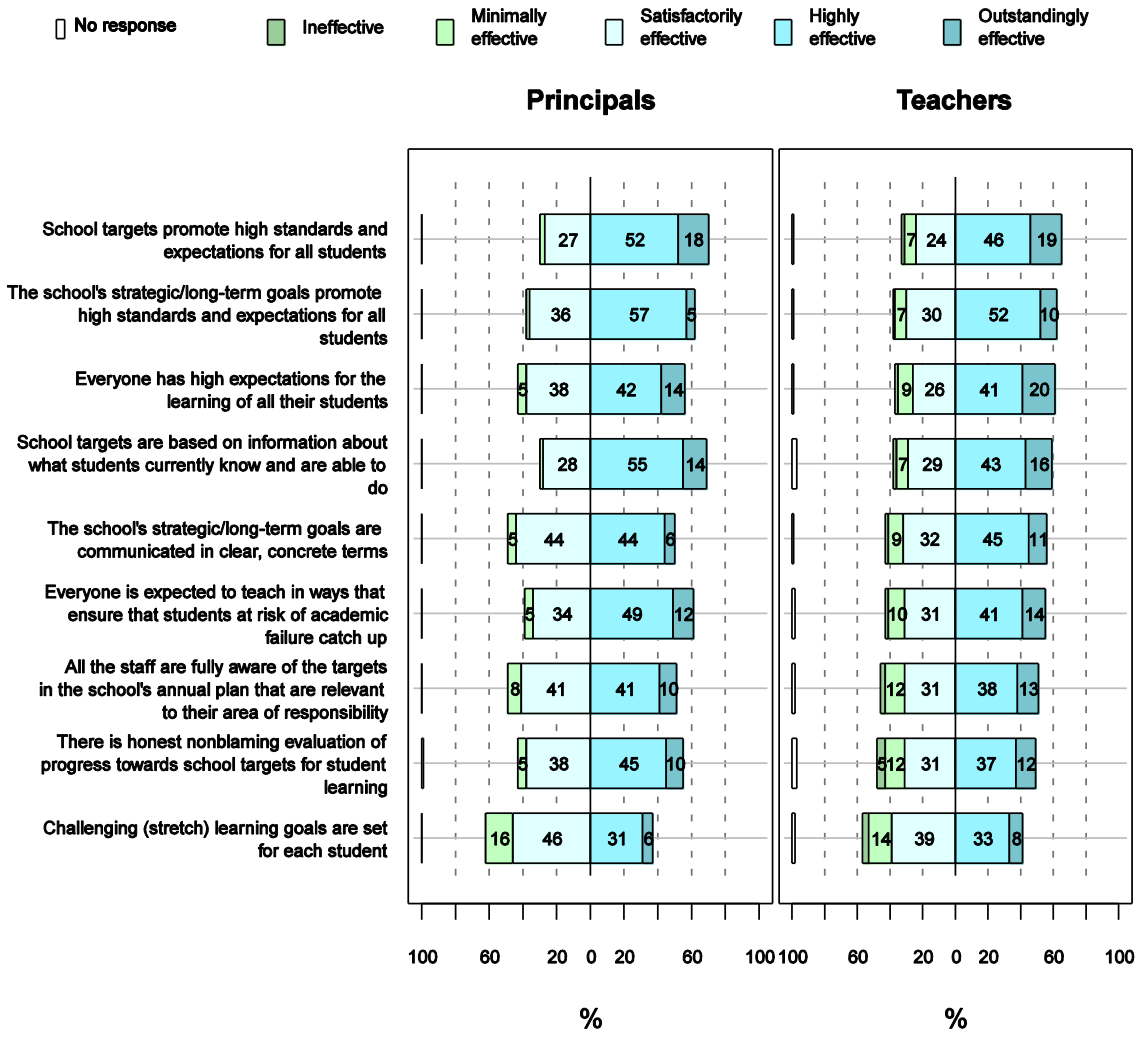


Goal Setting scale: Teacher and principal views

There was minimal difference between the overall ratings for principals and teachers for Goal Setting. When teacher scores are aggregated across schools (rather than looking at the mean score per school) there are some small differences in the order of effectiveness of items and differences between principals and teachers. Eighty percent of principals and 65 percent of teachers rated their school's leadership as highly or outstandingly effective in *ensuring school targets promote*

high standards and expectations for all students. At the other end of the continuum, just 40 percent of principals and 41 percent of teachers rated school leadership in *ensuring challenging (stretch) learning goals are set for each student* as highly or outstandingly effective (see Figure 3).

Figure 3 **Goal Setting scale: Principal and teacher ratings**



Strategic Resourcing scale

Evidence base

The questions in this scale cover the school leadership in relation to aligning resources with pedagogical purposes, rather than securing resources generally (e.g., fundraising etc.). Robinson et al. (2009) note a significant impact on student outcomes from this alignment.

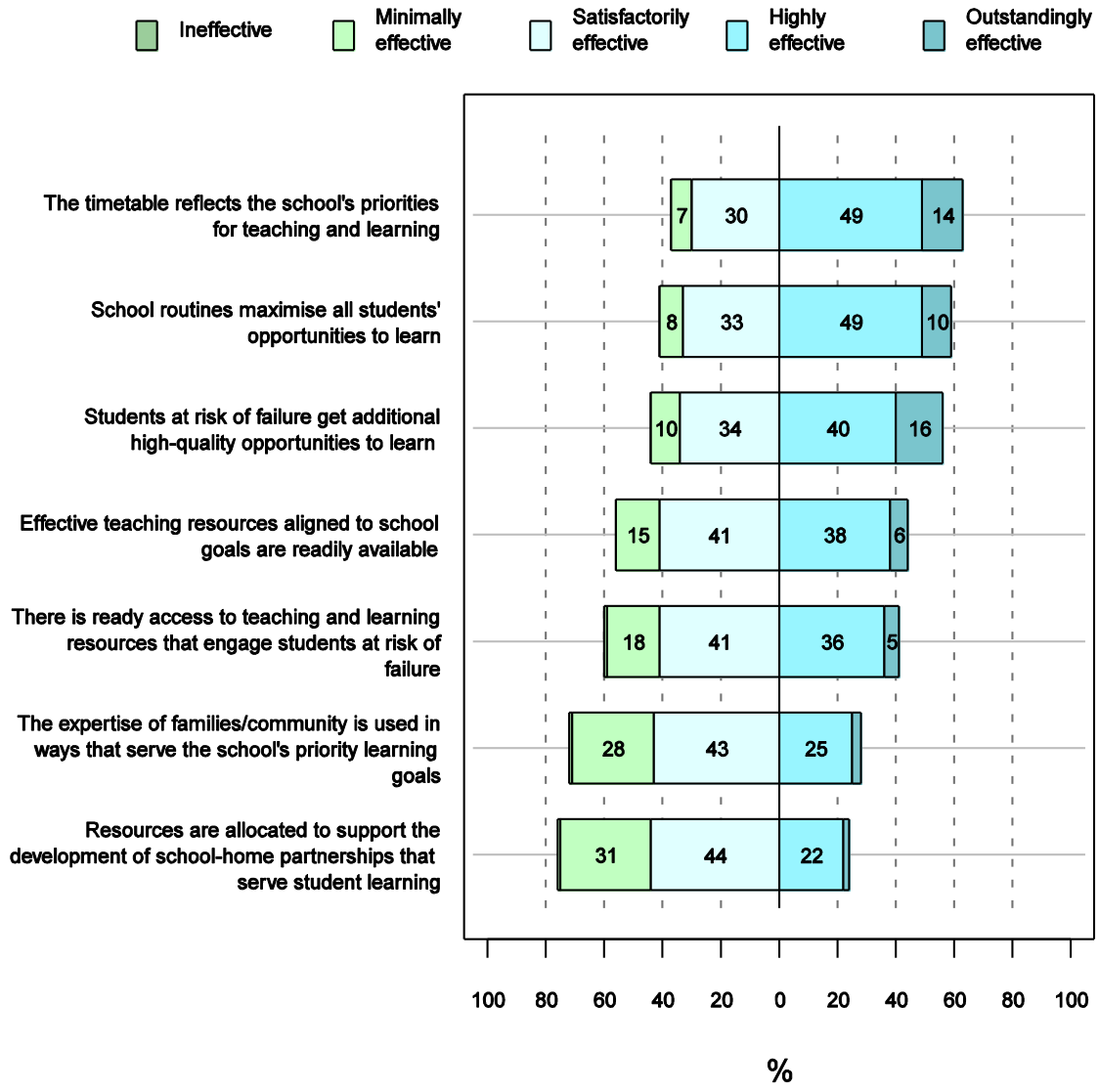
School views

Ratings of school leadership effectiveness were highest in relation to ensuring strategic use of time and teaching resources but lower in relation to the use of community resources to support student learning and school goals.

On average, 42 percent of schools rated their school leadership as highly or outstandingly effective in ensuring the practices included in the Strategic Resourcing scale. Ratings were highest in terms of *ensuring that the timetable reflects the school's priorities for teaching and learning* (63 percent). Just one-quarter of schools rated their school leadership as highly or outstandingly effective in *ensuring that the expertise of families/community is used in ways that serve the school's priority learning goals,* and *ensuring that resources are allocated to support the development of school–home partnerships that serve student learning.* See Figure 4 below for a detailed picture.

`

Figure 4 **Strategic Resourcing scale: School ratings**

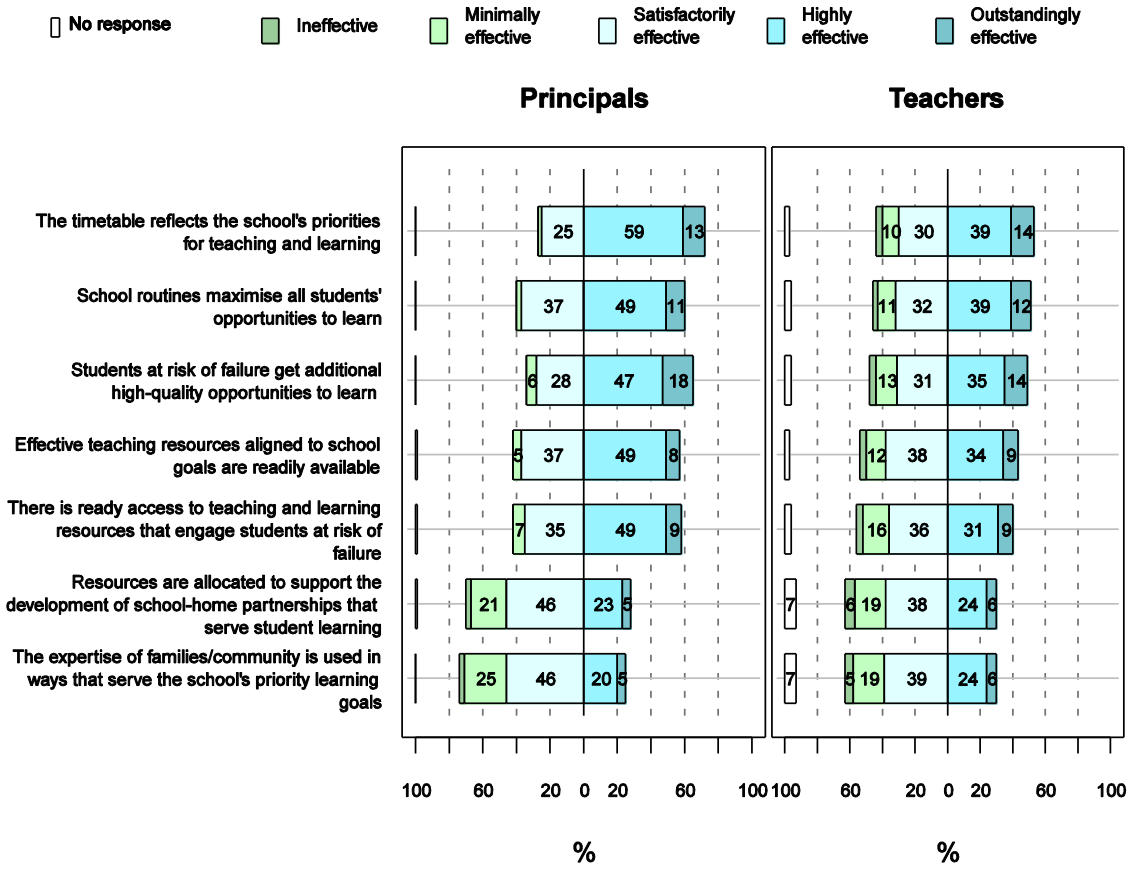


Teacher and principal views

Ratings of school leadership as highly or outstandingly effective were similar for both principals and teachers (aggregated across schools) for almost all the items for Strategic Resourcing, but teachers were somewhat more inclined to give "ineffective" ratings than principals on the items focused on use of school time and teaching resources.

Figure 5 gives the full picture of ratings for principals and teachers for Strategic Resourcing.

Figure 5 **Strategic Resourcing scale: Principal and teacher ratings**



Curriculum Quality scale

Evidence base

> Leaders in high performing schools are distinguished from their counterparts in otherwise similar low performing schools by their personal involvement in planning, coordinating, and evaluating teaching and the curriculum. (Robinson et al., 2009, p. 99)

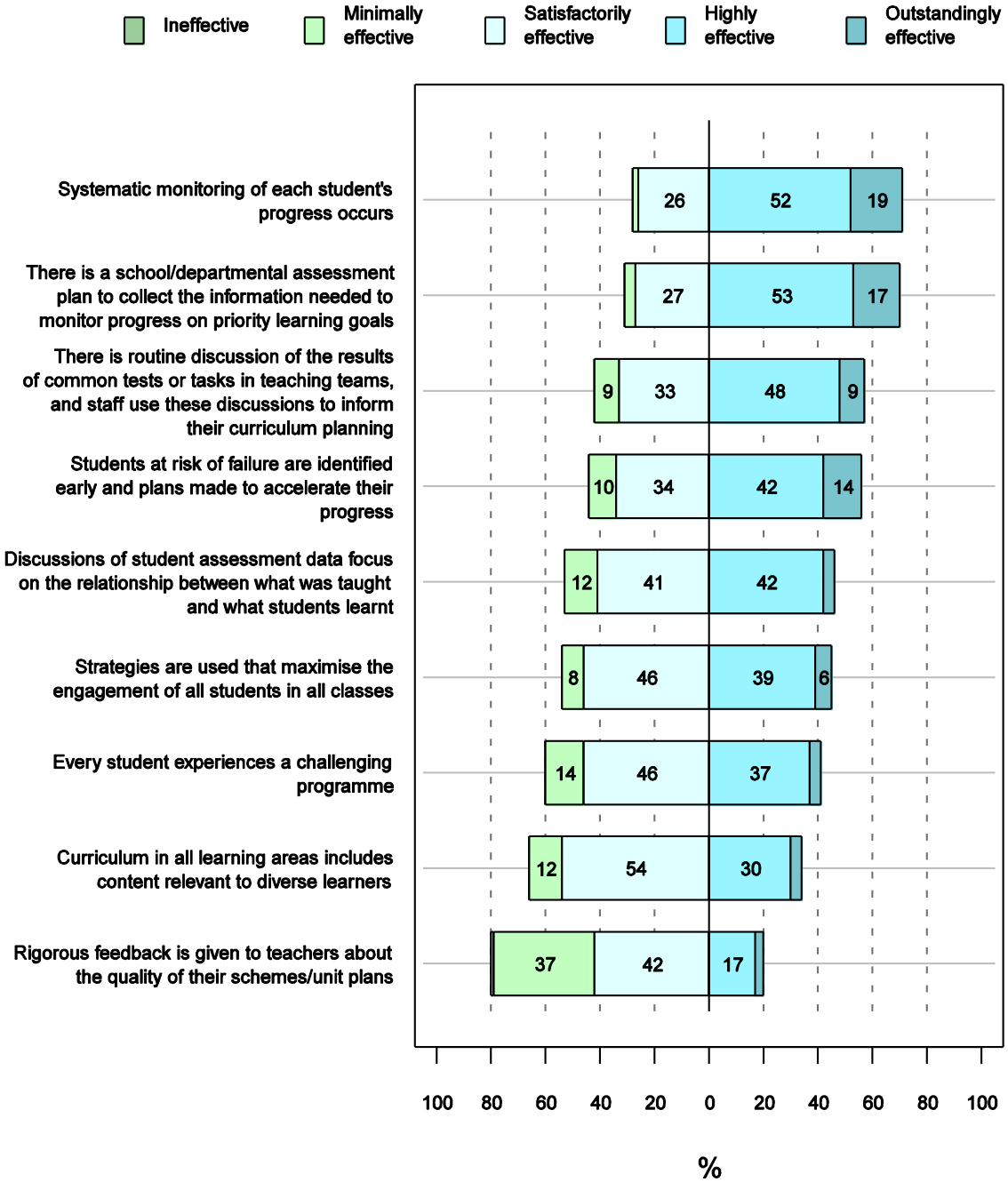
Educational leadership in ensuring curriculum quality is shown to be related to improved student outcomes when: leaders are actively involved in collegial discussion of instructional matters; leaders have active oversight of the instructional programme; and leaders are involved in classroom observation and subsequent feedback (Robinson et al., 2009, pp. 99–100). As noted in Robinson et al. (2009) and Notman (2010) there is more research based in primary schools available for analysis, and direct classroom observation may not be as strong an indicator at secondary school level.

School views

On average, 46 percent of schools rated their school leadership as highly or outstandingly effective in Curriculum Quality. Seventy percent or more rated their school leadership as highly

or outstandingly effective in *ensuring that systematic monitoring of each student's progress occurs*, and *ensuring that there is a school/departmental assessment plan to collect the information needed to monitor progress on priority learning goals*. Conversely, only one in five rated their school leadership as highly or outstandingly effective in *ensuring that rigorous feedback is given to teachers about the quality of their schemes/unit plans*. See Figure 6 below for the detailed information.

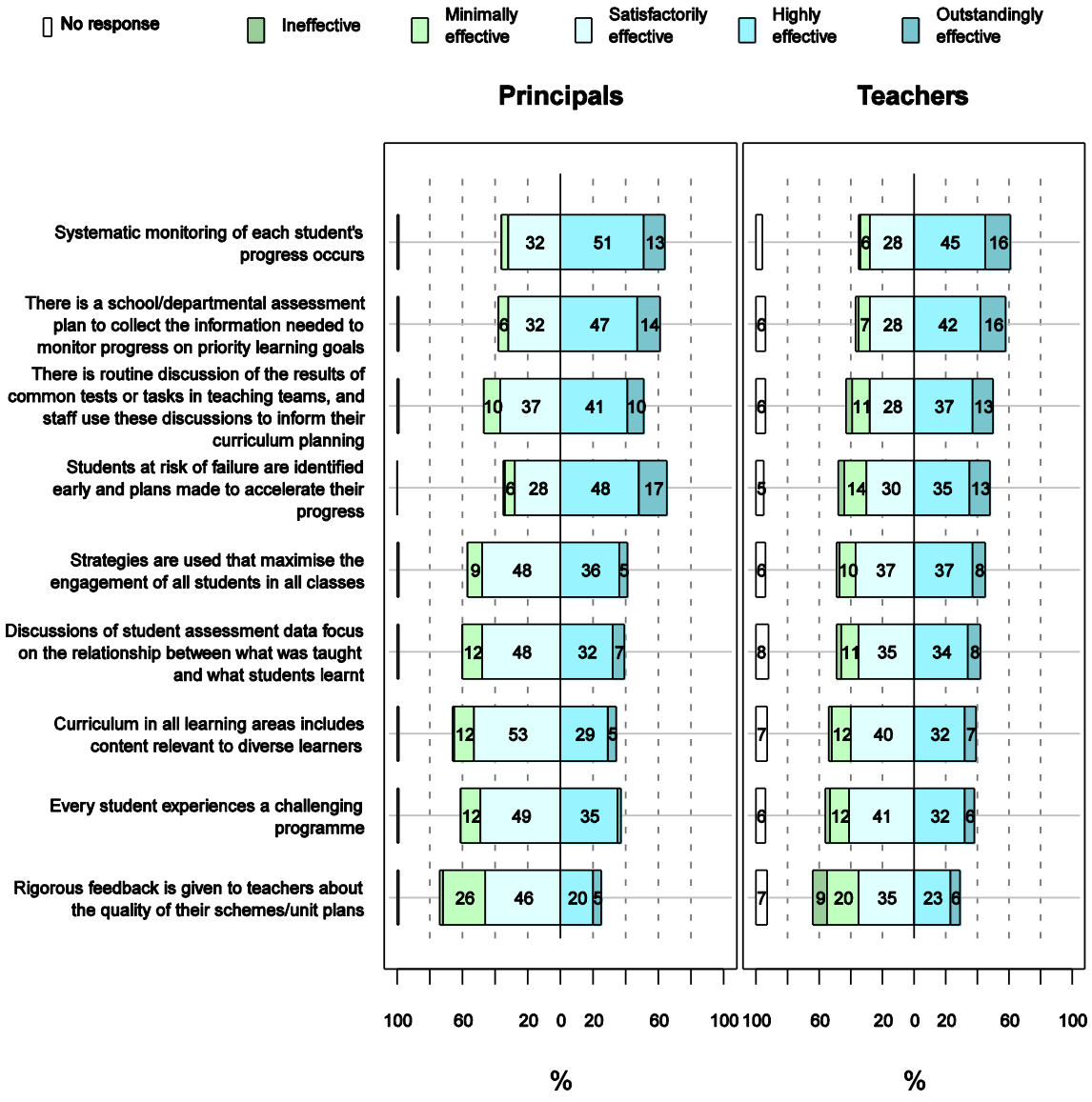
Figure 6 **Curriculum Quality scale: School ratings**



Teacher and principal views

When looking at the differences between principals' and teachers' ratings when teachers' scores are not aggregated by school, there was a high level of similarity in the spread of teachers' and principals' ratings for the Curriculum Quality items. Around 60 percent of teachers and principals felt that school leadership in *ensuring that systematic monitoring of each student's progress in their school occurs* was highly or outstandingly effective, and in *ensuring that there was a school/departmental assessment plan to collect the information needed to monitor progress on priority learning goals*. See Figure 7 below.

Figure 7 **Curriculum Quality scale: Principal and teacher ratings**



Quality of Teaching scale

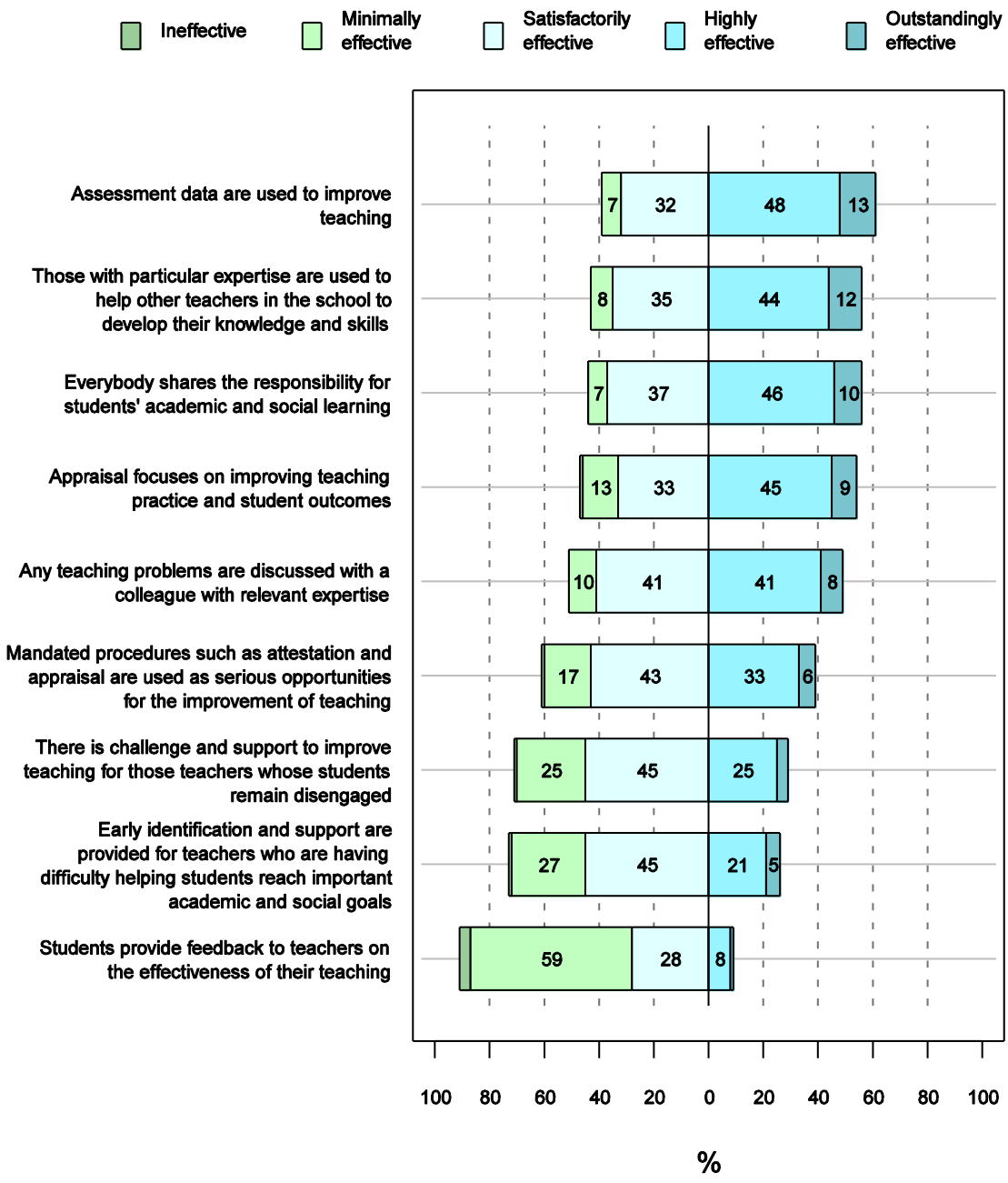
Evidence base

The Quality of Teaching scale draws on the same evidence as the Curriculum Quality scale. This scale has an emphasis on school leadership ensuring support and accountability for teaching, the way teachers use assessment and the way teaching issues are handled (through peer support and formal review) as well as provision for student feedback about effectiveness of teaching.

School views

There is a general trend that items related to school leadership ensuring collegial support practices are more highly rated than items about ensuring more formal individualised support and accountability, when individual teachers are identified as requiring improvement in their teaching. On average, 42 percent of schools rated their school leadership as highly or outstandingly effective in ensuring the quality of teaching. School leadership in e*nsuring that assessment data are used to improve teaching* was rated as highly or outstandingly effective by 61 percent of schools. Also of note is that school leadership in *ensuring that students provide feedback to teachers on the effectiveness of their teaching* was only rated as highly or outstandingly effective by around 10 percent of schools; more than 60 percent of schools rated this as minimally effective or ineffective. Figure 8 below provides the detailed ratings from schools.

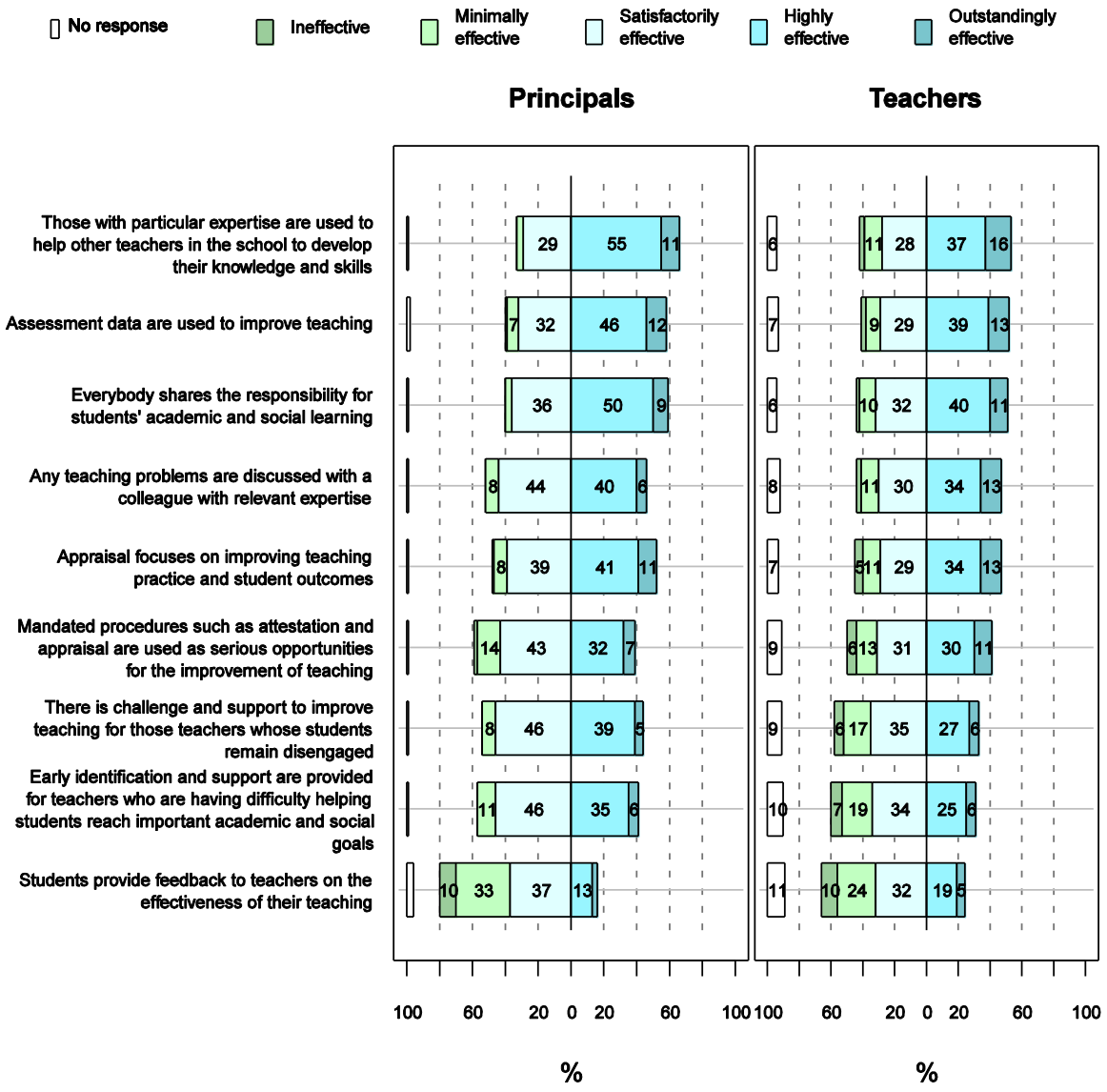
Figure 8 **Quality of Teaching scale: School ratings**



Teacher and principal views

Principals rated some items higher than teachers; particularly in items that were about systems to move support and advice to where it was needed: *ensuring that those with particular expertise are used to help other teachers in the school to develop their knowledge and skills*; *ensuring that appraisal focuses on improving teaching practice and student outcomes; ensuring that there is challenge and support to improve teaching for those students who remain disengaged;* and *ensuring that early identification and support are provided for teachers who are having difficulty helping students reach important academic and social skills.* See Figure 9 below.

Figure 9 **Quality of Teaching scale: Principal and teacher ratings**



Promoting and Participating in Teacher Learning and Development

Evidence base

Again, there are overlaps with the evidence related to the scale on a school leadership's role in teacher learning and development with the curriculum quality and quality of teaching scales. Robinson et al. (2009) note that their synthesis:

> … provides empirical support for calls for leaders to be actively involved with their teachers as the leading learners of their schools. Based on teachers' reports, leaders (usually the principal) in high-achieving and high-gain schools participate more actively in teacher learning and development than leaders in low-achieving or low-gain schools. They are also

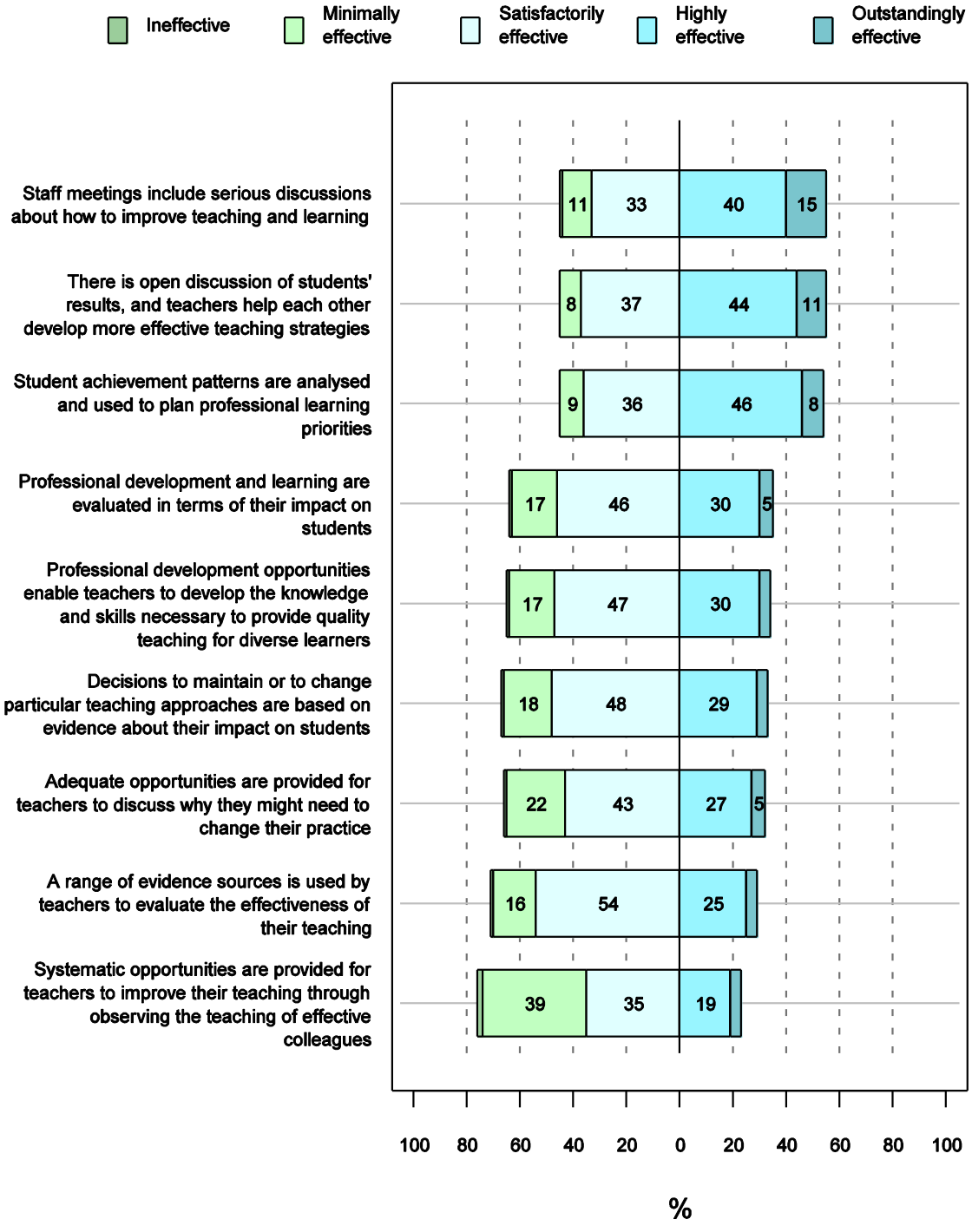
> more likely to promote and participate in staff discussion of teaching and teaching problems. (p. 101)

Robinson et al. (2009) report a large effect size for this dimension of educational leadership—the largest effect size of the dimensions identified in their synthesis. They also caution that in this dimension (along with some other dimensions) their evidence was largely from studies in primary schools. Notman (2010, p. 15) and Youngs (2011, p. 19) challenge some aspects of this dimension's evidence base in relation to large schools and secondary schools concerning actual within-class observations and direct participation in all professional development and learning. Certainly, the items in this scale emphasise more a systematic approach to providing opportunities for teacher learning, rather than direct principal involvement in the practices.

School views

On average, 37 percent of schools rated their school leadership in relation to Promoting and Participating in Teacher Learning and Development as highly or outstandingly effective. Just over half of the schools in the sample gave school leadership highly or outstandingly effective ratings on three items: *ensuring that staff meetings include serious discussions about how to improve teaching and learning; ensuring that there is open discussion of students' results and teachers help each other develop more effective teaching strategies;* and *ensuring that student achievement patterns are analysed and used to plan professional learning priorities.* High or outstanding ratings were given by only between a third and a quarter for all the other six items in the Promoting and Participating in Teacher Learning and Development scale. Lower ratings were associated with school leadership effectiveness in ensuring the use of student outcome evidence to evaluate teaching approaches/professional development approaches. Lower ratings were also associated with how school leadership organises time; approximately one-quarter of schools rated as highly or outstandingly effective *ensuring the effectiveness of the provision of systematic opportunities for teachers to improve their teaching through observing the teaching of effective colleagues.* See Figure 10 below.

Figure 10 **Promoting and Participating in Teacher Learning and Development scale: School ratings**

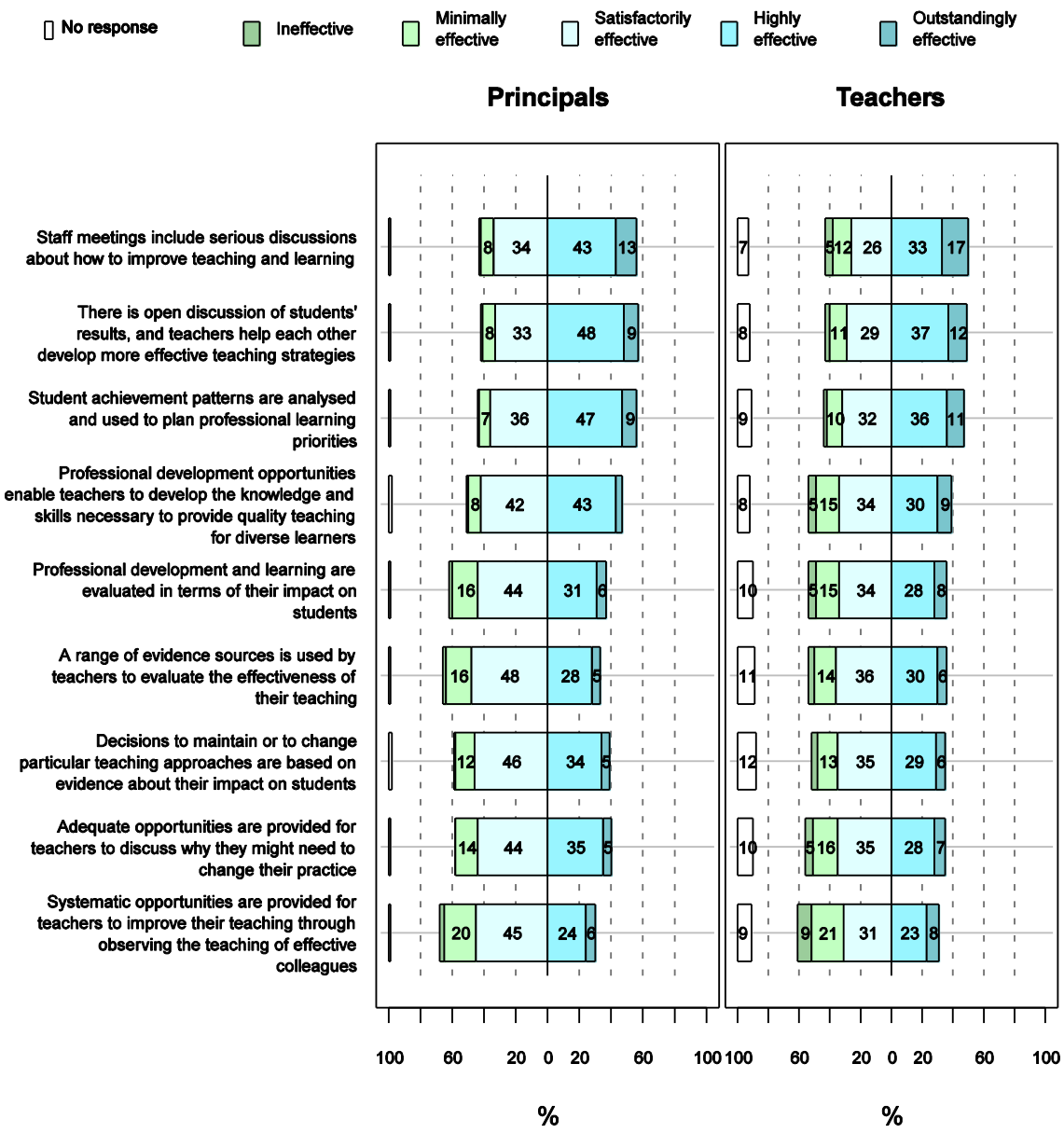


Principal and teacher views

Principals gave somewhat higher ratings than teachers on the items in the Promoting and Participating in Teacher Learning and Development scale. For example, 56 percent of principals and 50 percent of teachers rated their school leadership as highly or outstandingly effective in

ensuring that staff meetings included serious discussions about how to improve teaching and learning. See Figure 11 below.

Figure 11 **Promoting and Participating in Teacher Learning and Development scale: Principal and teacher ratings**



Ensuring a Safe and Orderly Environment scale

Evidence base

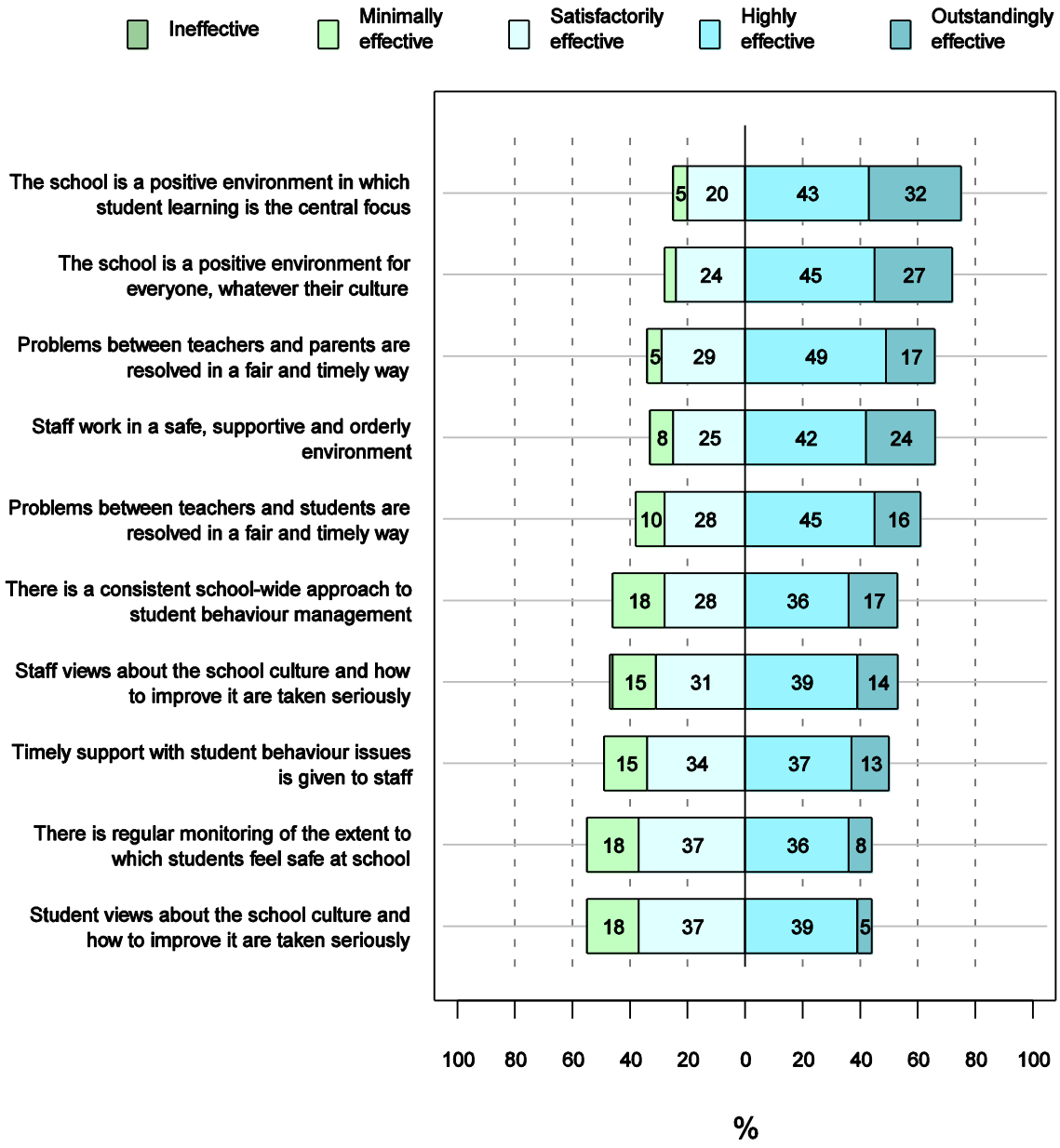
The evidence points to the need for the creation of an environment where key academic and social goals can be achieved. A safe and organised school lets students and teachers teach and learn without concern or distraction:

> The findings suggest that the leadership of effective schools is distinguished by an emphasis on, and success in establishing, a safe and supportive environment through clear, consistently enforced social expectations and discipline codes. (Robinson et al., 2009, p. 101)

School views

Overall, schools give high ratings to the items on the Ensuring a Safe and Orderly Environment scale, with an average of 58 percent of schools giving high or outstandingly effective ratings. Seventy-five percent of schools rated ensuring effectiveness of *ensuring that the school is a positive environment in which student learning is the central focus* as highly or outstandingly effective, and 72 percent *ensuring that the school is a positive environment for everyone, whatever their culture.* Ensuring effectiveness in checking that school policies were working over a period of time was less highly rated. Ensuring *that there is regular monitoring of the extent to which students feel safe at school* was rated highly or outstandingly effective by 42 percent of schools and *ensuring that student views about the school culture and how to improve it are taken seriously* by 44 percent of schools. See Figure 12 below for the detailed picture on schools' ratings of the effectiveness of their leadership in ensuring a safe and orderly environment.

Figure 12 **Ensuring a Safe and Orderly Environment scale: School ratings**



Teacher and principal views

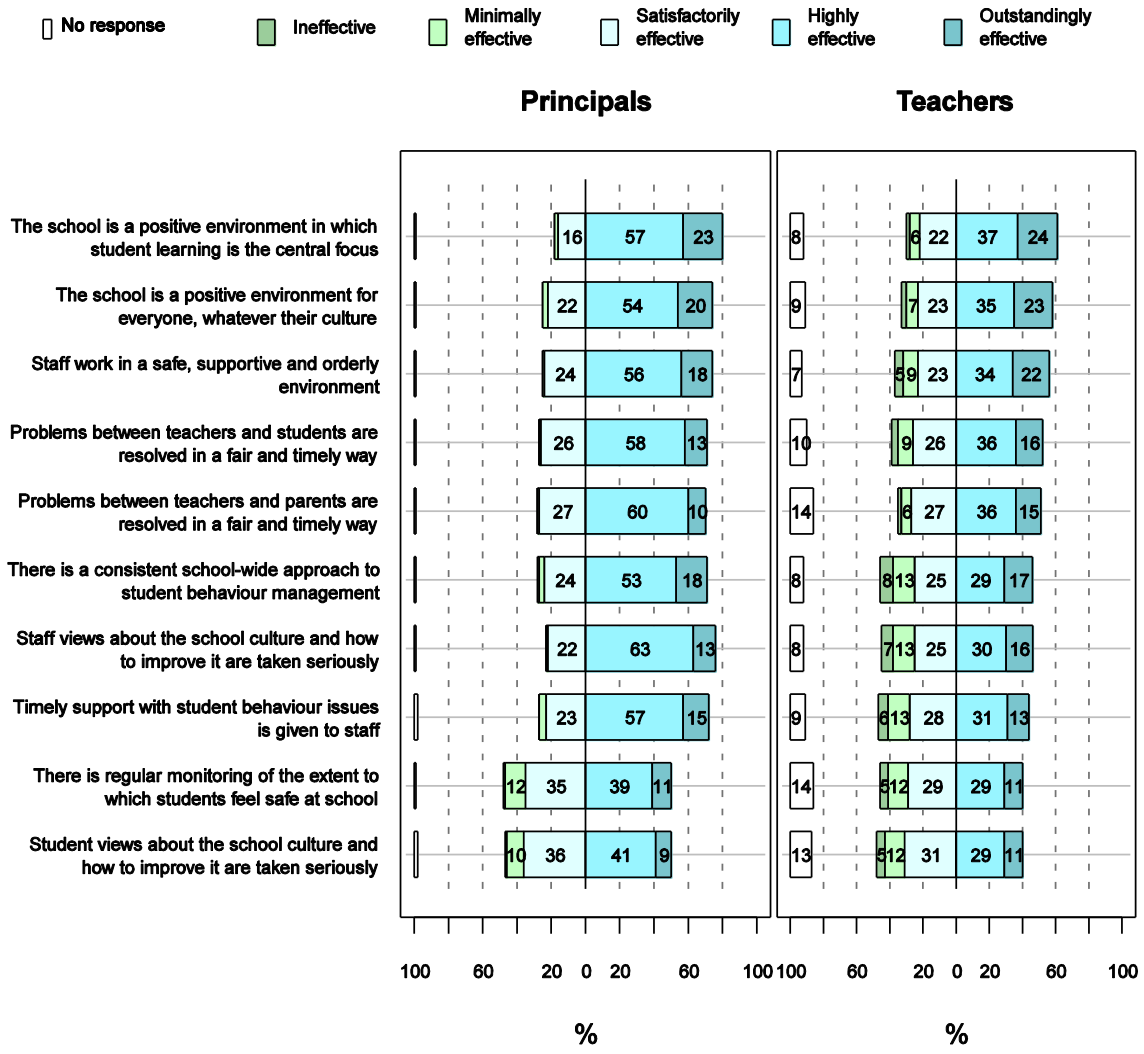
The Safe and Orderly Environment scale was one where teachers gave lower ratings than principals. This pattern is evident in the items below where less than 50 percent of teachers (but a higher percentage of principals) rated their school leadership as highly or outstandingly effective in ensuring that:

- *staff views about the school culture and how to improve it are taken seriously* (45 percent of teachers and 76 percent of principals)
- *there is a consistent school-wide approach to student behaviour management* (46 percent of teachers and 71 percent of principals)

- *timely support with student behaviour issues is given to staff* (44 percent of teachers and 72 percent of principals)
- *there is regular monitoring of the extent to which students feel safe at school* (40 percent of teachers and 50 percent of principals)
- *student views about the school culture and how to improve it are taken seriously* (40 percent of teachers and 50 percent of principals).

See Figure 13 for the detailed picture of teacher and principal views.

Figure 13 **Safe and Orderly Environment scale: Principal and teacher ratings**



Ensuring Educationally Powerful Connections with Families, Whānau and Community scale

Evidence base

> … leaders create educationally powerful connections when they:
>
> - Establish continuities between student identities and school practices;
> - Develop continuities and coherence across teaching programmes;
> - Ensure effective transitions across educational settings. (Robinson et al., 2009, p. 116)

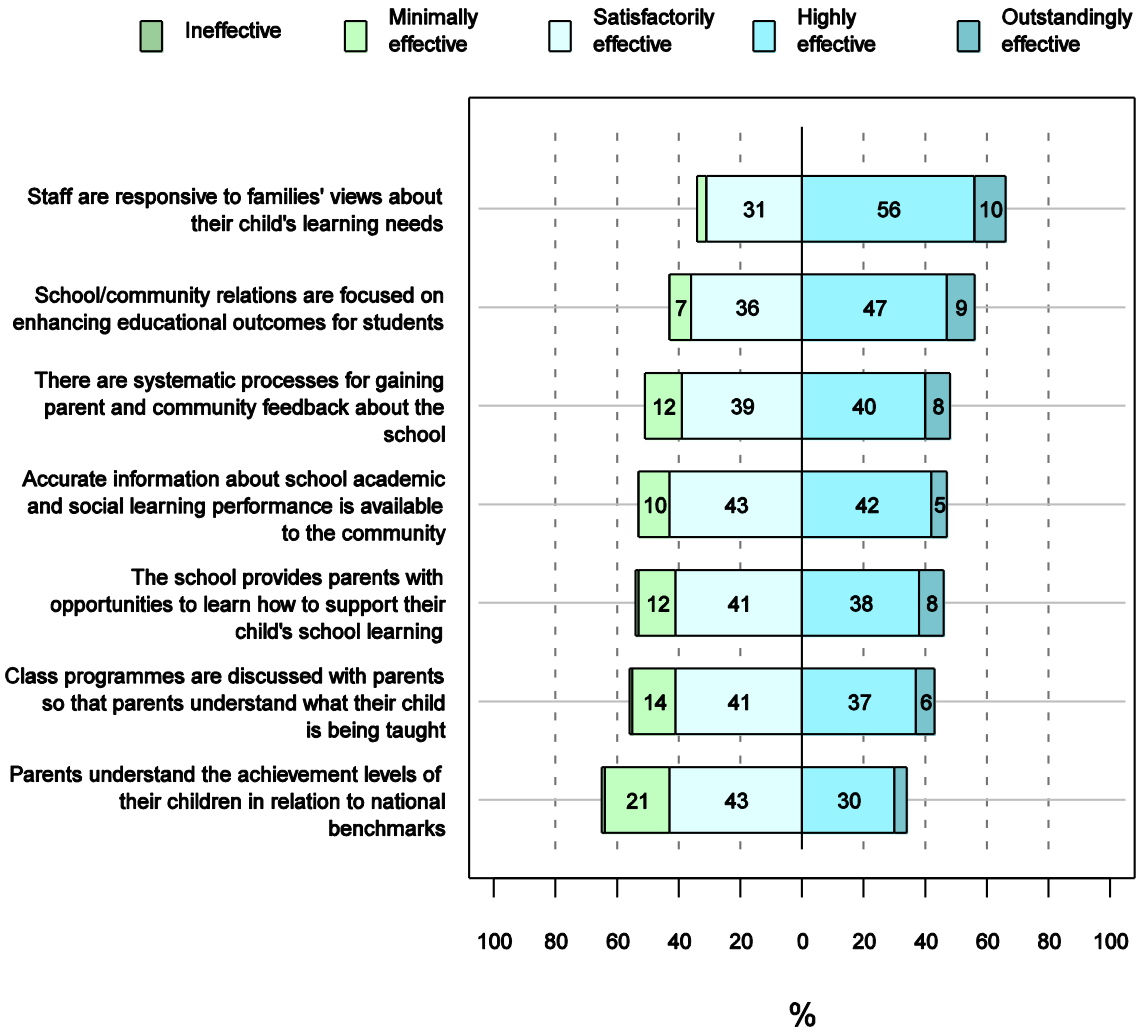
While *The New Zealand Curriculum* (Ministry of Education, 2007) is germane across all the scales of the ELP, it is worth noting in particular the *Community engagement* principle:

> The curriculum has meaning for student, connects with their wider lives, and engages the support of their families, whānau, and communities. (p. 9)

School views

Overall, 45 percent of schools rated school leadership in this area as highly or outstandingly effective. Almost all schools rated their educational leadership in relation to *ensuring that staff are responsive to families' views about their child's learning needs* highly or outstandingly effective. However, only two of the seven items on this scale were rated as highly or outstandingly effective by 50 percent or more of the schools in the sample. More than one in 10 schools gave school leadership a minimally effective rating for five items, and one in five did so for the item *ensuring that parents understand the achievement levels of their children in relation to national benchmarks.* (This sample includes data collected prior to the introduction of National Standards.) See Figure 14 below.

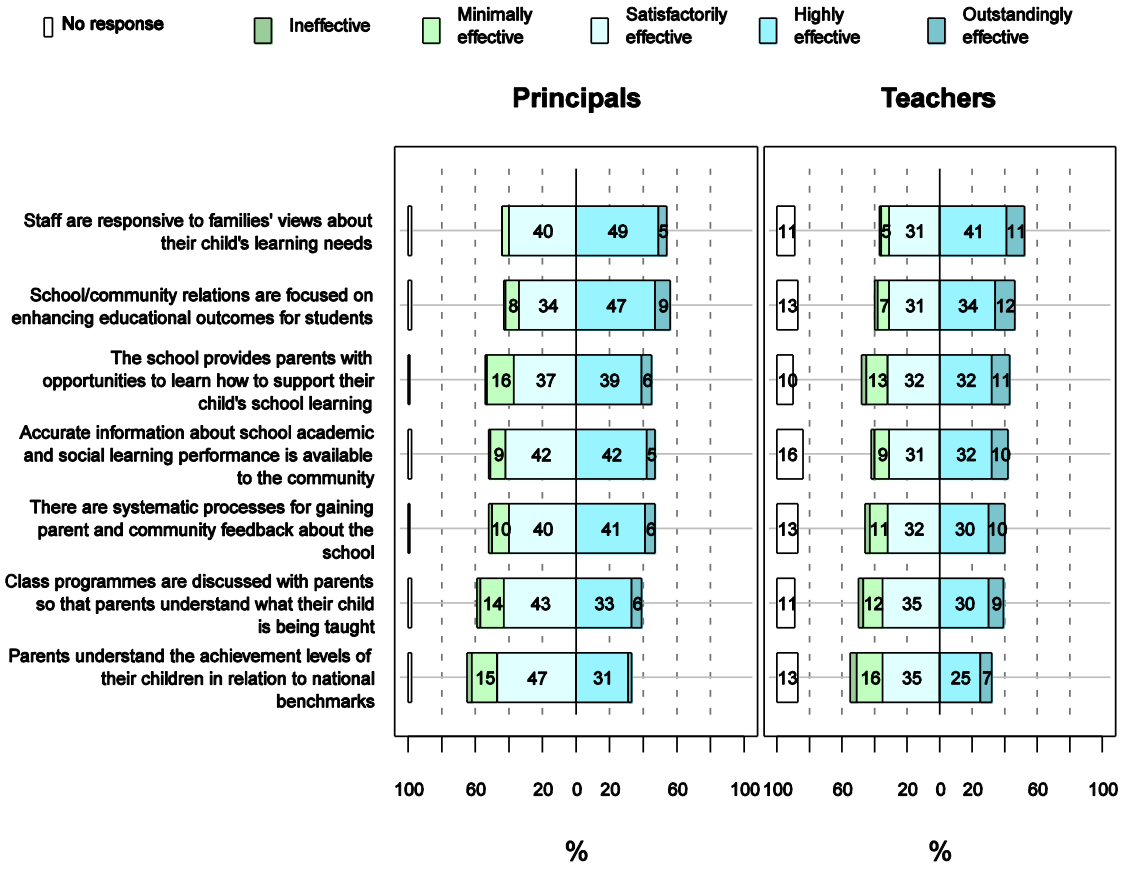
Figure 14 **Ensuring Educationally Powerful Connections with Families, Whānau and Community scale: School ratings**



Teacher and principal views

Similar percentages of teachers and principals gave minimally effective ratings. Teachers were slightly more likely than principals to give high ratings for the items on this scale. See Figure 15 below.

Figure 15 **Ensuring Educationally Powerful Connections with Families, Whānau and Community scale: Principal and teacher ratings**



Māori success scale

As noted in Section 1, this set of items is critical in the New Zealand context:

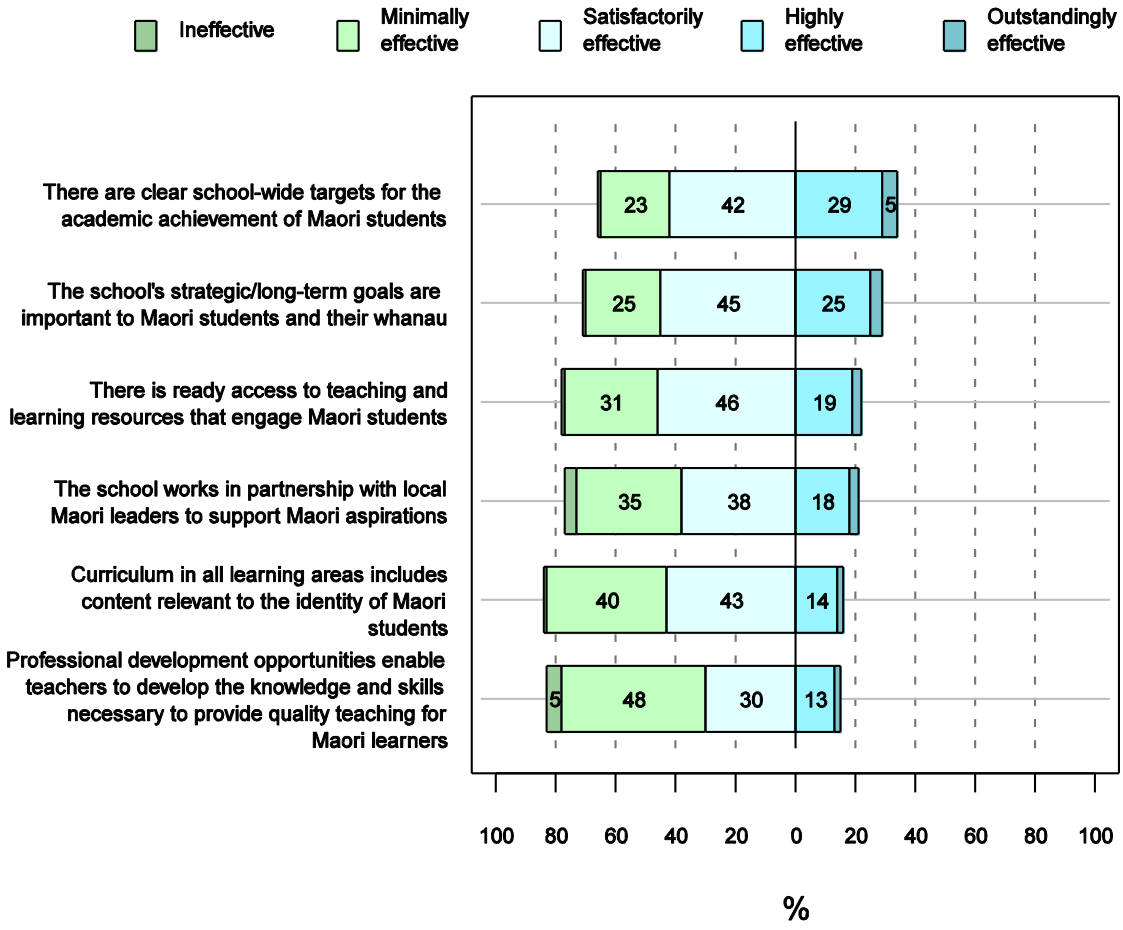
> The Treaty of Waitangi is central to, and symbolic of our national heritage, identity, and future. Our commitment to the principles of the Treaty obliges a distinctive focus on ensuring excellent education outcomes for Māori. Educational success is the key to enabling Māori to live as Māori in tea Ao Māori and in the wider world. Our task is to expand on emerging successes for Māori. This is fundamental to an equitable education system. (Ministry of Education, 2007, p. 10)

The items related to Māori students are distributed throughout the questionnaire and were intended to relate to a number of different aspects of school leadership. However, as also noted in Section 1, psychometric analysis indicated that they showed a distinct additional aspect of educational leadership. The score distribution on this scale was different from the aspects of the main scale, with more describing their school's education leadership as ineffective or minimally effective in ensuring that the practices asked about occurred. Just 17 percent of schools rated this area as highly or outstandingly effective.

School views

Schools were most likely to rate their school leadership as highly or outstandingly effective in relation to having clear school-wide targets for the academic achievement of Māori students (around 33 percent), and least likely to rate them so for ensuring that there were *professional development opportunities that enabled teachers to develop the knowledge and skills necessary to provide quality teaching for Māori learners* (around 15 percent). See Figure 16 for the detailed view of school ratings on ensuring effectiveness in Māori success.

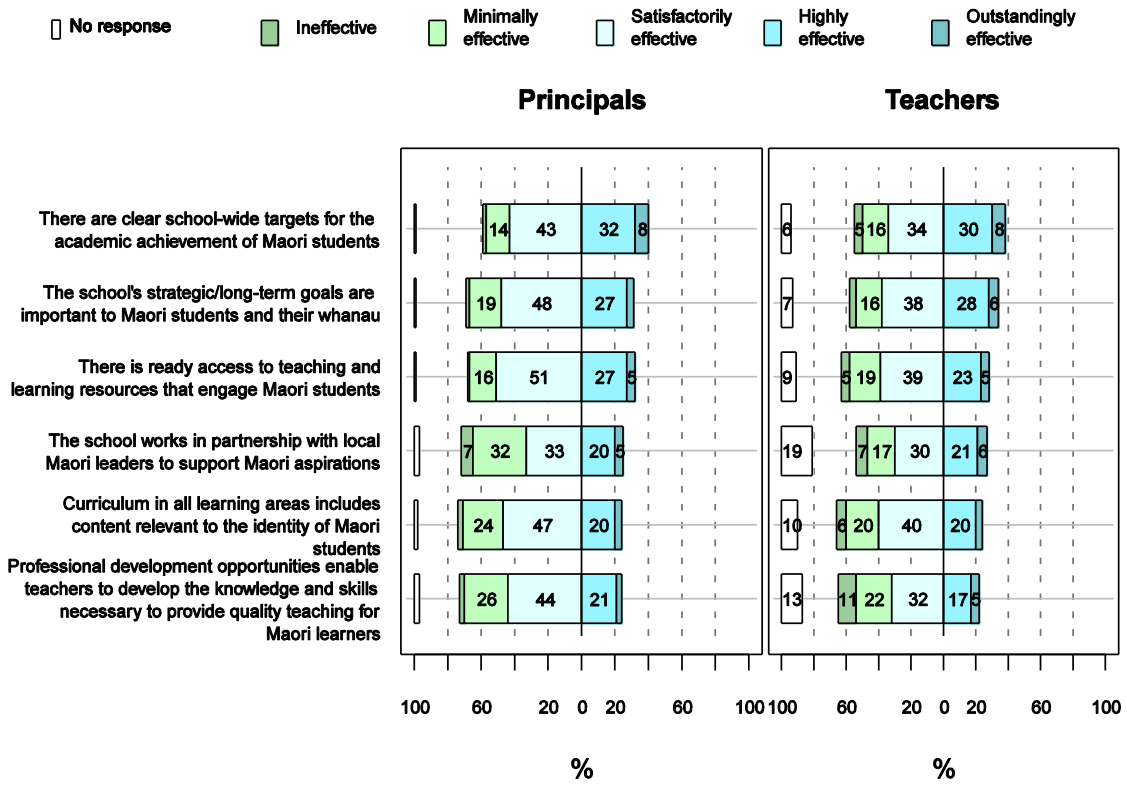
Figure 16 **Māori Success scale: School ratings**



Principal and teacher ratings

School means were lower than principal means in this scale. See Figure 17 below for the detailed responses for both teachers and principals on the Māori subscale.

Figure 17 **Māori Success scale: Principal and teacher ratings**



Teacher workload and morale

Evidence base

Analysis of the NZCER 2009 National Secondary Survey indicates that high morale is associated with factors that have been shown to be positively linked to good student outcomes:

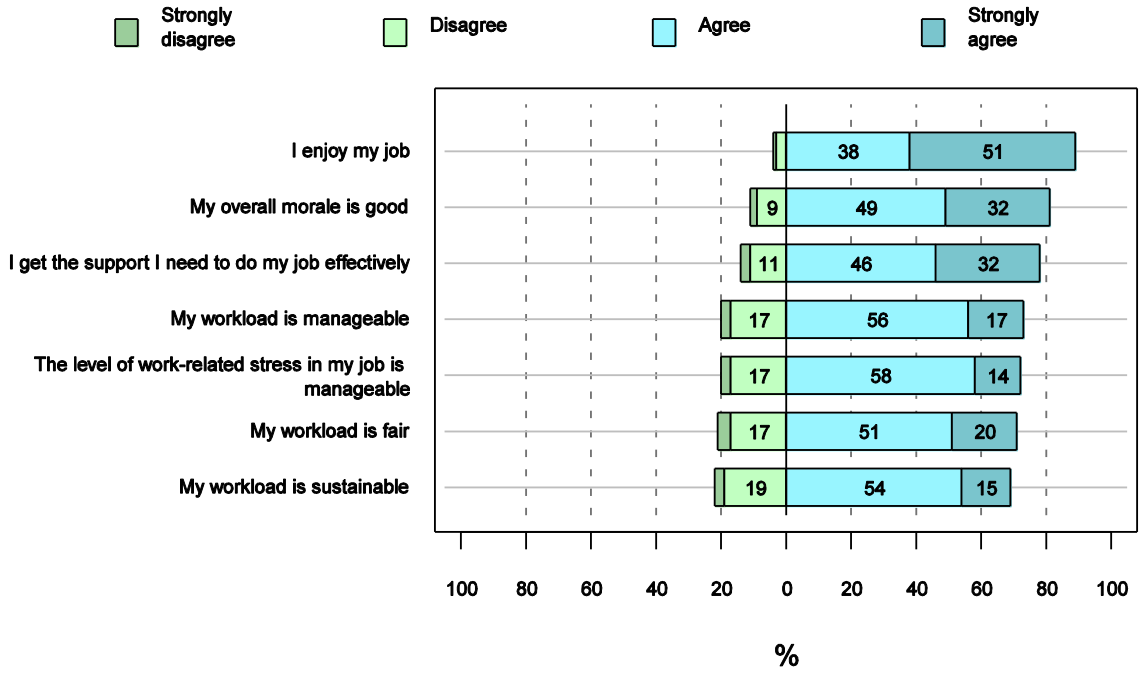
> High morale levels were linked to high levels of experience of collegiality and collective professional work. (Wylie, 2010, p. 60)

In the ELP national sample, just over 80 percent of teachers agreed or strongly agreed that their overall morale was good, and 89 percent agreed or strongly agreed that they enjoyed their job. When scores were aggregated by schools (excluding principals as per the process described in Section 1), 84 percent of schools agreed or strongly agreed that their overall morale was good, and 99 percent agreed that they enjoyed their job. Teachers in the ELP sample were somewhat less sanguine about workload: 72 percent of teachers (82 percent of schools—average teacher score per school) agreed or strongly agreed that their workload was fair and 73 percent of teachers (and 82 percent of schools) that it was sustainable.

Teacher workload and experience is significantly correlated with ratings on the ELP scales and with their perceptions of principal effectiveness. In other words, teachers whose morale is lower than others and who are less satisfied that their workload is sustainable or fair give lower ratings to their school leadership for all the ELP items, and lower ratings for the effectiveness of their

principal. Teacher workload and morale is rated lowest in composite and secondary schools and is most highly rated in the smallest schools (U1 and U2). There is a slight significant difference in teacher views in relation to decile—teachers in higher decile schools give higher ratings to their job enjoyment and getting support to do their job effectively.

Figure 18 **Teacher workload and morale**



Principal Leadership scale

Evidence base

The Kiwi Leadership for Principals framework (Ministry of Education, 2008, pp. 18–20) outlines areas of practice that principals work in to lead change and to solve problems in their schools:

- culture (what we value around here)
- pedagogy (knowledge about teaching and learning)
- systems (how things work around here)
- partnership and networks (creating positive links to support learning).

The Kiwi Leadership for Principals framework also talks about the critical nature of relationships (p. 13) and the qualities of effective school leaders (pp. 22–23) (outlined below):

- manaakitanga (leading with moral purpose)
- pono (having self-belief)
- ako (being a learner)
- awhinatanga (guiding and supporting).

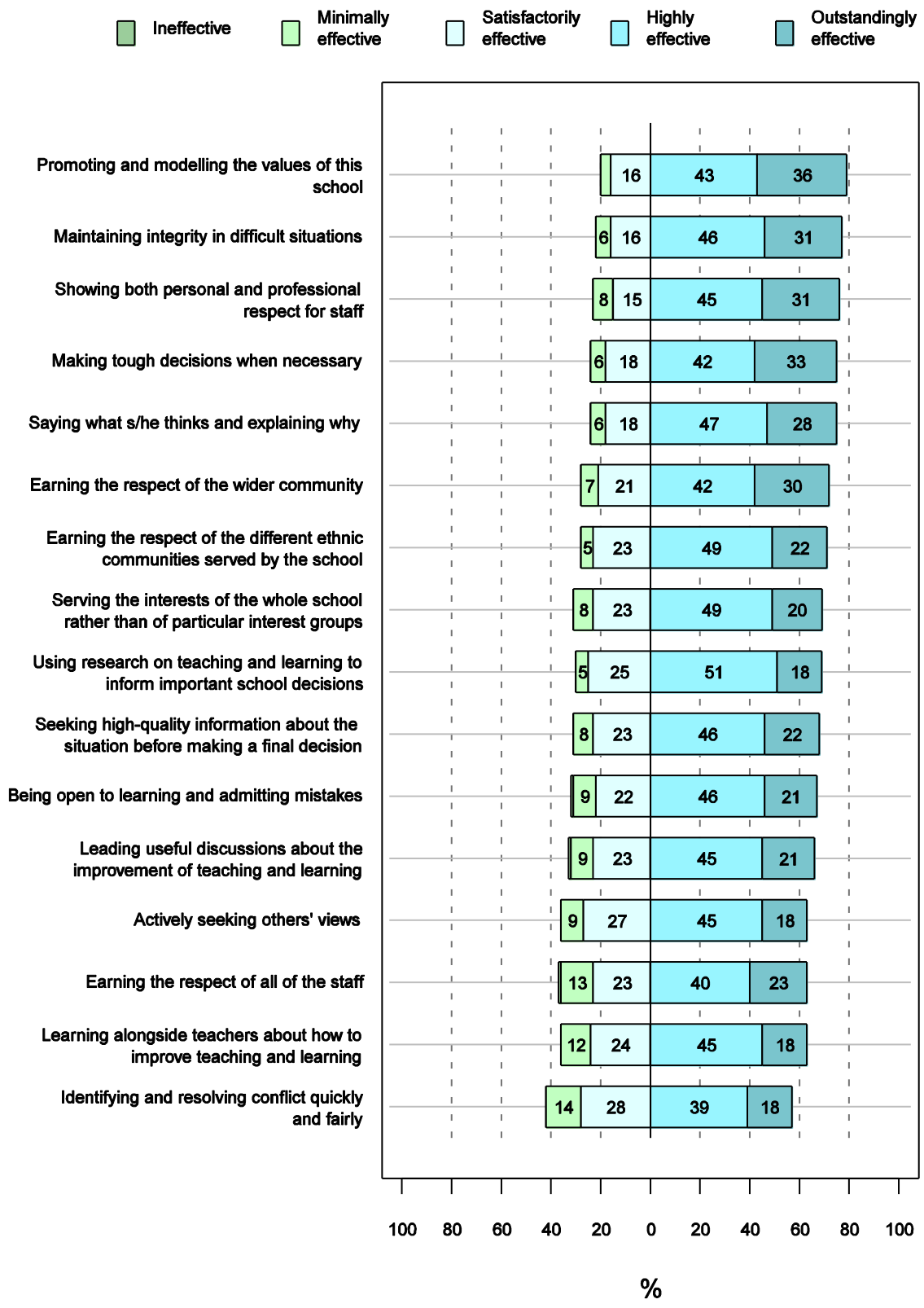
Robinson et al. (2009) also outline the knowledge, skills and dispositions involved in effective educational leadership, and the need to emphasise the elements outlined above. In addition, they note the importance of a principal's ability to identify and resolve conflict (p. 102).

The Principal Leadership scale captures these elements of effective principal leadership.

School views

Overall, 69 percent of schools rated Principal Leadership highly or outstandingly effective. The most highly rated aspects of principal effectiveness were around general patterns of behaviour—modelling the values of school, maintaining integrity in difficult situations and showing both professional and personal respect for staff. Less frequently rated as highly or outstandingly effective were: *identifying and resolving conflict quickly and fairly; learning alongside teachers about how to improve teaching and learning; earning the respect of all the staff;* and *actively seeking others' views.* See Figure 19 for the details.

Figure 19 **Principal Leadership scale: School ratings**



Principal and teacher views

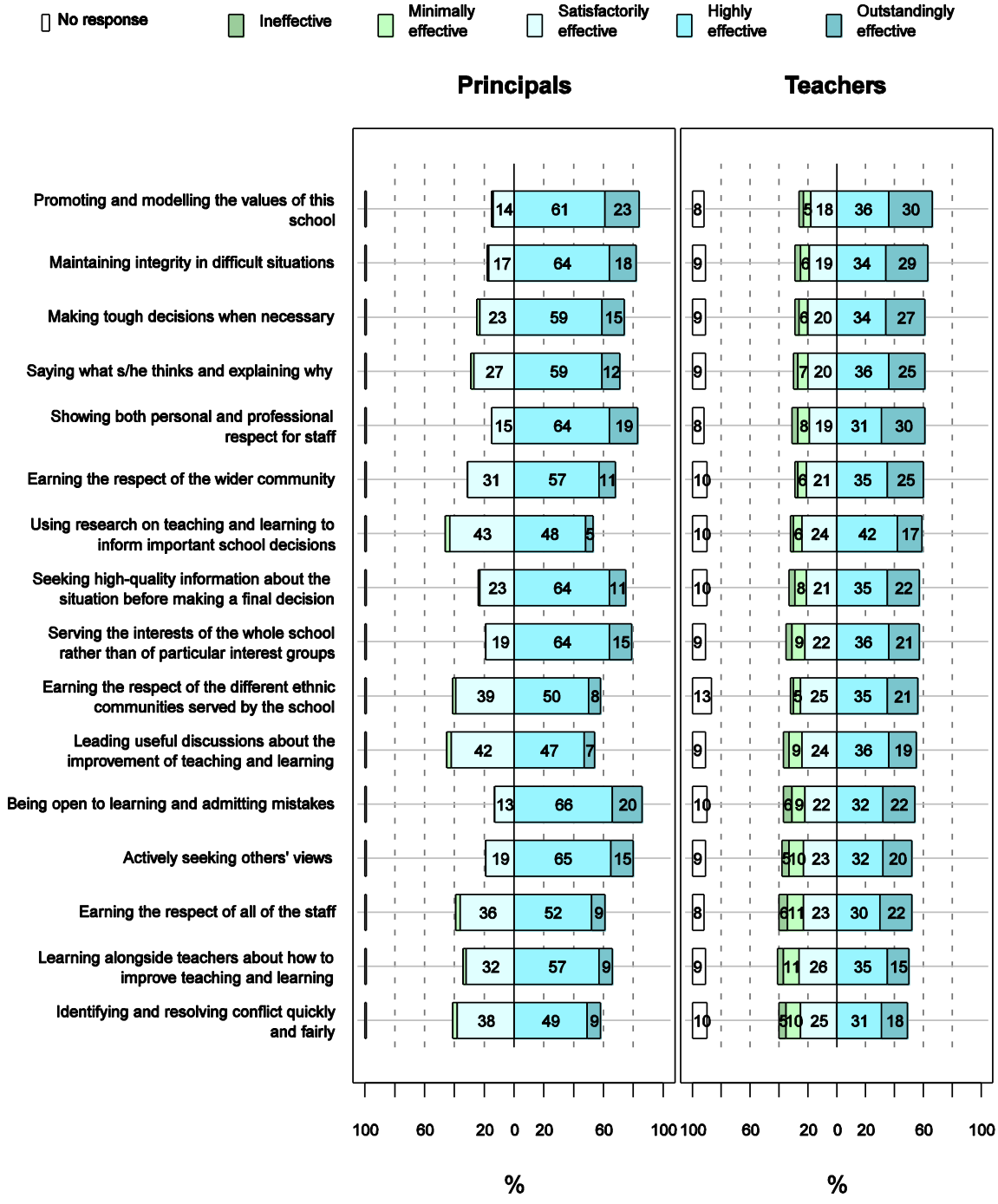
The patterns of response between teachers and principals show both similarities and differences. The rank order of effectiveness rating is similar across the two groups. However, teachers tended to give ratings more frequently at the extreme ends of the scale than did principals. By contrast, very few principals rated themselves as minimally effective, and none as ineffective.

Principals tended to give high ratings more often than teachers in terms of their effectiveness in:

- *actively seeking others' views* (52 percent of teachers rated this as highly or outstandingly effective cf. 80 percent of principals)
- *earning the respect of all the staff* (52 percent of teachers cf. 61 percent of principals)
- *identifying and resolving conflict quickly and fairly* (49 percent of teachers cf. 58 percent of principals).

See Figure 20 for the detailed description of teachers' and principals' ratings of principal effectiveness.

Figure 20 **Principal Leadership scale: Principal and teacher ratings**



Context for Pedagogical Leadership scale

Evidence base

While devolution has allowed principals to develop systems in response to the needs of their students, it has also increased their administrative workload. Some research (e.g. Hodgen and Wylie 2005) suggests that New Zealand principals spend almost twice as much time on

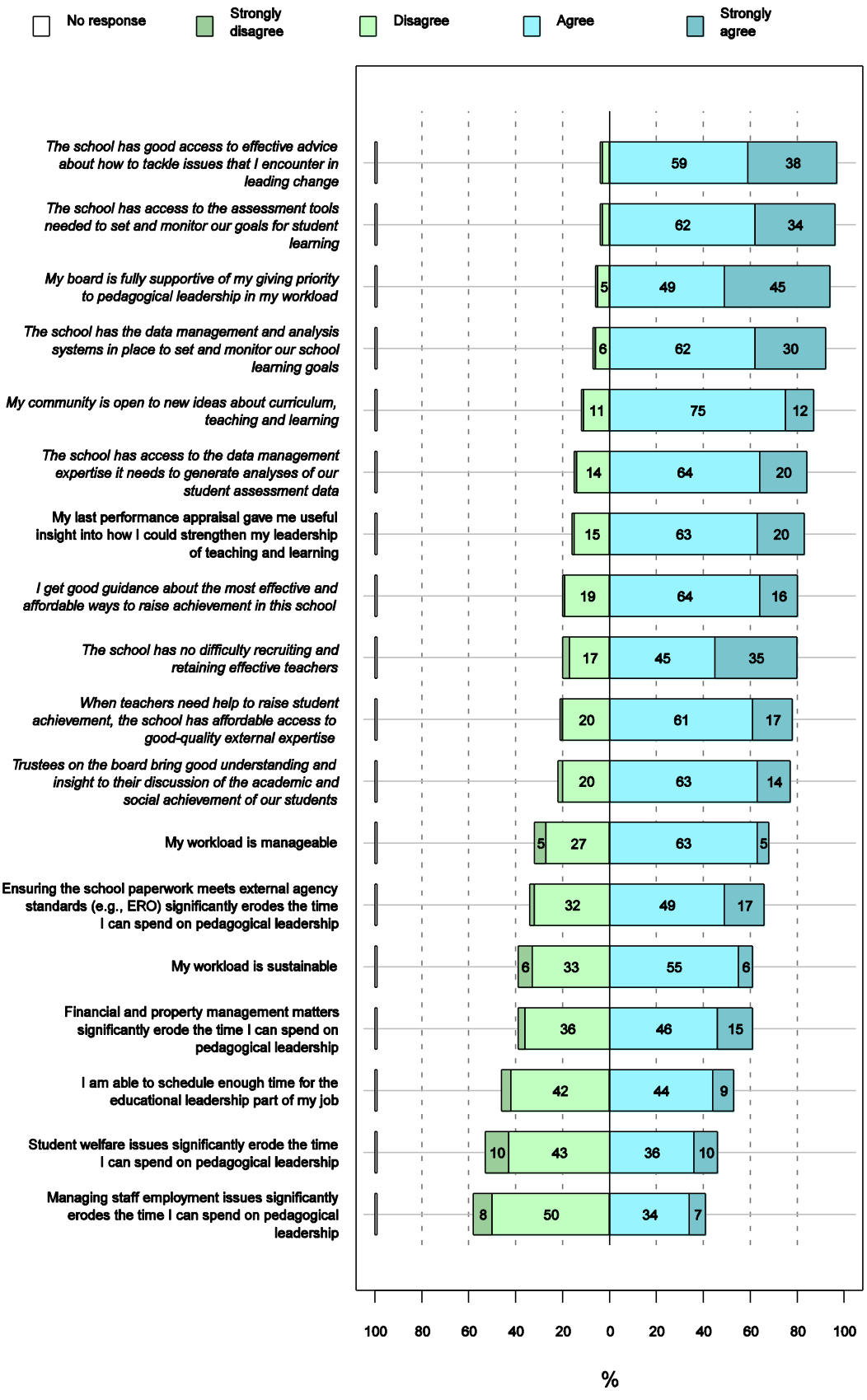
> administration as do their international counterparts. The multiple demands of leadership and administration can be a source of tension for the principal when deciding how to prioritise time and attention. (Ministry of Education, 2008, p. 10)

If we want to further develop educational leadership in New Zealand schools, it is also important to know about the supports and constraints that are experienced in the school setting. The context scale built on prior research in this area, particularly other aspects of the principal role in a self-managing system, and incorporating aspects of policy that are intended to support school development, such as principal appraisal and analysis of student achievement data.

The context for pedagogical leadership scale is answered only by principals. Principals were asked to rate their agreement with 17 items, using a 4-point scale. The items were designed to assess a range of both support and barriers to their pedagogical leadership. More than 90 percent of principals agreed or strongly agreed that the school had access to *effective advice* and *assessment tools*, and that the *board was supportive of the principal giving priority to pedagogical leadership* and that the school had *data management and analysis systems in place to monitor school learning goals.* On the challenges side of the ledger, manageable and sustainable workload, external agency paperwork demands, time, finance and property demands, student welfare and staff issues were challenges for a significant number of principals.

Figure 21 shows their responses in detail.

Figure 21 **Context for Pedagogical Leadership scale**



Summary of patterns across the scales

Goal Setting: There appears to be more confidence about the role of leadership in relation to schools' guiding frameworks than about the embedding of the goals into ongoing use and evaluation—linking the goals back to individual teacher roles and individual student learning.

Strategic Resourcing: The schools gave highest ratings to the effectiveness of their school leadership in ensuring that the timetable reflected the school's priorities for teaching and learning, and lowest to items related to working with families and communities. In between come items related to teaching resource relevance and availability.

Curriculum Quality: School leadership was seen as most effective in ensuring the systematic monitoring of each student's progress and the existence of assessment plans to collect the information needed to monitor progress on priority learning goals, and least effective in ensuring that rigorous feedback was given to teachers about the quality of their schemes or unit plans, that all students experience challenging programmes and that all curriculum included content relevant to diverse learners.

Quality of Teaching: More than half the schools in this national sample thought that their school leadership was highly or outstandingly effective in ensuring that everyone shared responsibility for student learning, that assessment data were used to improve teaching and that those teachers with particular expertise were used in the school to help other teachers' development. The lowest rating item was students providing feedback to teachers on the effectiveness of their teaching, followed by support provided for teachers having difficulty helping students reach important academic and social goals, and challenge and support to improve teaching for teachers whose students remain disengaged.

Promoting and participating in teacher learning and development discussion of student results and teachers helping each other develop more effective teaching strategies, serious discussions of how to improve teaching and learning in staff meetings and analysis and use of student achievement patterns to plan professional learning priorities were the items most likely to attract highly or outstandingly effective ratings of school leadership. Schools were much less likely to give such ratings to the provision of systematic opportunities to improve teaching through observing effective colleagues at work, teachers' use of a range of evidence sources to evaluate the effectiveness of their teaching and adequate opportunities are provided for teachers to discuss why they might need to change their practice.

Ensuring a Safe and Orderly Environment: Most schools gave their school leadership highly or outstandingly effective ratings for ensuring they had positive environments for learning in which student learning is the central focus, and that the school was a positive environment for everyone, irrespective of culture. The gathering and use of student views in relation to school safety and culture were the two items on which the school leadership was least likely to be rated as highly or outstandingly effective.

Ensuring Educationally Powerful Connections with Family, Whānau and Community: Two-thirds of the schools thought their leadership was highly or outstandingly effective in ensuring that staff were responsive to families' views about their child's learning needs. Only around one-third of schools thought that their school leadership was highly or outstandingly effective in ensuring that parents understand the achievement levels of their children in relation to national benchmarks.

Māori Success: Schools were markedly less sanguine in this area. Schools were most likely to rate their school leadership as highly or outstandingly effective in relation to having clear school-wide targets for the academic achievement of Māori students (just over one-third of schools in the national sample), and least likely to rate them so for ensuring that there were professional development opportunities that enabled teachers to develop the knowledge and skills needed to provide quality teaching to Māori learners (only around 15 percent of schools).

Principal Leadership: The top items in this scale were mostly related to the principal promoting the values of the school, having integrity and gaining and showing others respect; it also included making tough decisions when necessary. Identifying and resolving conflict quickly and fairly was the item with the lowest proportion of schools rating their principal as showing highly or outstandingly effective leadership.

3. School characteristics and ELP scores

Introduction

How different are the ELP scores in schools with different characteristics? The first report on the baseline EPD sample showed that school scores were affected by how complex the school setting was, with higher scores more likely to occur in primary, small, rural and/or high-decile schools. Here we provide more detail of the differences for the national norms sample associated with the school characteristics of type, size, school socioeconomic decile and proportion of Māori enrolment, along with a picture of what it looks like when school characteristics are combined (e.g., decile and school type). This will enable schools to compare their own results with "schools like us".

In a series of figures in this section we describe the variability overall in scores according to school type, size, decile and percentage of Māori involvement. The first set of figures in this section shows the means (dots for principals and triangles for schools) for the overall ELP scale, and two standard errors[3] above and below the mean (the lines above and below the dot/triangle). On some of the figures, large errors of measurement are evident. These can be accounted for by high variability within that part of the sample and/or small samples.

The focus of this section and indeed this report is about schools and "schools like us". We have accordingly used the averaged ratings for each school as the base unit for aggregation (as noted previously, these are school averages and do not include principals' ratings). For additional comparison/commentary, we have added the separate principals' means and standard errors to the figures in this section.

We start with the overall scale, and then look at each of the ELP scales in turn.

Overall scale—differences related to school characteristics

School type

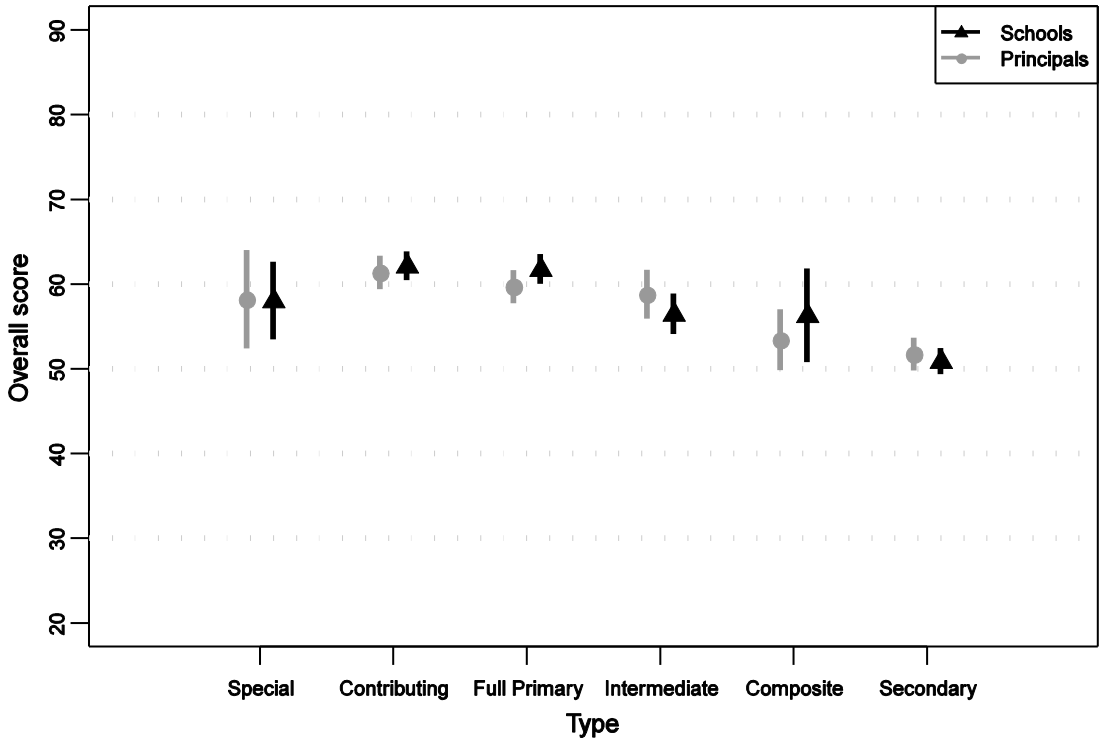
We compare school type: composite; contributing primary; full primary; intermediate; secondary; and special schools. Contributing and full primary schools have higher average mean scores than do composite and secondary schools. High standard errors of measurement are noted in composite

[3] The standard error (SE) is a measure of the variability of the mean. Our sample gives us an estimate of the true mean for all schools in the country. The SE gives us an idea about how accurate our estimate is likely to be. There is a 68 percent chance the true mean lies within 1 SE of the mean, and a 95 percent chance it lies within 2 SE of the mean.

and special schools, which are smaller in number. Generally, means are similar for principals and schools, with a slightly higher mean rating for principals (over school means) in intermediate schools (see Figure 22).

How would this be interpreted by a secondary school with an overall scale score of 50, or one of 55? The sample mean for secondary schools was just over 50, so a school with an overall scale score of 50 has leadership that is about as effective as that in most other similar schools. A secondary school with a mean of 55, however, has a mean outside the range of the error bars, so leadership in that school has been rated as being more effective than that in most other secondary schools.

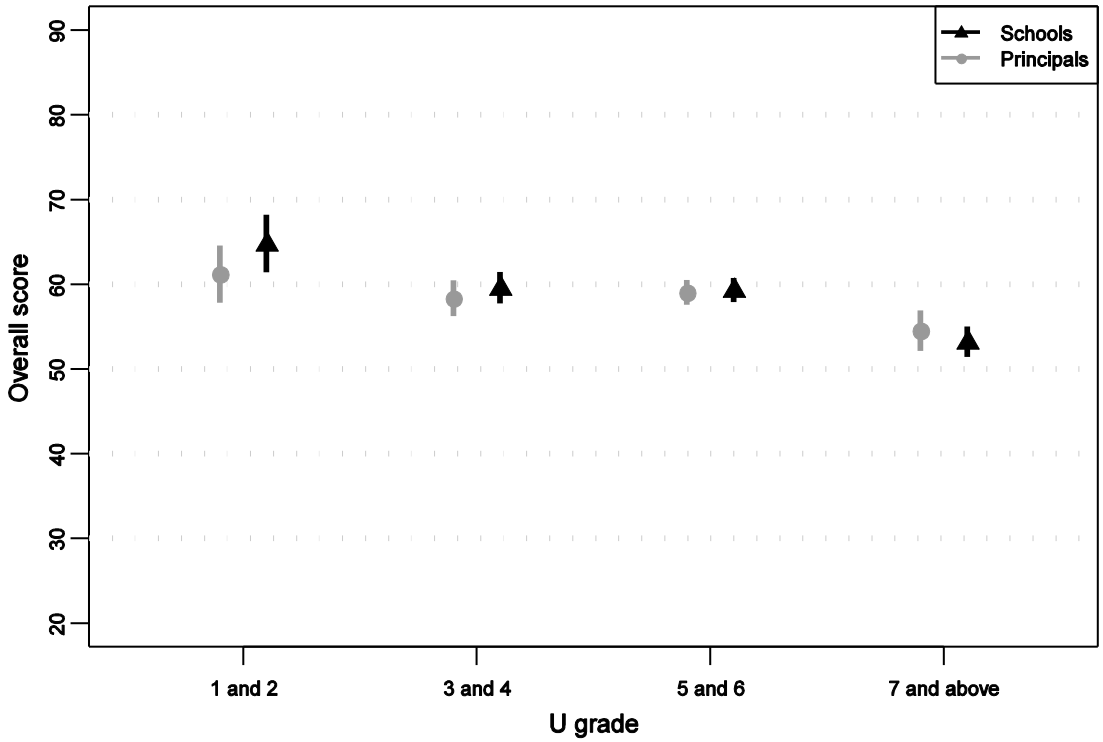
Figure 22 **School and principal overall scaled scores according to school type**



School size

School size has been analysed according to U grade, the grading used in principal salary scales. U grades run from U1 (1–50 students) to U16 (2,400+ range). A U4 school has a roll of 151–300 students; a U5 school a roll of 301–500; and a U7 school has a roll of 676–850. Because of the relatively small number of schools in our sample in the higher U grades we have grouped all the larger schools into the category "U7 and above" in the figures in this section. There were similar patterns for schools larger than U7, with the **size** of the difference in some areas increasing more above U7. Larger school size (U7 and above) is associated with lower ratings of effectiveness on the ELP (see Figure 23 below).

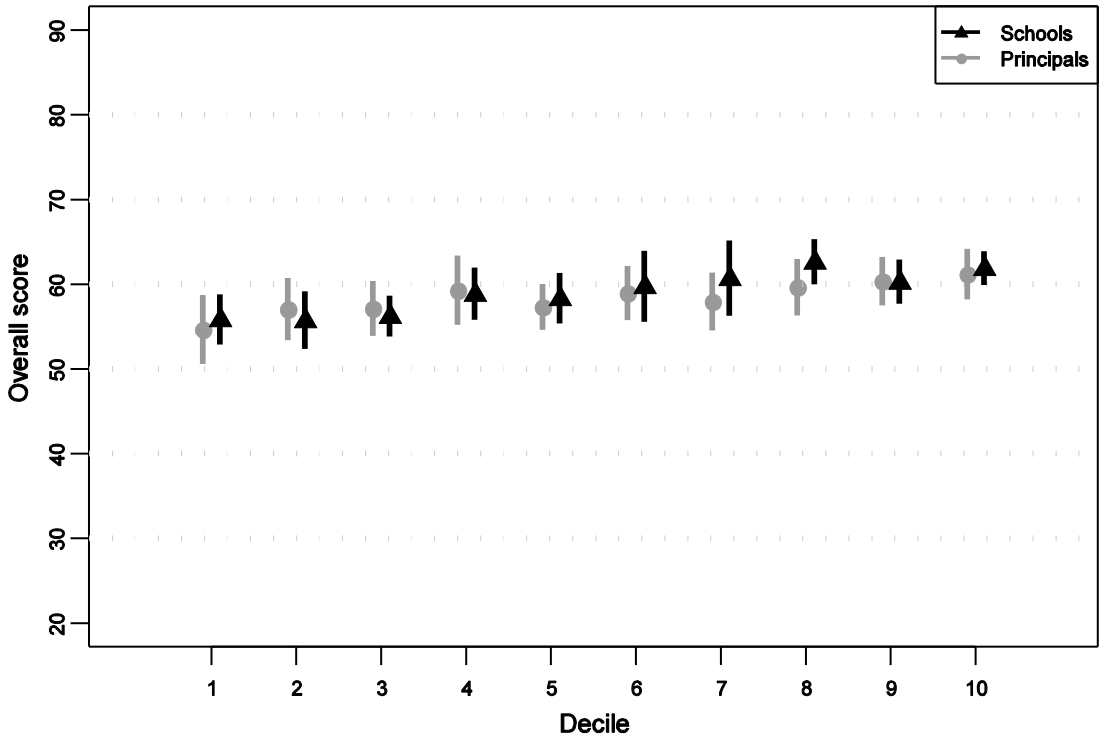
Figure 23 **School and principal overall scaled scores according to school size**



School decile

School socioeconomic decile is clearly associated with differences in scores on the ELP. The trend evident in Figure 24 shows lowest scores on the overall scale were in deciles 1–3 schools, slightly higher scores in deciles 4–5 schools, slightly higher scores again in decile 6+ schools. The highest average overall score was in decile 8 schools, although on closer examination this is likely to be a feature of the sample composition—with more primary schools in the decile 8 group of schools.

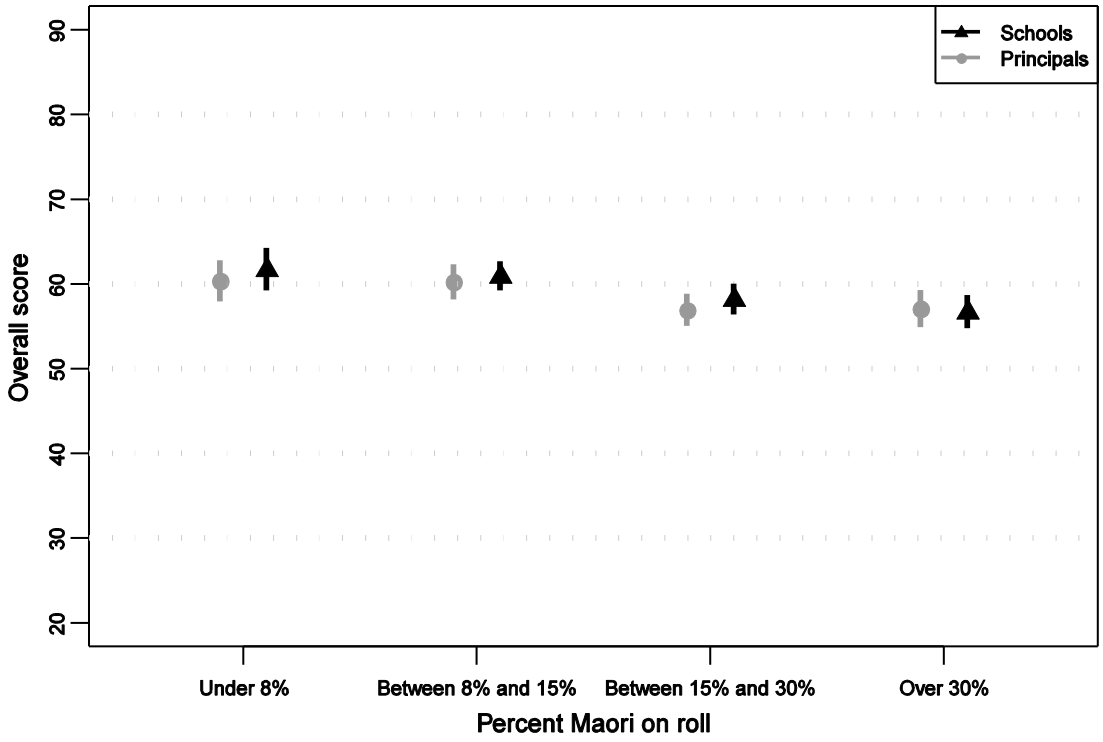
Figure 24 **School and principal overall scaled scores by school decile**



Percentage of Māori on the school roll

A slight decrease in overall school scaled scores is visible when the proportion of Māori students on the school roll rises above 15 percent and again when the proportion of Māori students on the roll is above 30 percent (see Figure 25 below).

Figure 25 **School and principal overall scaled scores by percentage of Māori on the school roll**

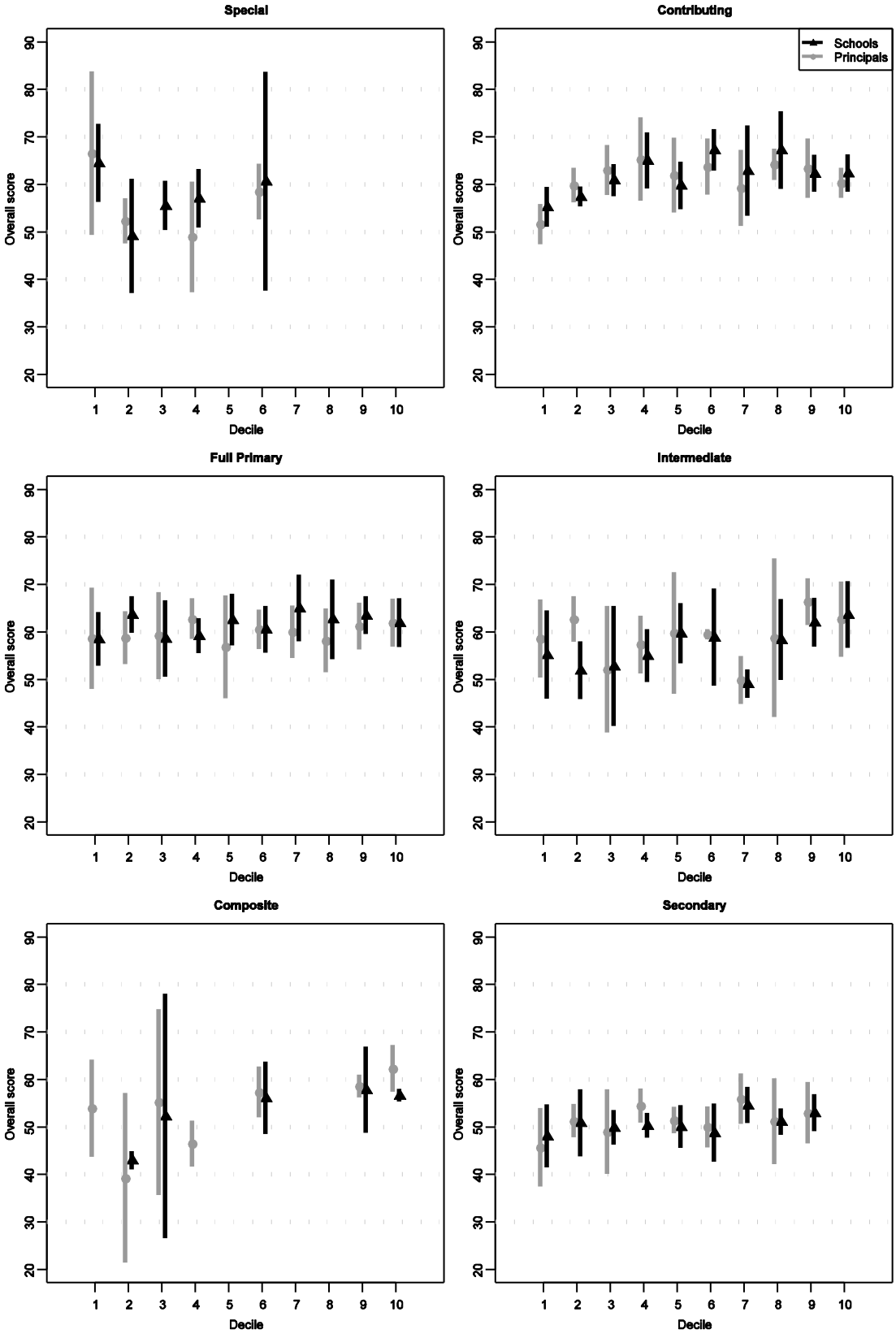


Differences within types of school by decile

It is possible to look at profiles for schools when factors are combined (e.g., school type and decile). Figure 26 gives a graphical representation of this. Variability is noted, but the size of the standard error of measurement suggests caution with this level of fine-grained analysis. We would need a much larger sample to investigate the interaction between sample size and variability within each subgroup before we could draw any in-depth conclusions about the differences between, say, a decile 3 secondary school or a decile 6 secondary school.

However, while there are larger standard errors evident on the patterns evident in Figure 26, similar trends can be observed between Figures 22 to 25 above, and Figure 26 below. Decile patterns appear to be still grouped into low, mid and high patterns of means. The lowest means are in low-decile secondary and composite schools, and the highest means are in high-decile primary schools. So, for example, a low-decile secondary school might expect to have a lower mean score than a high-decile secondary school, if school size was constant at, say, U7 across schools.

Figure 26 **Differences in overall scaled score by decile and school type**

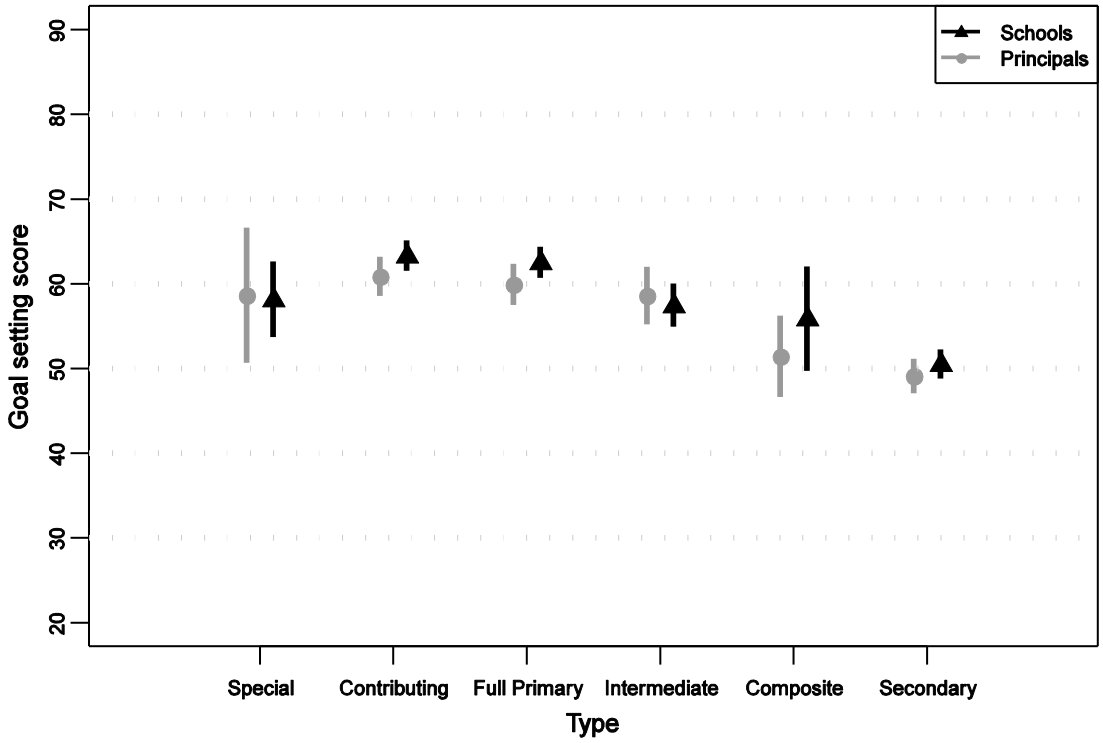


Goal Setting scale

School type

Secondary and composite schools gave the Goal Setting scale items lower ratings, at around 10 scaled score points lower than other school types. In most school types, the schools rated the goal setting items higher than did principals. There were significant differences related to school type on all but one item on the Goal Setting scale. Principals were also significantly less likely to give high ratings on the goal setting items if they were at a secondary or a composite school. Wider variability is noted in special schools and in composite schools—most likely due to a combination of small sample sizes in these school types and a possible wide variation across schools within those types (see Figure 27).

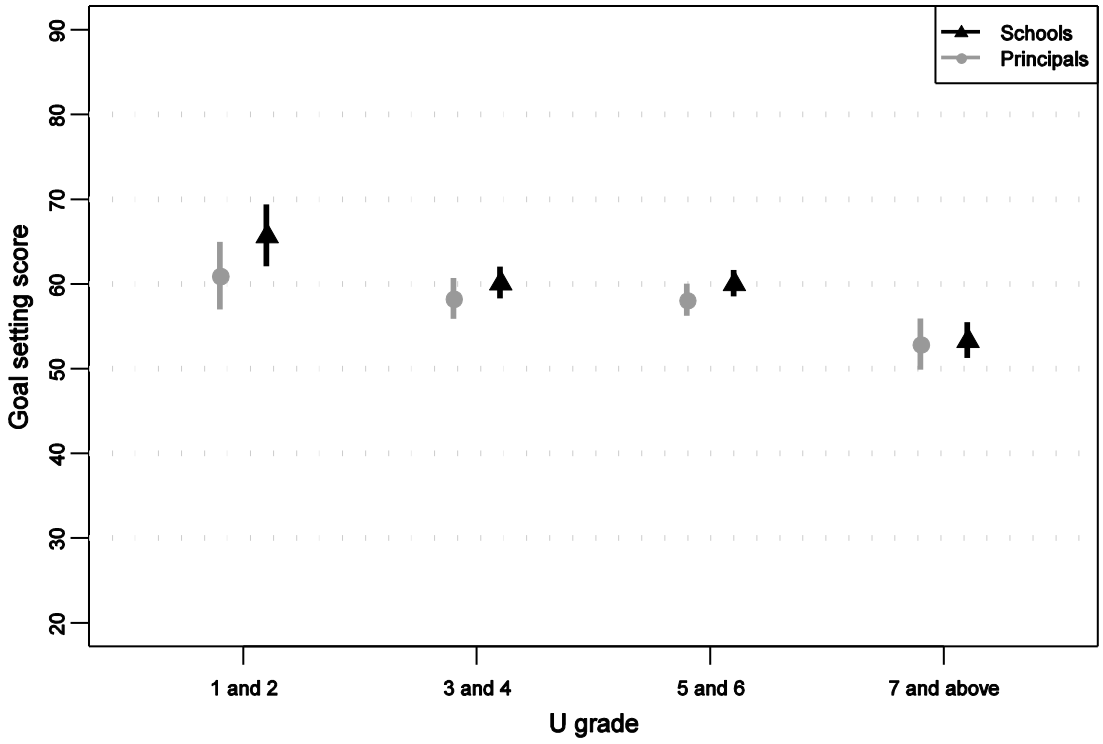
Figure 27 **Goal Setting: Schools and principals, according to school type**



School size

There are approximately 10 scaled score points between the smallest schools' and the larger schools' means on the Goal Setting scale.

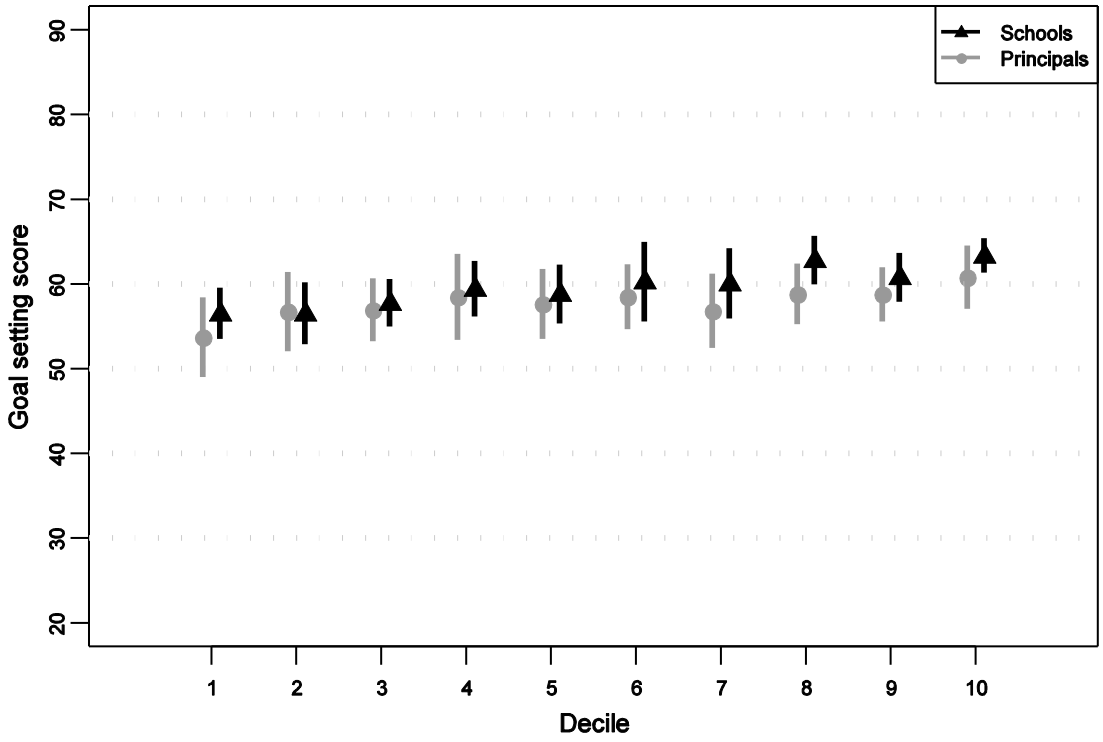
Figure 28 **Goal Setting: School and principal scaled scores for goal setting according to school size**



Decile differences

Trends across low-, middle and high-decile schools are noted: Similar school scaled scores on the Goal Setting scale across deciles 1–3; slightly higher school scaled scores for deciles 4–7; and higher again for deciles 8–10 schools. (A slightly higher mean for decile 8 schools is most likely explained by the higher proportion of primary schools in this group.)

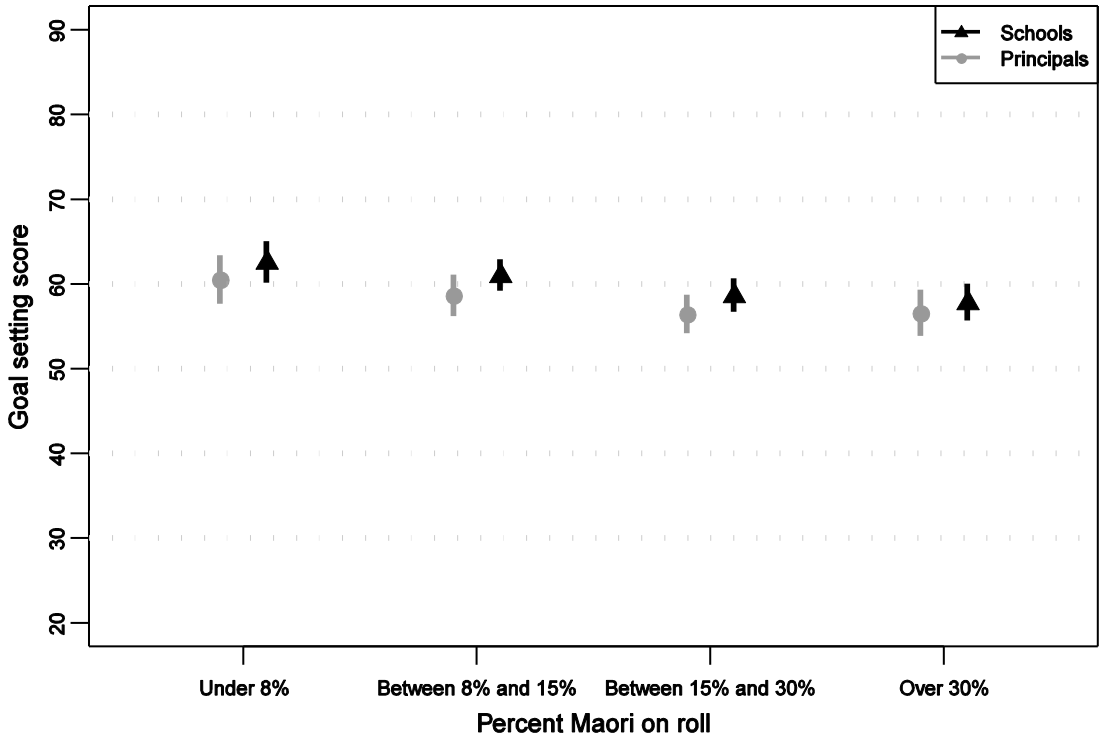
Figure 29 **Goal Setting: School and principal scaled scores' differences according to decile**



Percentage of Māori students on the school roll

Differences in the scaled scores on the Goal Setting scale were less marked in relation to the proportion of Māori enrolment than in relation to the other school characteristics reported above. Schools with less than 8 percent of Māori on the school roll were a little higher in ratings; there was little difference between the categories 15–30 percent and over 30 percent. (There is, of course, an overlap with decile and percentage of Māori on the roll.)

Figure 30 **Goal Setting: School and principal scaled scores by percentage of Māori on the school roll**

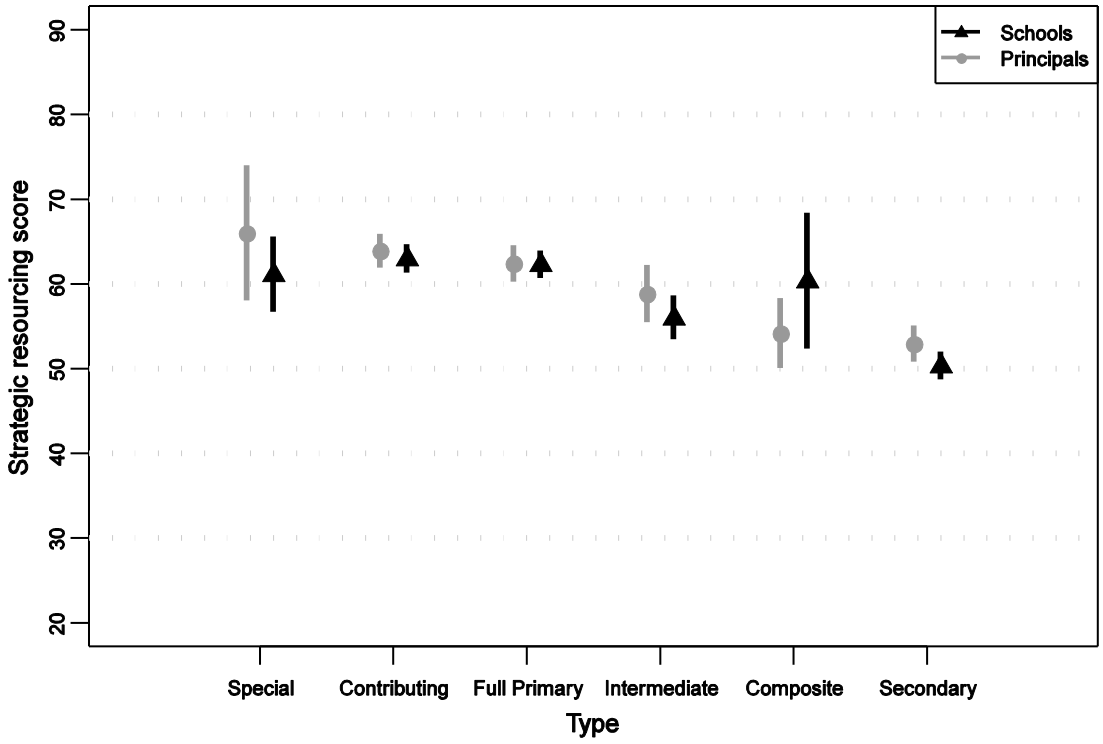


Strategic Resourcing scale

School type

There was around 10 scaled score points difference between the mean for primary schools (both contributing and full primaries) and the mean for secondary schools. Again, special schools and composite schools showed a great deal of variability.

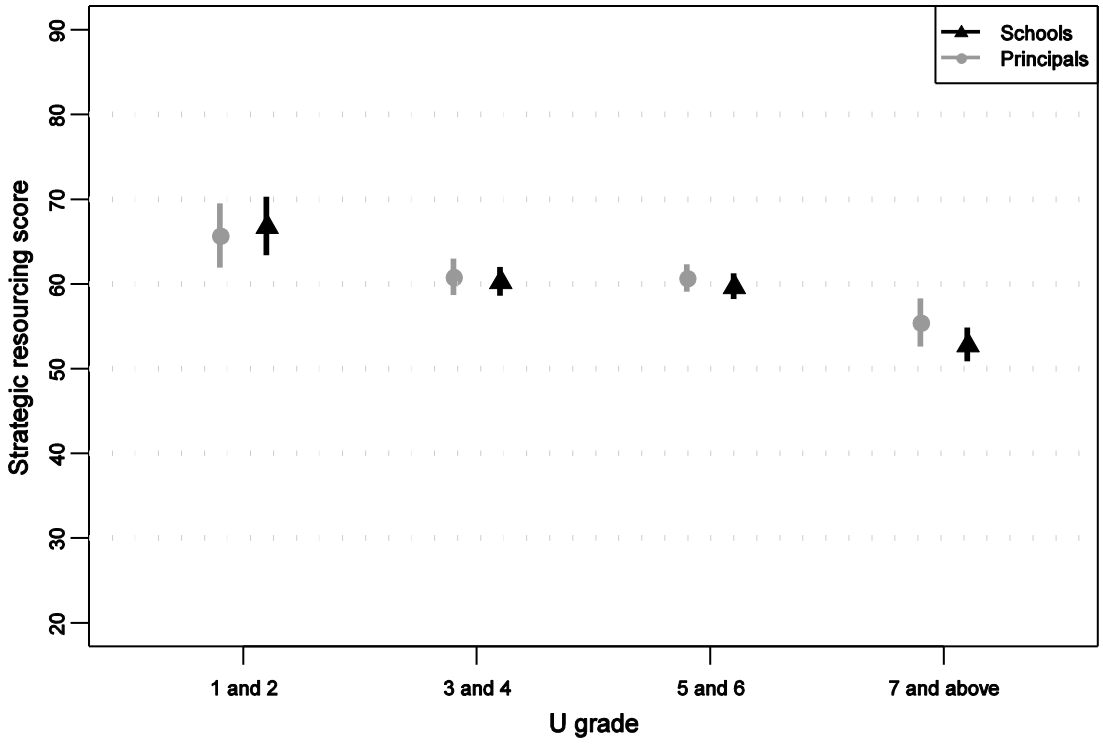
Figure 31 **Strategic Resourcing: School and principal scaled score differences according to school type**



School size

The smallest schools (U1 and U2) have the highest mean scaled scores, there is a dip down to the mean scores for the U3 to U6 schools and U7 and above schools have lower scores (we found small further dips in U9 and above schools, not shown here). See Figure 32 below.

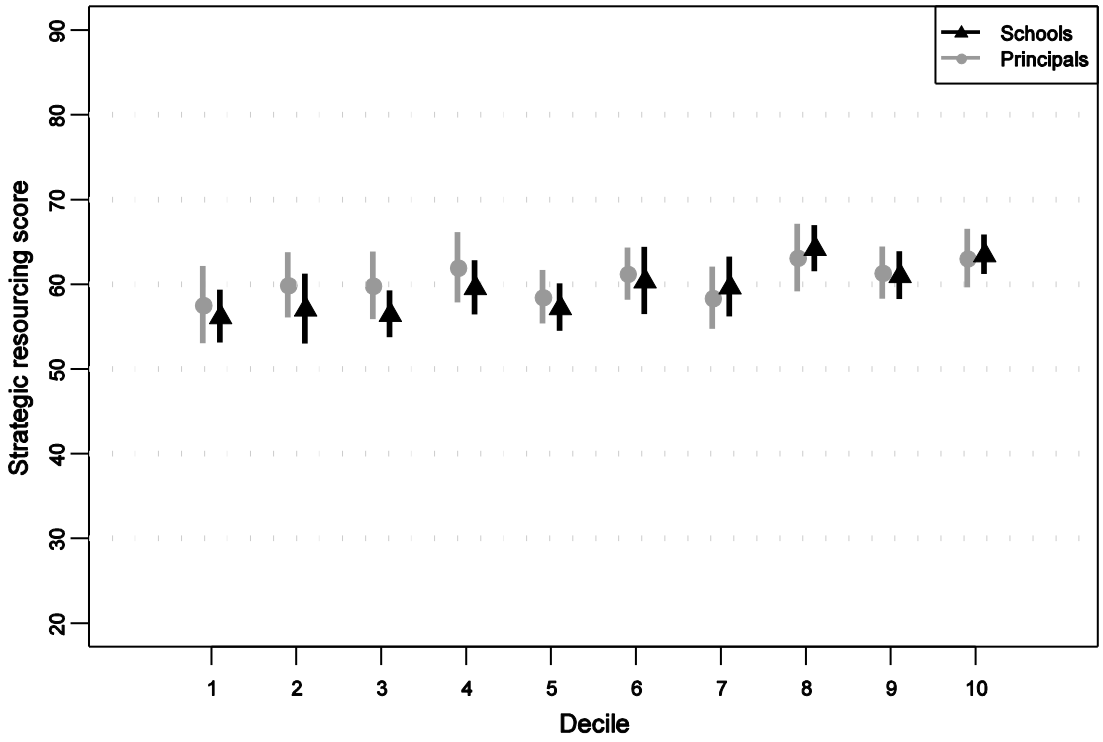
Figure 32 **Strategic Resourcing: School and principal scaled scores according to school size**



Decile differences

While there are higher scaled scores on the Strategic Resourcing scale for decile 10 schools than for decile 1 schools, the association between scores on this scale and school decile is not as clear as we have seen for Goal Setting. For example, there is little difference between decile 4 and decile 9 schools. Looking at principals, there is a trend in deciles 1–6 schools for principals to give higher average ratings than schools.

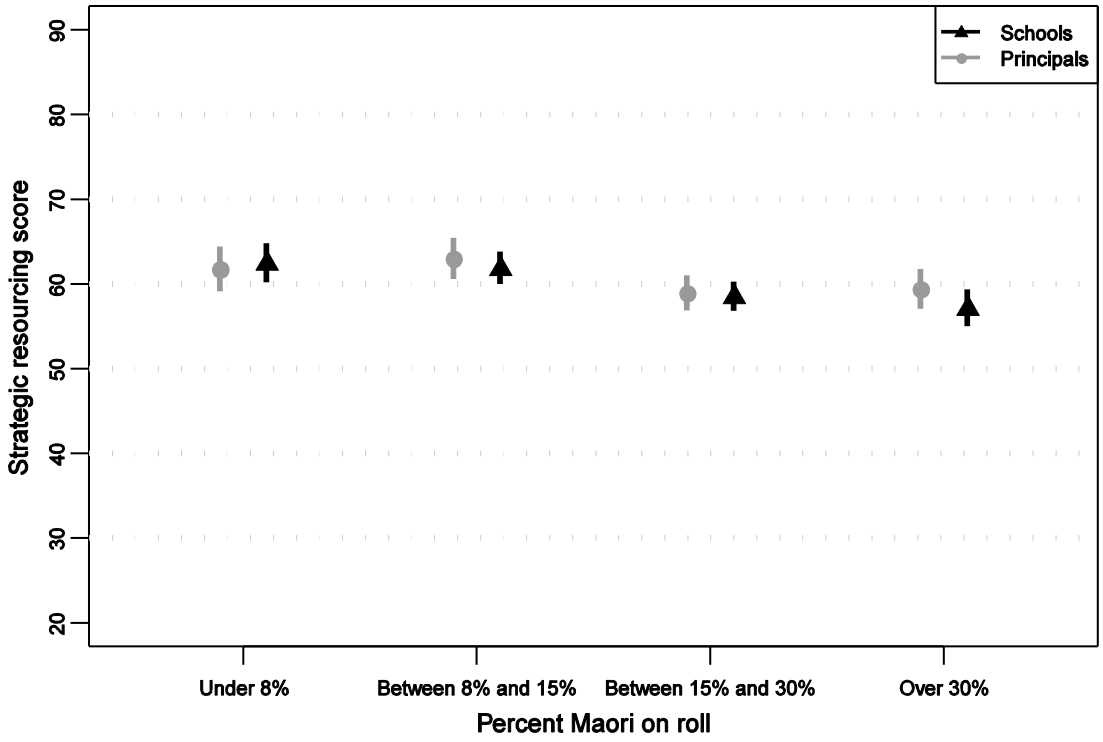
Figure 33 **Strategic Resourcing: School and principal scaled scores according to decile**



Percentage of Māori students on the school roll

Again, the small differences in ratings according to the percentage of Māori on the school roll were in two clusters—under 15 percent Māori, and over 15 percent Māori on the school roll. Principals tended to give lower ratings than schools on the Strategic Resourcing scale in schools with less than 8 percent Māori, but higher ratings in schools with over 30 percent Māori enrolment.

Figure 34 **Strategic Resourcing: School and principal scaled scores according to percentage of Māori on the school roll**

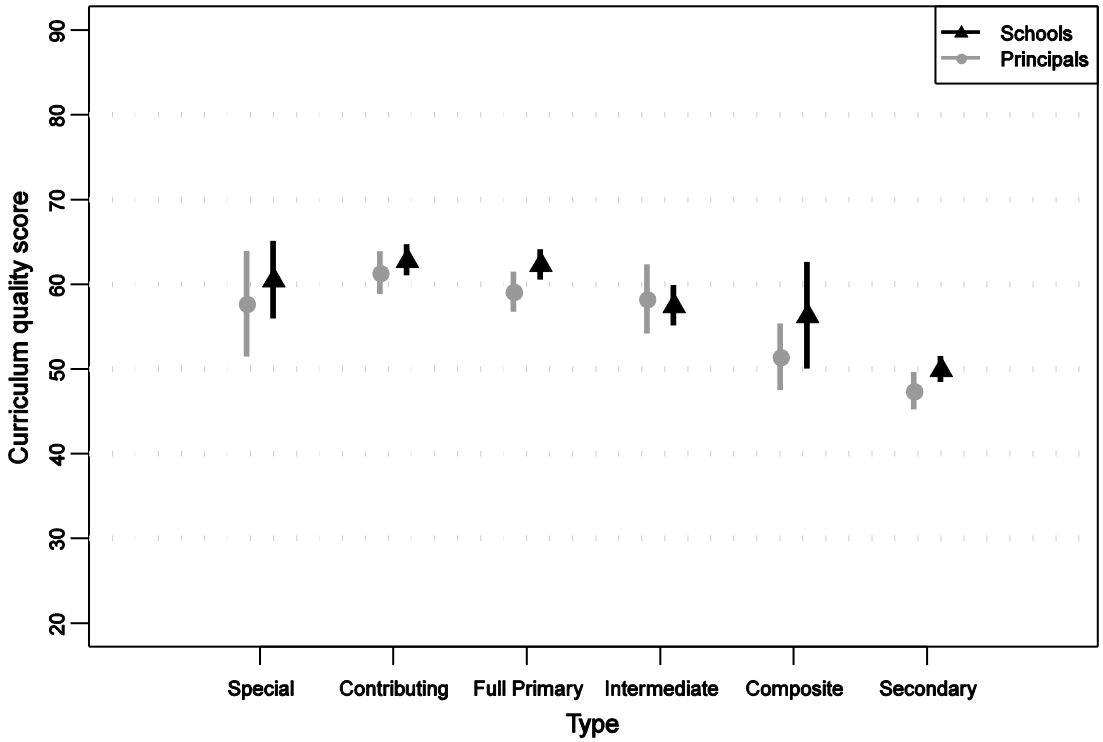


Curriculum Quality scale

School type

Secondary schools again had the lowest mean scaled scores on the Curriculum Quality scale—more than 10 scaled score points less than contributing and full primary schools. Composite schools sat between primary and secondary in ratings, though a wide error of measurement is noted in the composite school scores.

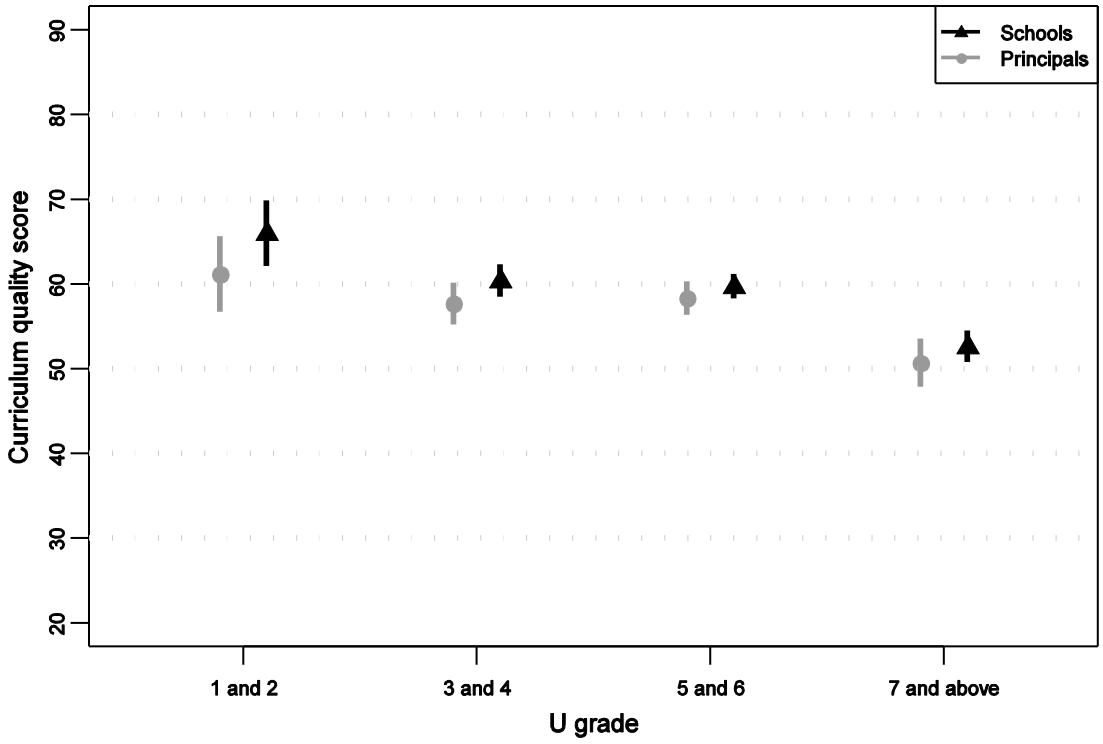
Figure 35 **Curriculum Quality: School and principal scaled scores according school type**



School size

School size shows a clear relationship with scores on the Curriculum Quality scale, with more than 15 scaled score points between the means of U1 and U2 schools and U7 and above schools.

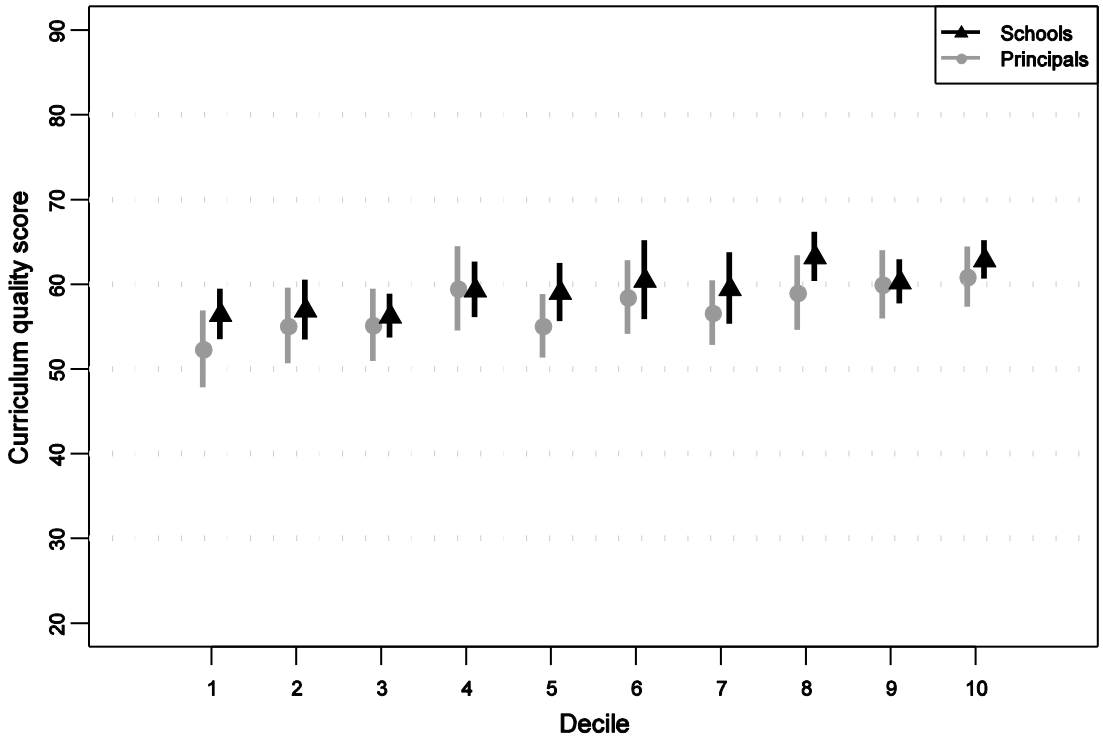
Figure 36 **Curriculum Quality: School and principal scaled scores according to school size**



Decile differences

We see the same pattern we saw for the other scales in relation to the Curriculum Quality scale, with lower average scaled scores in the low-decile schools, and higher average scaled scores in the high-decile schools just under 10 scaled score points between deciles 1 and 2 schools, and decile 10 schools (with some variation in differences in deciles 3–8 schools).

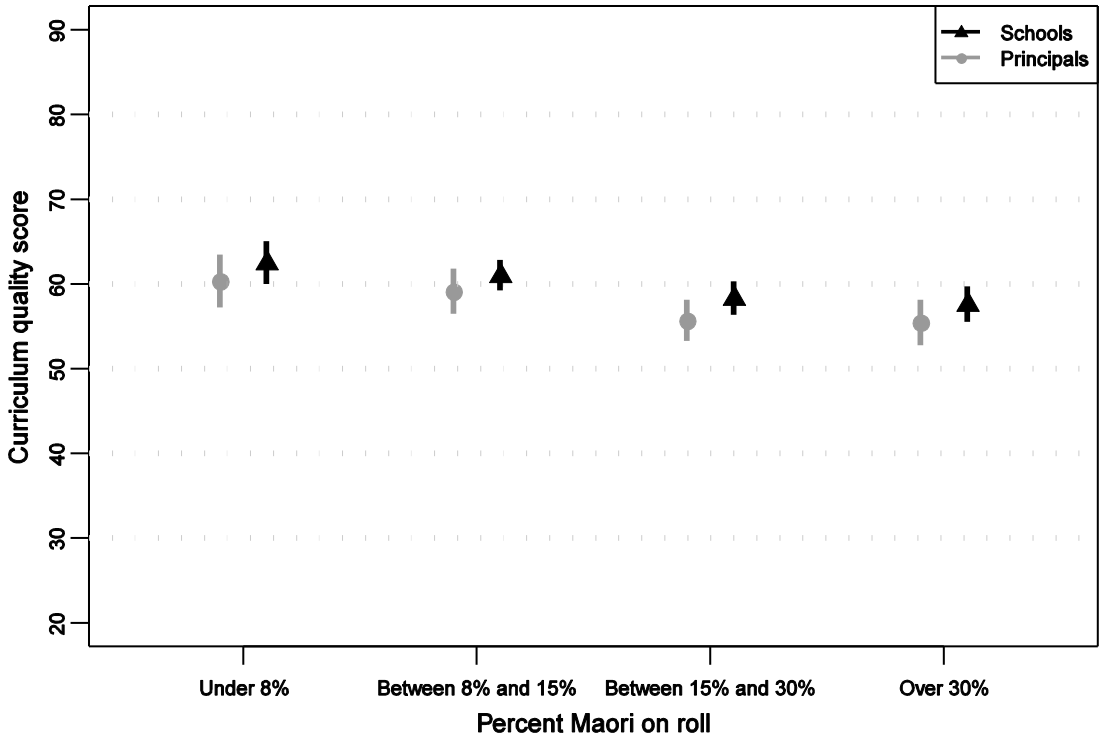
Figure 37 **Curriculum Quality: School and principal scaled scores according to decile**



Percentage of Māori students on the school roll

Differences in Curriculum Quality scaled scores were relatively small when we examined scores in relation to different proportions of Māori enrolment. They were also a little less marked between schools than between principals.

Figure 38 **Curriculum Quality: School and principal scaled scores according to percentage of Māori on the school roll**

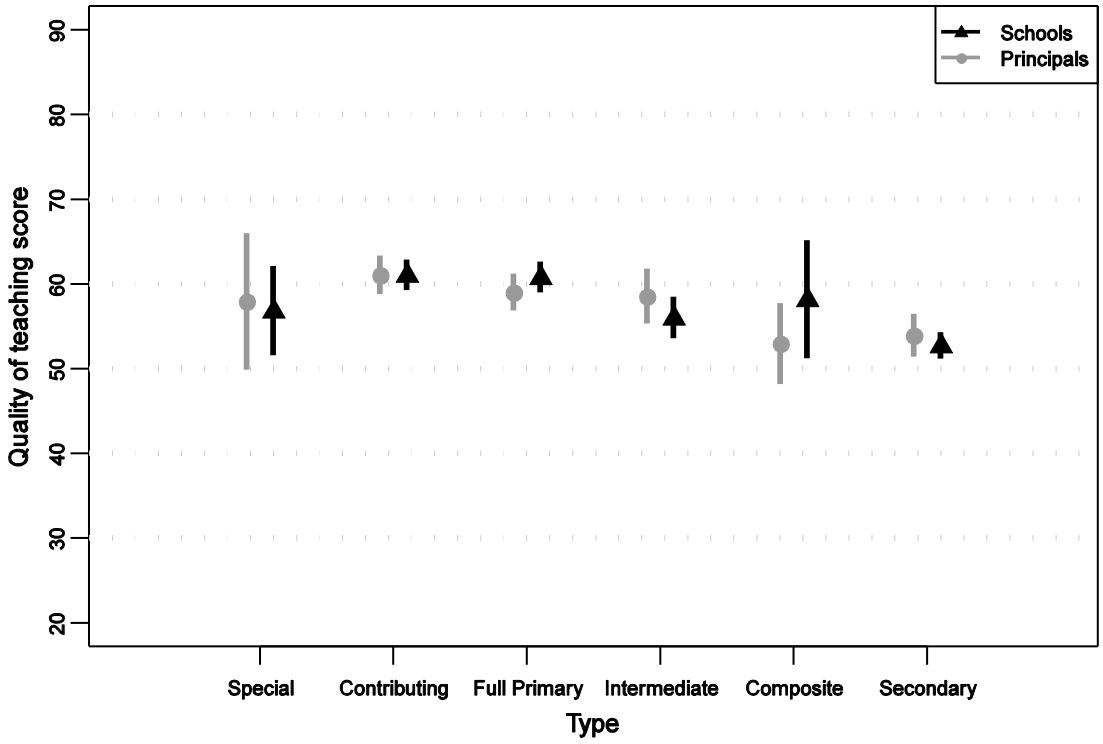


Quality of Teaching scale

School type

We see similar trends in this scale as in other scales—secondary schools had lower means in scaled scores—around 8 scaled score points between secondary schools and full and contributing primary schools, with the intermediate school scaled score mean sitting between primary and secondary school means. Secondary and intermediate principals tended to give higher ratings on the Quality of Teaching scale than secondary and intermediate schools. (The same difference is apparent for special school principals, but the error of measurement is too high to be confident about this difference.)

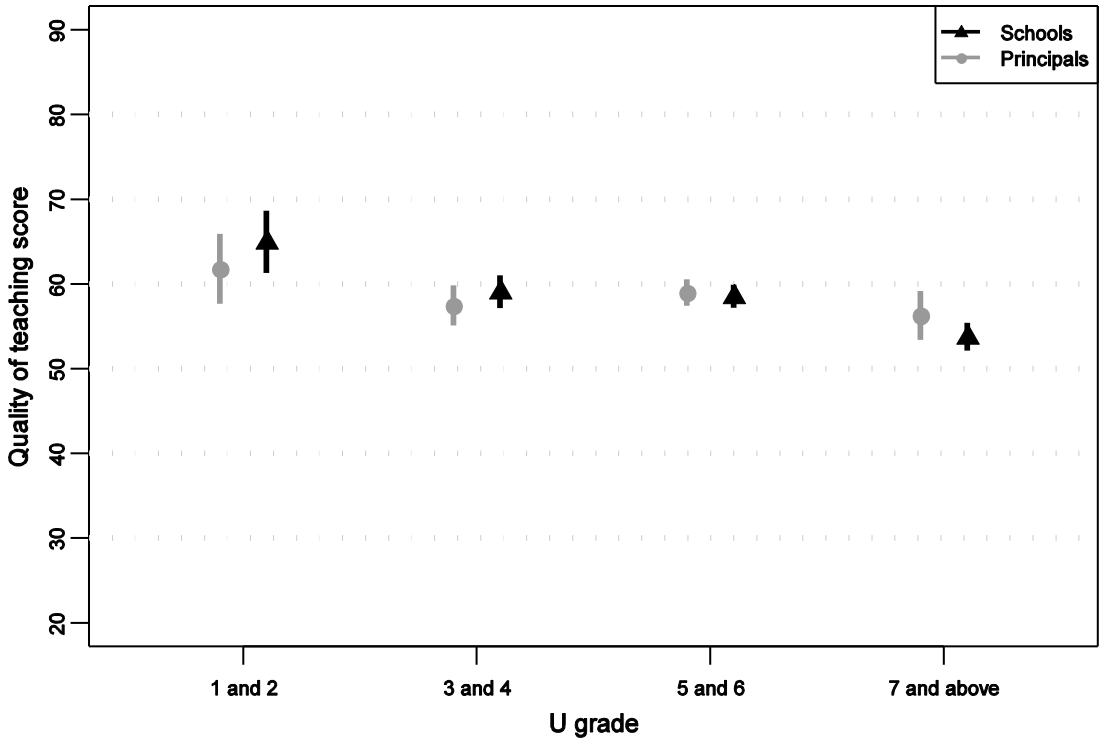
Figure 39 **Quality of Teaching: School and principal scaled scores according to school type**



School size

Principals were more positive in their ratings in the U7 and above schools—this is similar to the findings for school type, which is not surprising since a large proportion of the U7 and above schools would be secondary schools. But, as we have seen with the other ELP scales, ratings in the largest schools were the lowest on average.

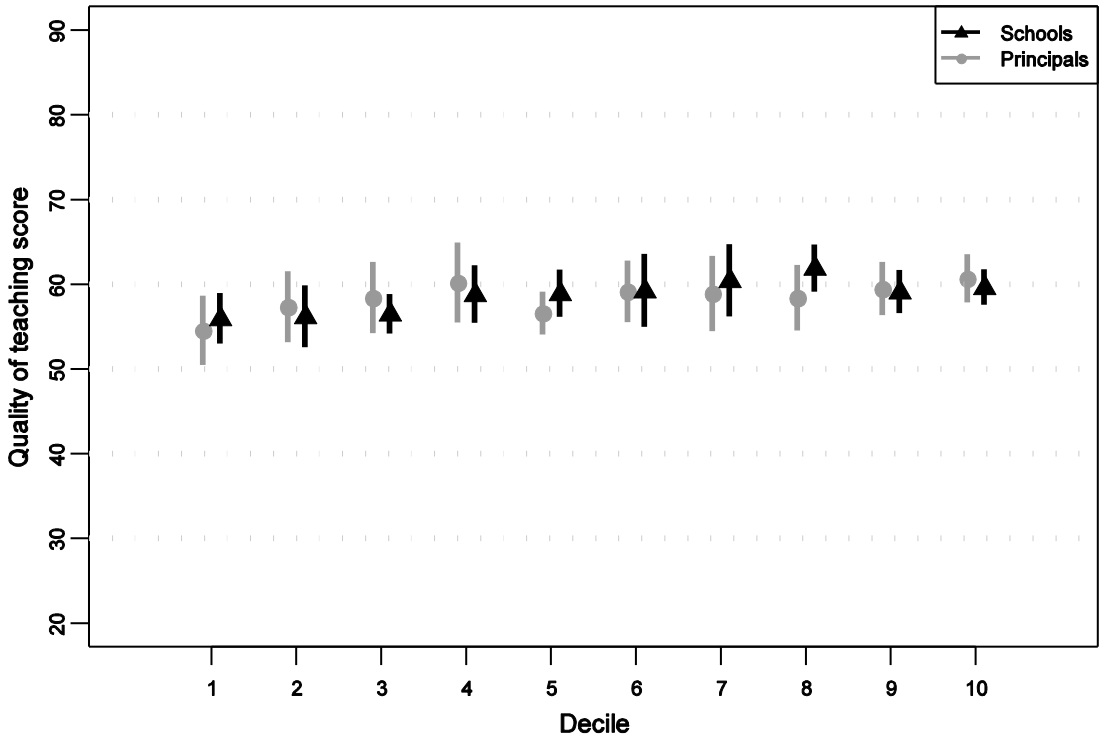
Figure 40 **Quality of Teaching: School and principal scaled scores according to school size**



Decile differences

The mixed pattern across schools according to decile is again evident in relation to Quality of Teaching, as shown in Figure 39 below. Deciles 1–3 schools have somewhat lower ratings than others.

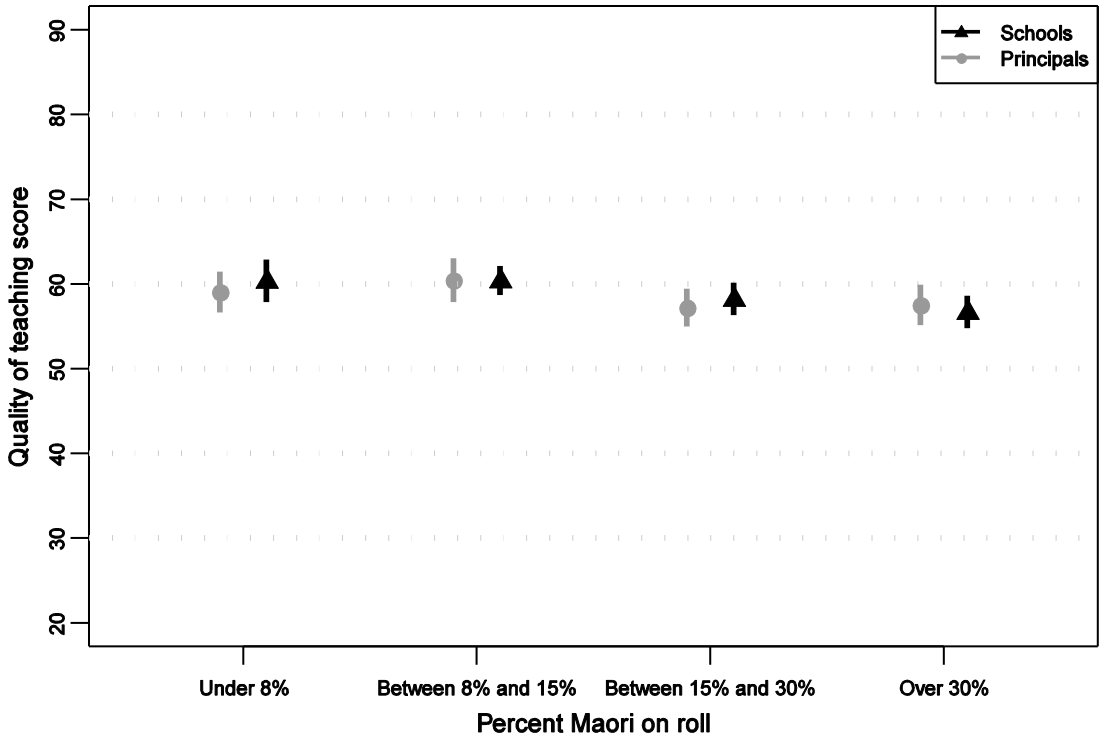
Figure 41 **Quality of Teaching: Principal and school scaled scores by decile**



Percentage of Māori students on the school roll

There was a small difference between schools with less than 15 percent Māori enrolment, and those with more than 15 percent Māori enrolment in relation to the Quality of Teaching scale.

Figure 42 **Quality of Teaching: School and principal views according to percentage of Māori on the school roll**

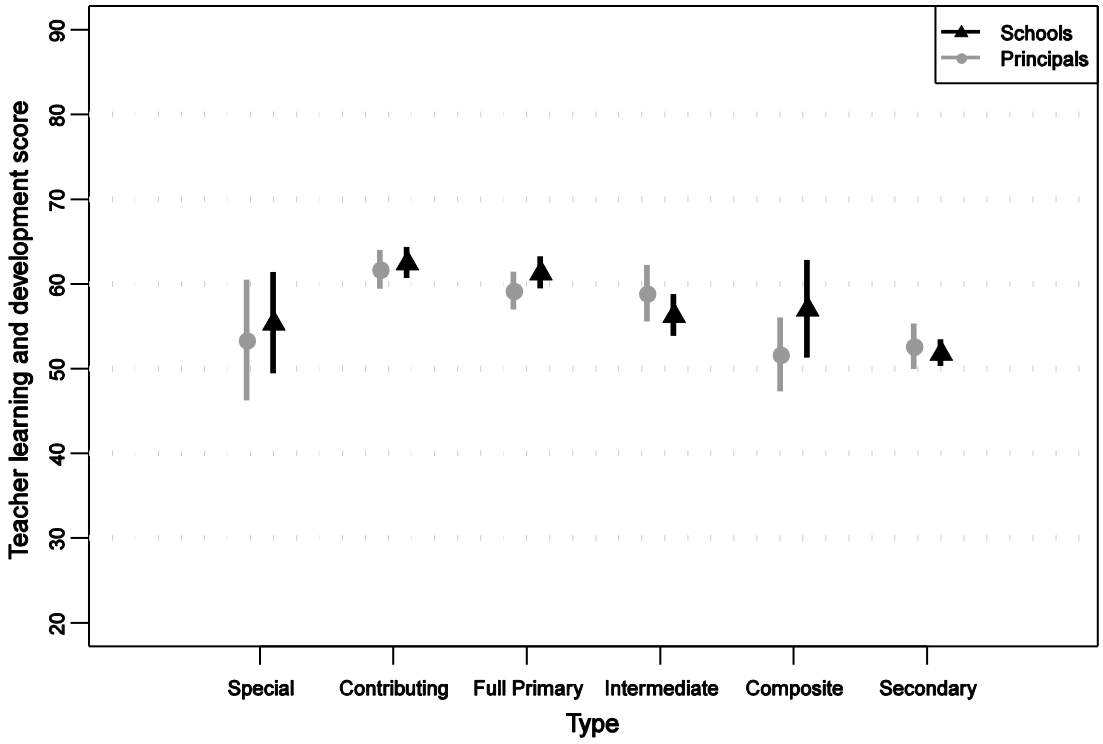


Promoting and Participating in Teacher Learning and Development scale

School type

Secondary schools' average score on the Promoting and Participating in Teacher Learning and Development scale was approximately 10 scaled score points lower than primary schools. Secondary and intermediate principals tended to give somewhat higher ratings than secondary schools (staff).

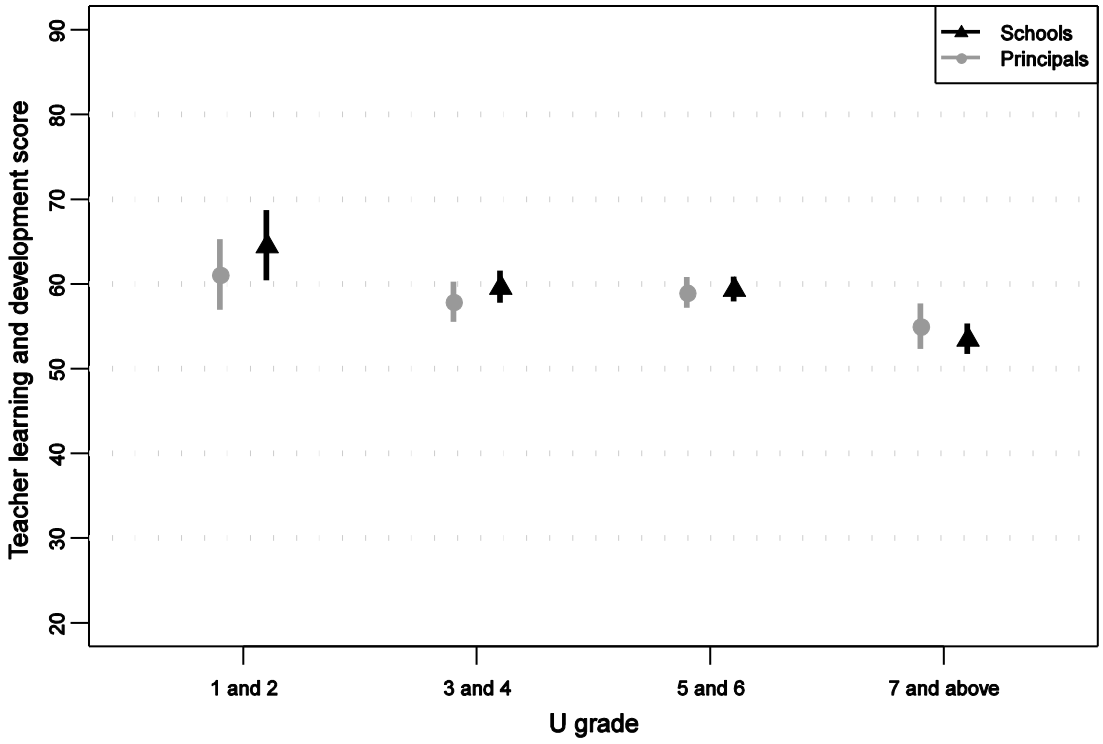
Figure 43 **Promoting and Participating in Teacher Learning and Development: School and principal scaled scores according to school type**



School size

School size patterns are very similar to the school type patterns again most likely because of the high number of larger schools that are secondary schools.

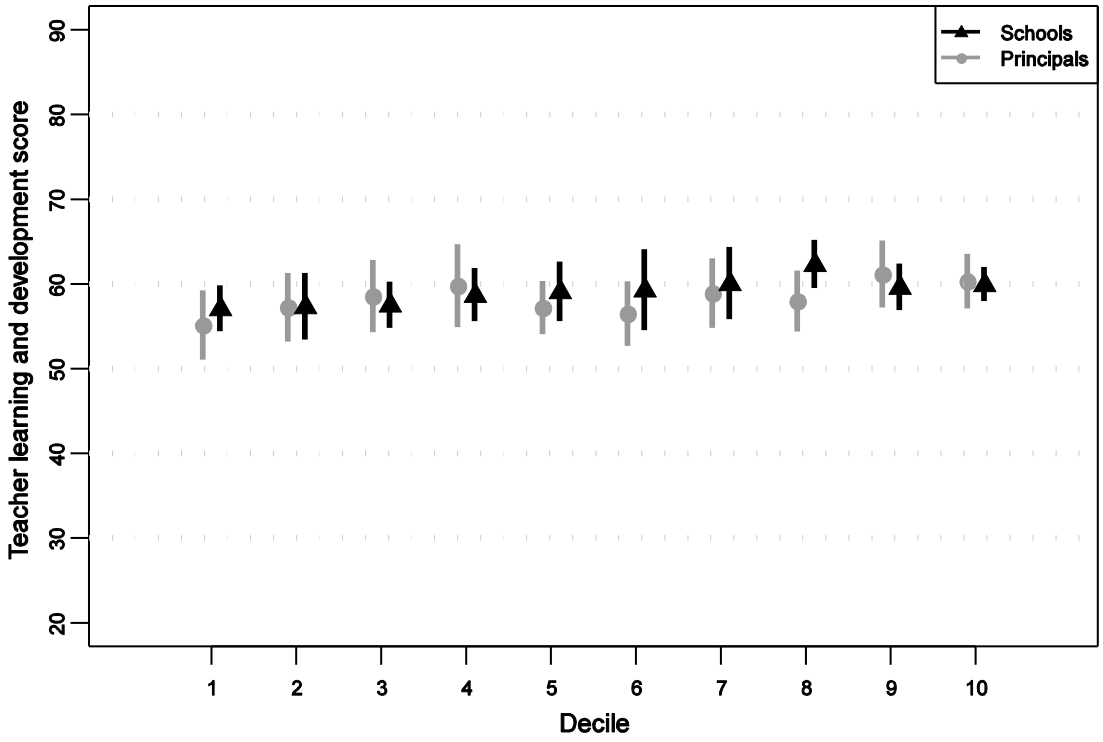
Figure 44 **Promoting and Participating in Teacher Learning and Development: School and principal scaled scores according to school size**



Decile differences

The patterns evident in the other scales in relation to decile were also evident in relation to differences in average scaled scores on the Promoting and Participating in Teacher Learning and Development scale. The differences, though, are not large.

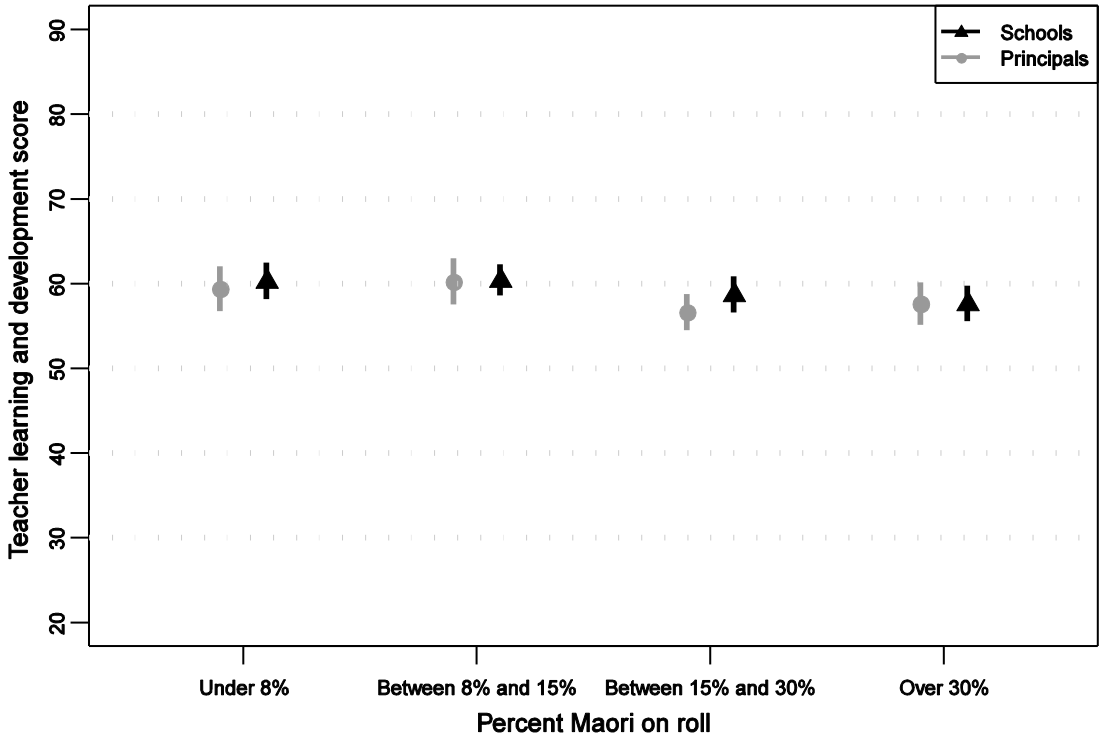
Figure 45 **Promoting and Participating in Teacher Learning and Development: School and principal scale scores according to decile**



Percentage of Māori students on the school roll

The differences across schools according to the percentage of Māori students on the school roll were minimal—again with a small trend towards a slightly lower average scaled score when the percentage of Māori on the school roll was over 15 percent.

Figure 46 **Promoting and Participating in Teacher Learning and Development: School and principal scaled scores according to percentage of Māori on the school roll**

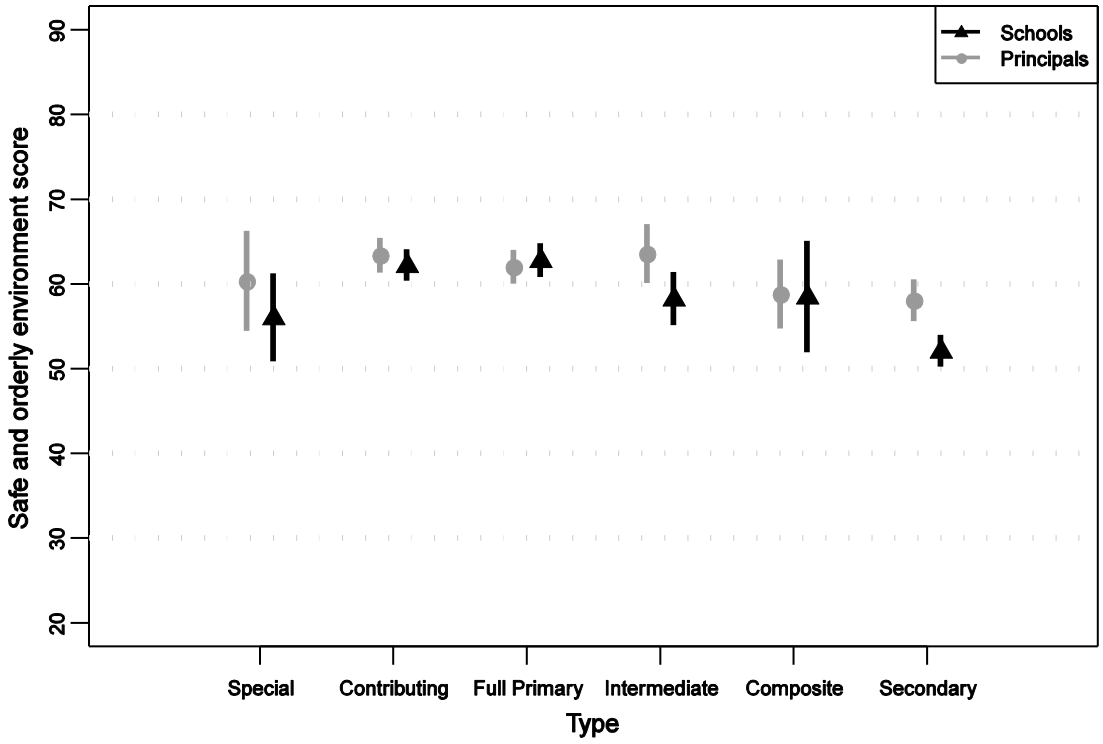


Ensuring a Safe and Orderly Environment scale

School type

On this scale there were more than 10 scaled score points between the means for secondary and primary schools. Secondary and intermediate principals were more likely to give higher ratings to the Ensuring a Safe and Orderly Environment scale items than secondary and intermediate schools.

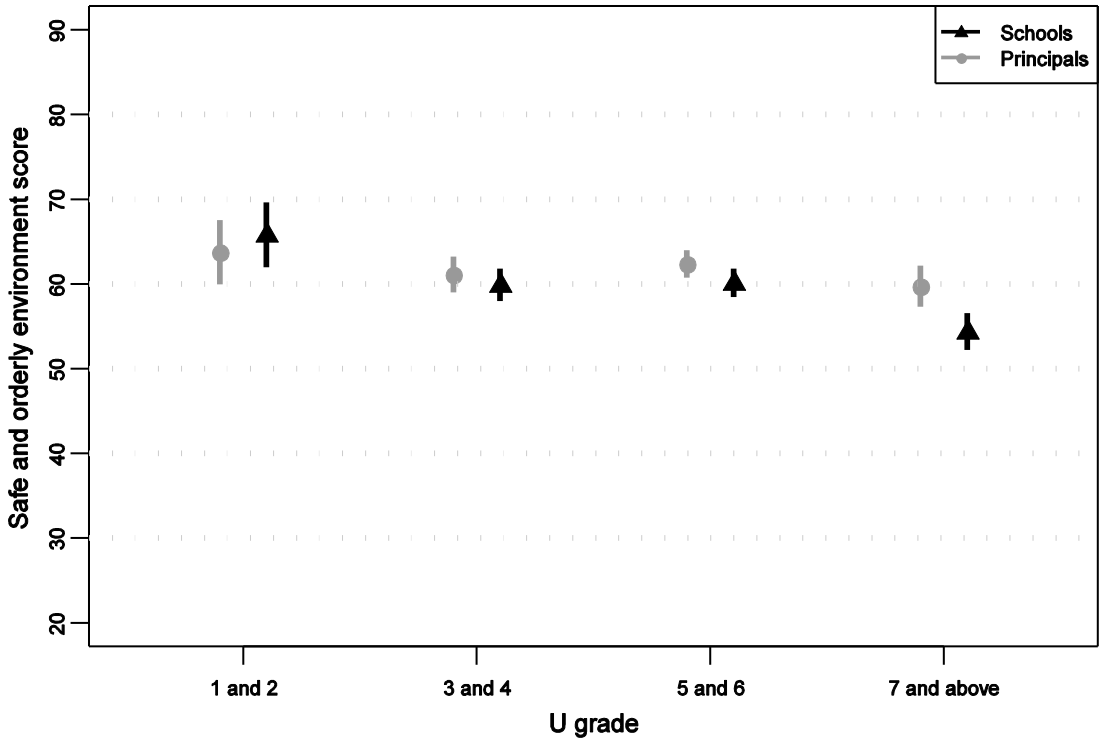
Figure 47 **Ensuring a Safe and Orderly Environment: School and principal mean scaled scores according to school type**



School size

Once again the patterns are similar to those for school type; around 10 scaled points between the smallest and largest schools. Principals gave higher ratings than schools in U7 and above schools (where most of the secondary schools are grouped).

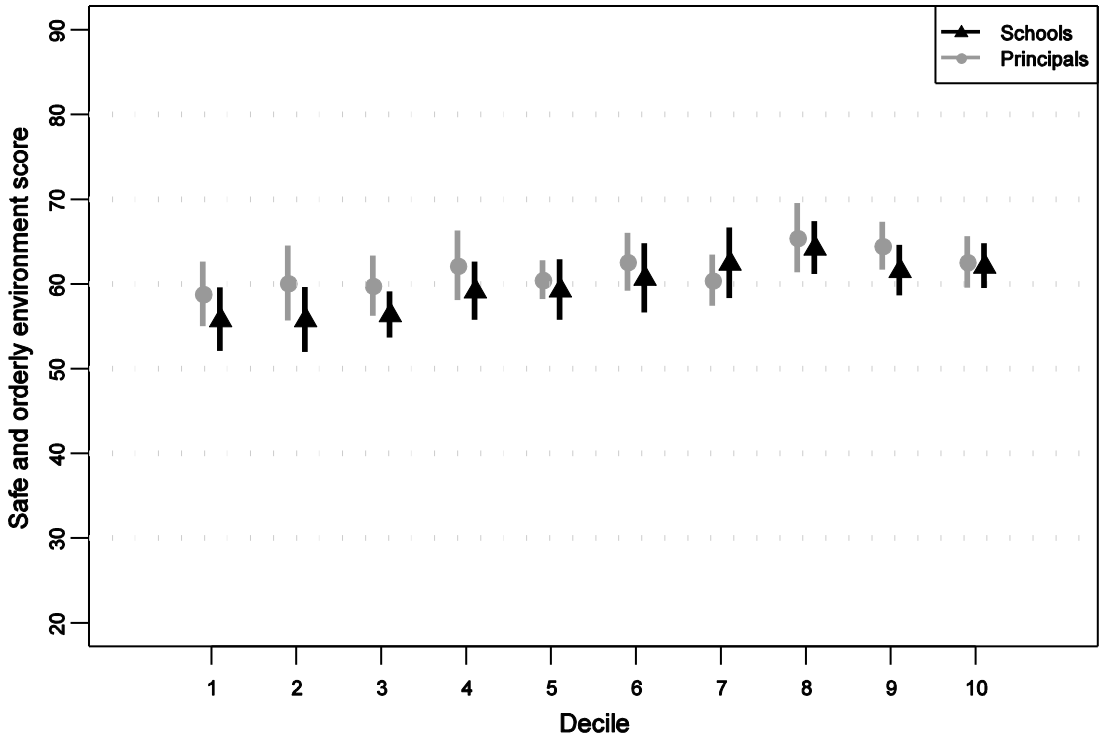
Figure 48 **Ensuring a Safe and Orderly Environment: School and principal mean scaled scores according to school size**



Decile differences

We see the same pattern we have seen for the other ELP scales with lowest ratings given by low-decile schools. Principals tended to give higher ratings than staff across deciles, but particularly so in deciles 1–4 schools.

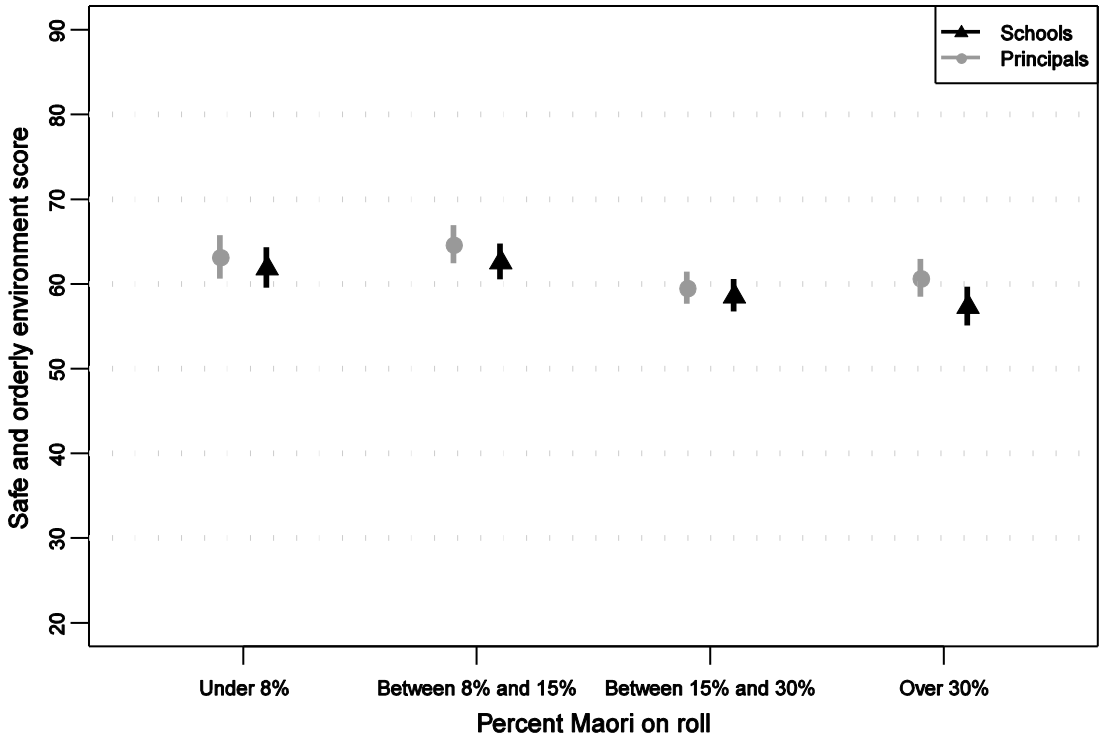
Figure 49 **Ensuring a Safe and Orderly Environment: School and principal mean scaled scores according to decile**



Percentage of Māori students on the school roll

Principals again tended to give higher ratings on the effectiveness of school leadership in ensuring a safe and orderly environment than schools. Again, there were two groups in terms of ratings of effectiveness—schools with less than 15 percent Māori on the school roll gave slightly higher ratings than schools with more than 15 percent Māori students on the school roll.

Figure 50 **Ensuring a Safe and Orderly Environment: School and principal mean scaled scores according to percentage of Māori on the school roll**

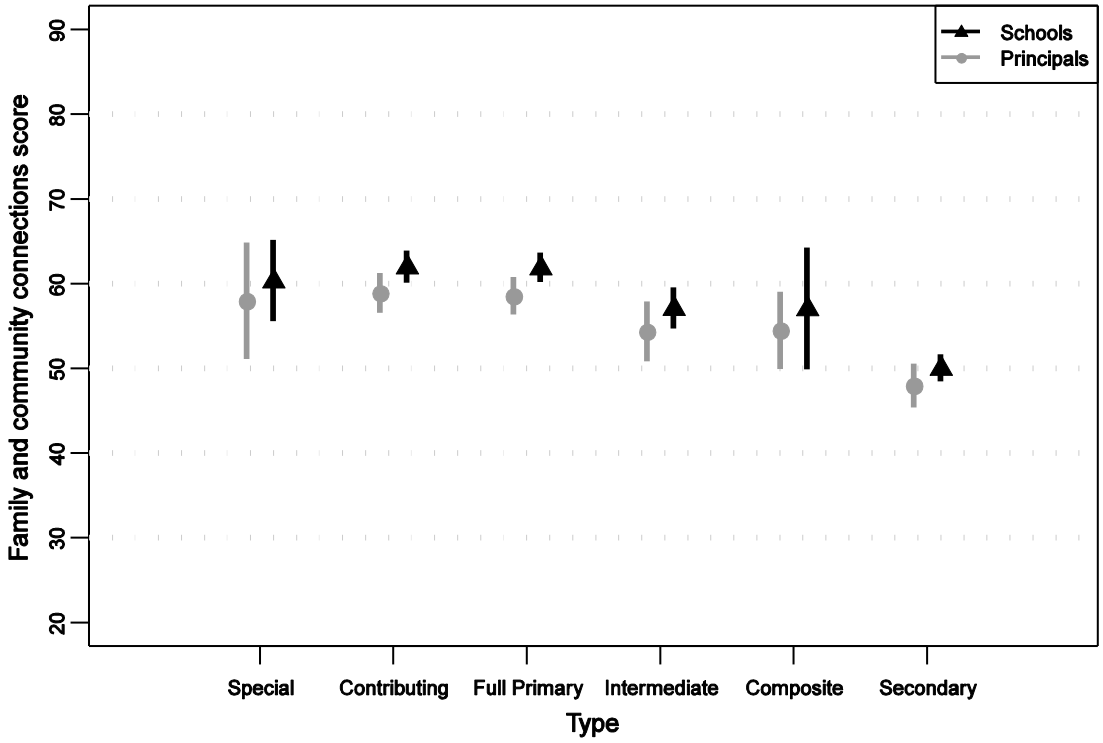


Ensuring Educationally Powerful Connections with Families, Whānau and Community scale

School type

Schools consistently rated more highly than principals the school's effectiveness in educational leadership in Ensuring Educationally Powerful Connections with Family, Whānau and Community. Again, there were more than 10 scaled point differences in the means between primary schools and secondary schools.

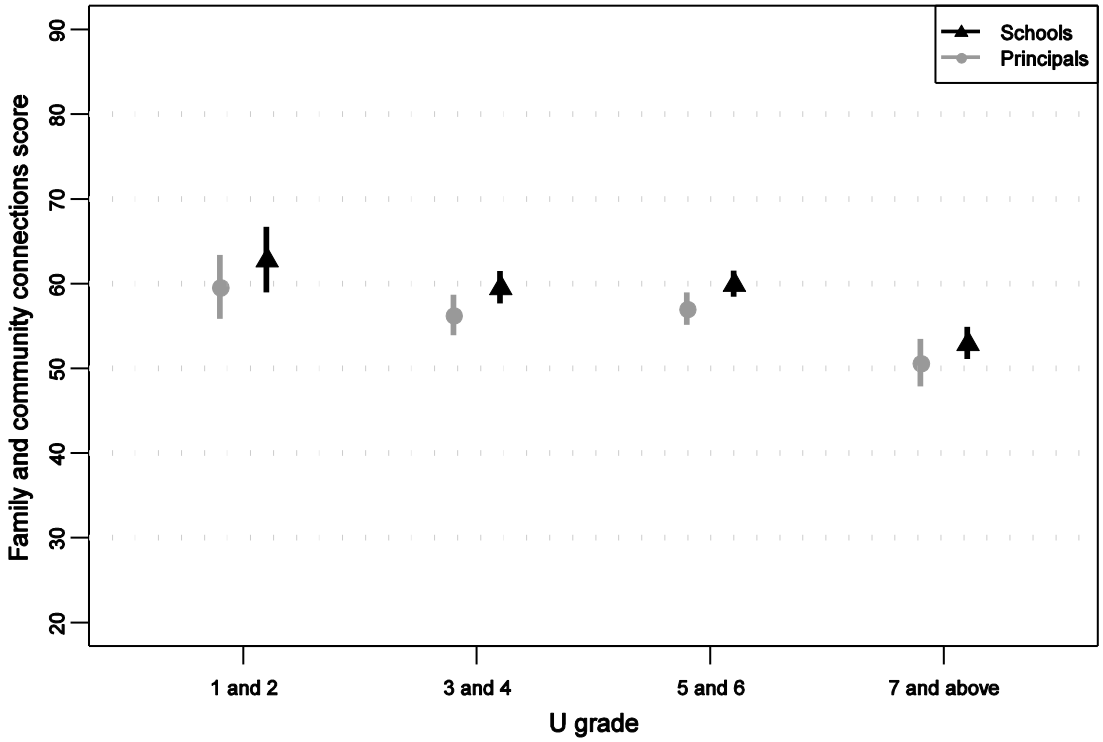
Figure 51 **Ensuring Educationally Powerful Connections with Family, Whānau and Community: School and principal mean scaled scores according to school type**



School size

There were similar mean scaled scores for school leadership effectiveness in Ensuring Educationally Powerful Connections with Family, Whānau and Community across U1 to U6 schools with a noticeable dip at U7 and above-sized schools.

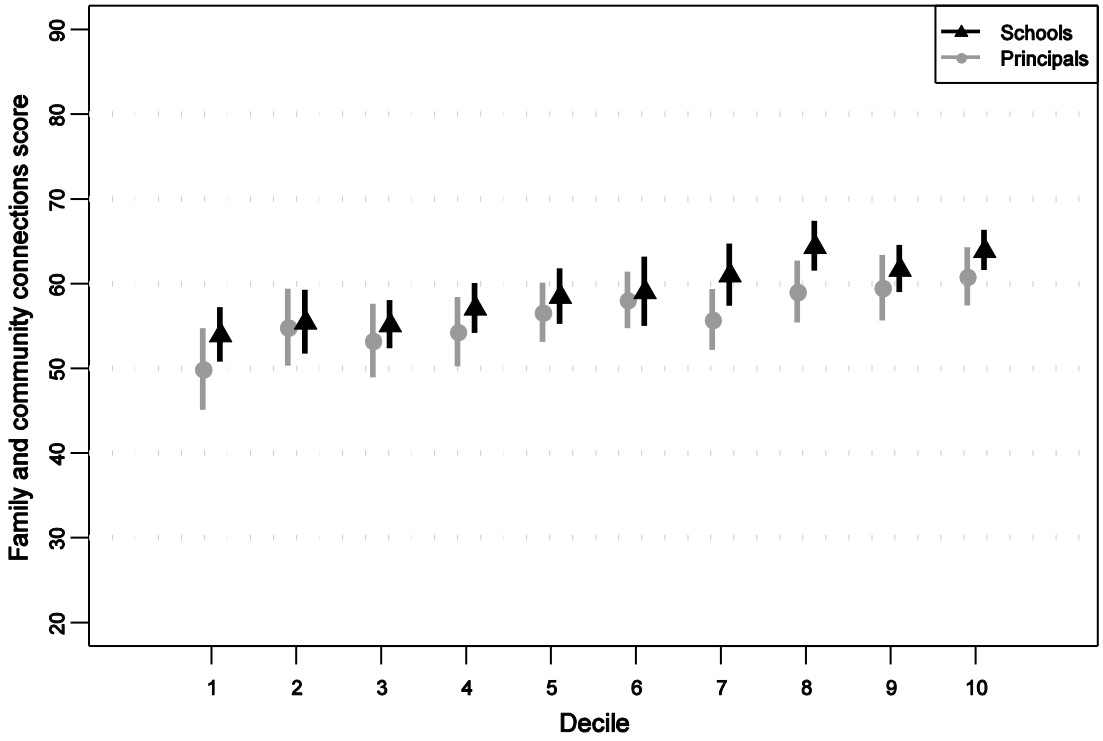
Figure 52 **Ensuring Educationally Powerful Connections with Family, Whānau and Community: School and principal mean scaled scores according to school size**



Decile differences

A consistent trend towards higher ratings in school leadership ensuring effectiveness in Ensuring Educationally Powerful Connections with Family, Whānau and Community across decile is noticeable; deciles 1–3 schools again group in one cluster with their scaled score means around 55 points. Deciles 6–10 schools' mean scaled score ranged between 60 and 63. Principals rated school leadership higher than schools in this scale in deciles 1–4 schools.

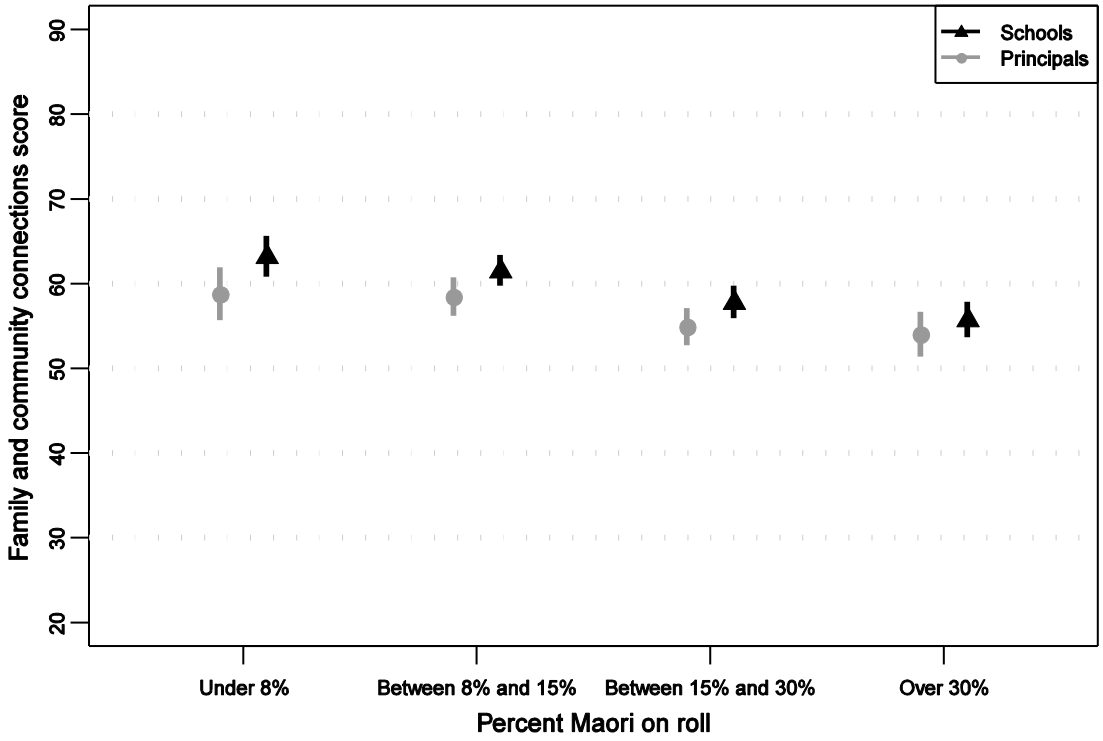
Figure 53 **Ensuring Educationally Powerful Connections with Family, Whānau and Community: School and principal mean scaled scores according to school decile**



Percentage of Māori students on the school roll

The most marked differences here in effectiveness ratings were between schools with less than 8 percent Māori on the school roll, and schools with more than 30 percent Māori on the school roll with around 5 scaled score points difference in the means.

Figure 54 **Ensuring Educationally Powerful Connections with Family, Whānau and Community: School and principal mean scaled scores accordingly to percentage of Māori on the school roll**



Māori success scale

In the area of Māori success we have used the 1–5 scale as the unit of measurement. As noted in Section 2, the overall means for this aspect of educational leadership were lower. Figures 55 to 58 show little variation related to school characteristics, unlike the other ELP scales. The smallest schools do have higher scores than others, but there is no difference between average secondary and primary school scores. There is a small variation across decile with the trend in the other direction from other ELP scales: higher decile schools have slightly lower means than deciles 1–3 schools. Schools with more than 30 percent Māori enrolment had (only) slightly higher means than other schools.

Figure 55 **Māori Success: School and principal ratings according to school type**

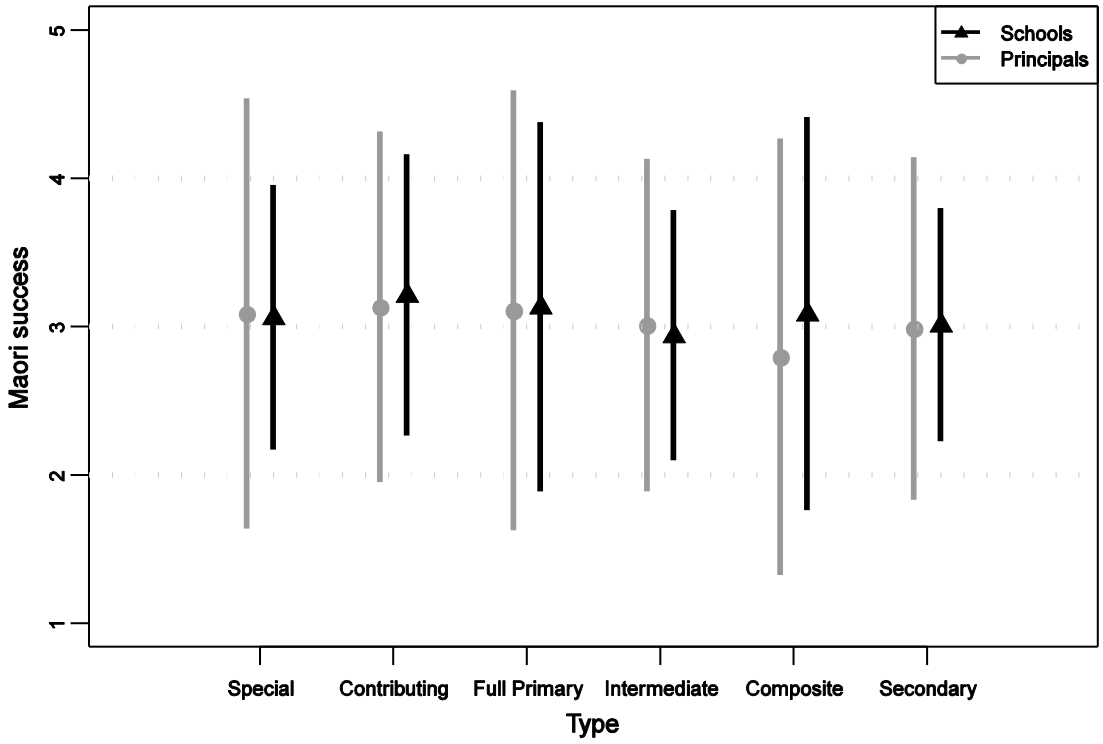


Figure 56 **Māori Success: School and principal ratings according to school size**

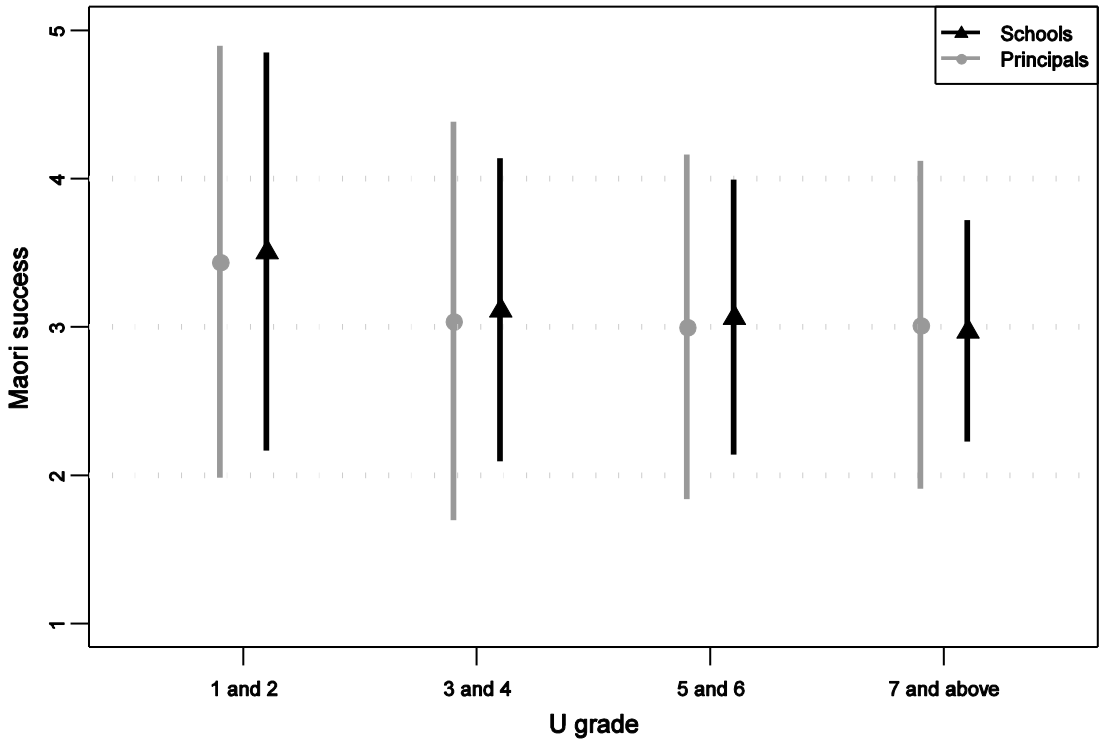


Figure 57 **Māori Success: School and principal ratings according to school decile**

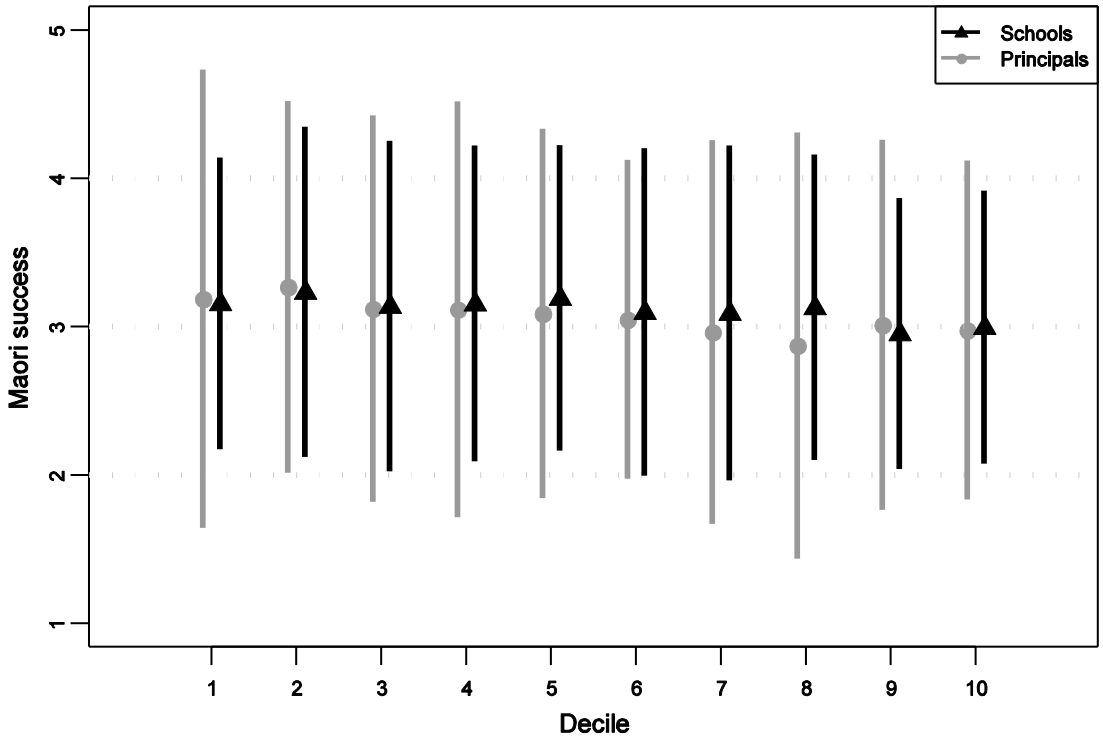
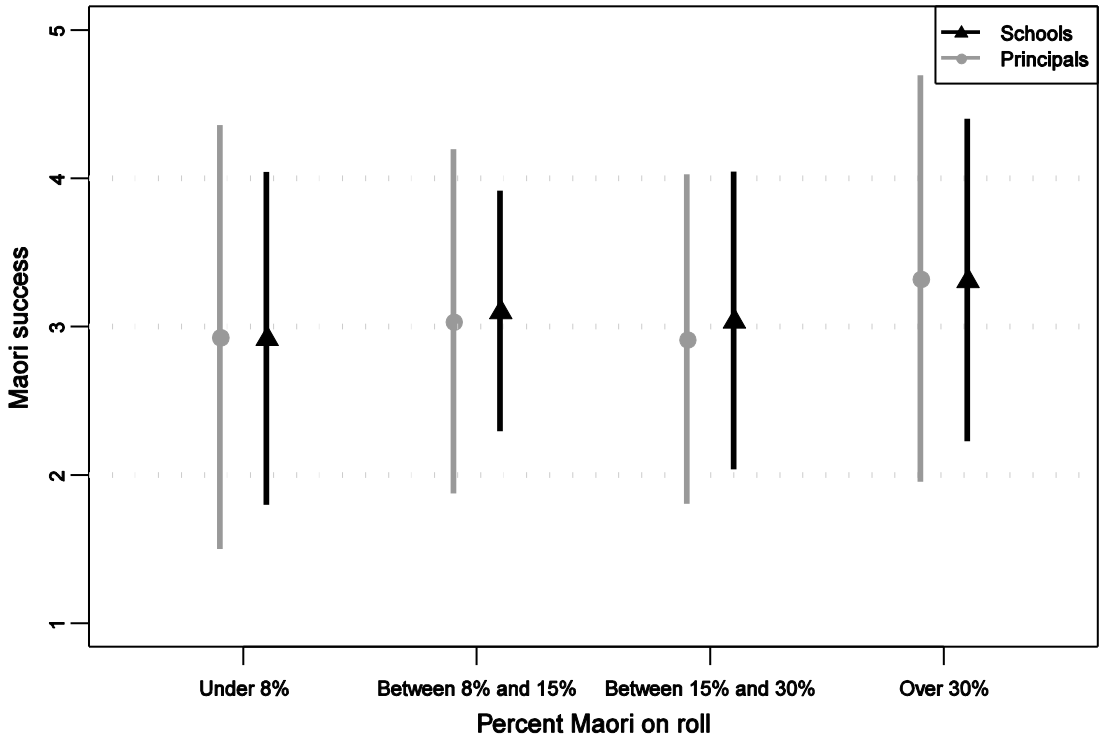


Figure 58 **Māori Success: School and principal ratings according to percentage of Māori on the school roll**

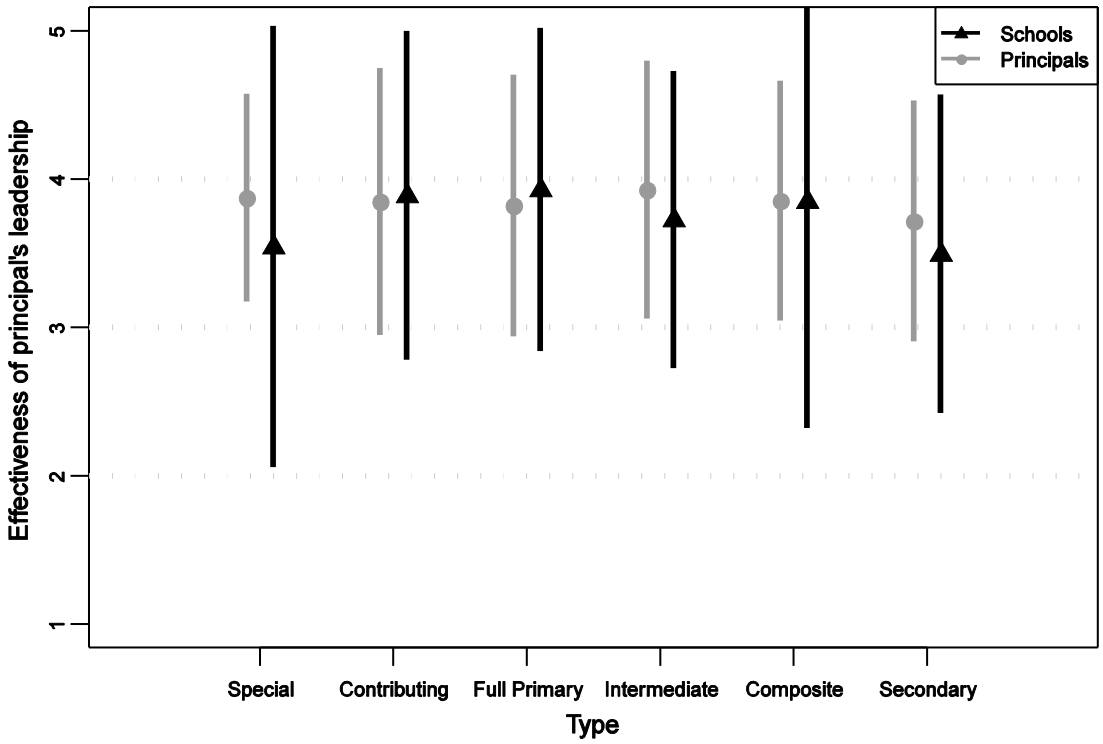


Principal Leadership scale

School type

Secondary and special school Principal Leadership was less highly rated by schools, with little difference evident between primary and composite schools. There was a slight tendency for secondary, special and intermediate school principals to rate Principal Leadership higher than schools did (see Figure 59).

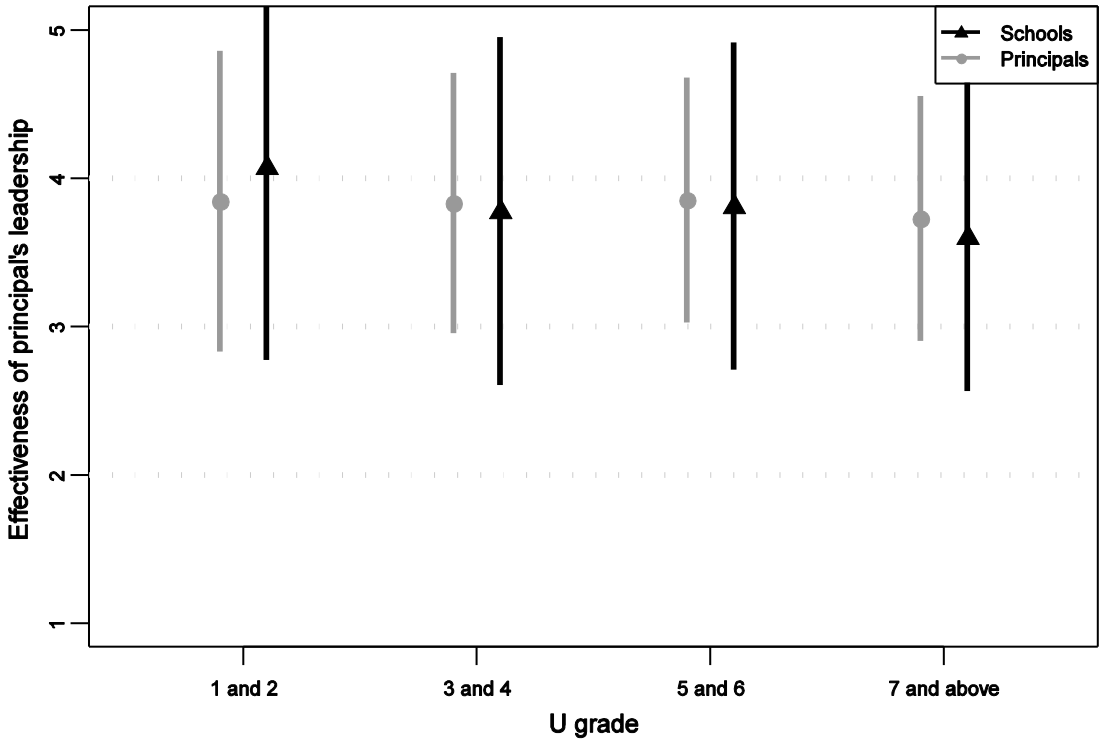
Figure 59 **Principal Leadership according to school type**



School size (U-grade) differences

There were no differences in principals' ratings according to school size, but larger schools gave lower ratings. Although not shown on Figure 60 below, U9 and above means were very similar to the means of U7 and above schools (in other words the lowering trend flattened out around U7 size) (see Figure 60).

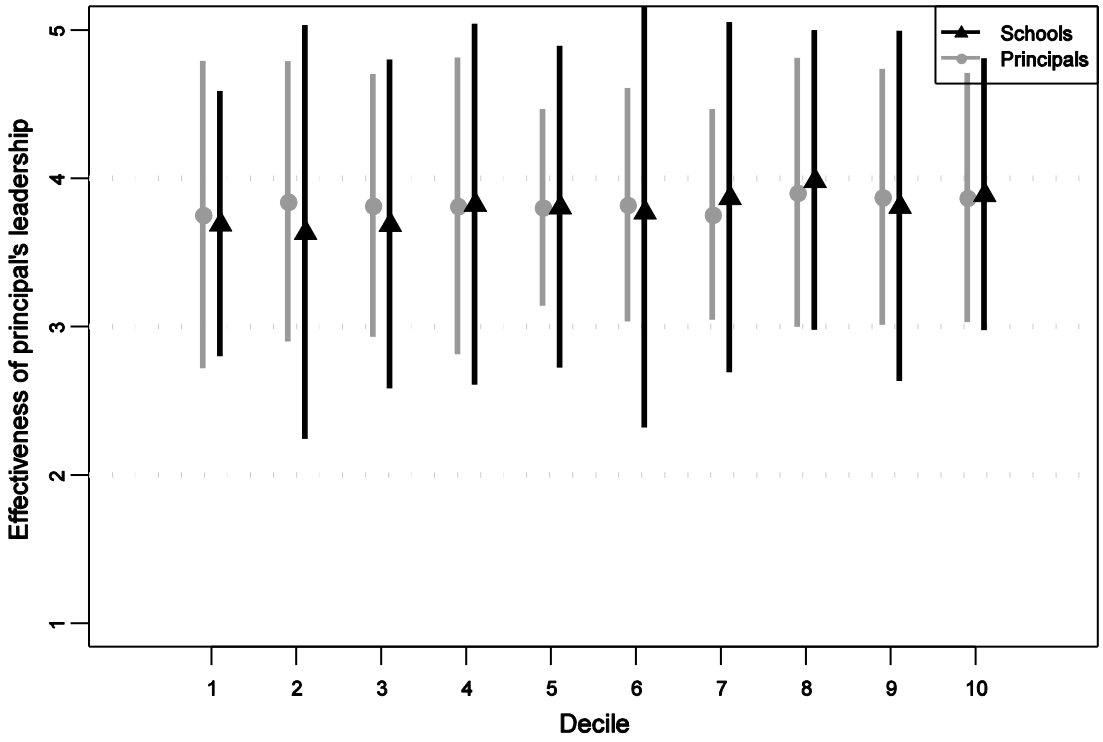
Figure 60 **Principal Leadership according to school size**



Decile differences

Slight differences are apparent in school ratings between deciles 1–3 schools and higher decile schools. Principal ratings, however, showed no clear patterns of difference across deciles in relation to their ratings of the effectiveness of their own leadership (see Figure 61 below).

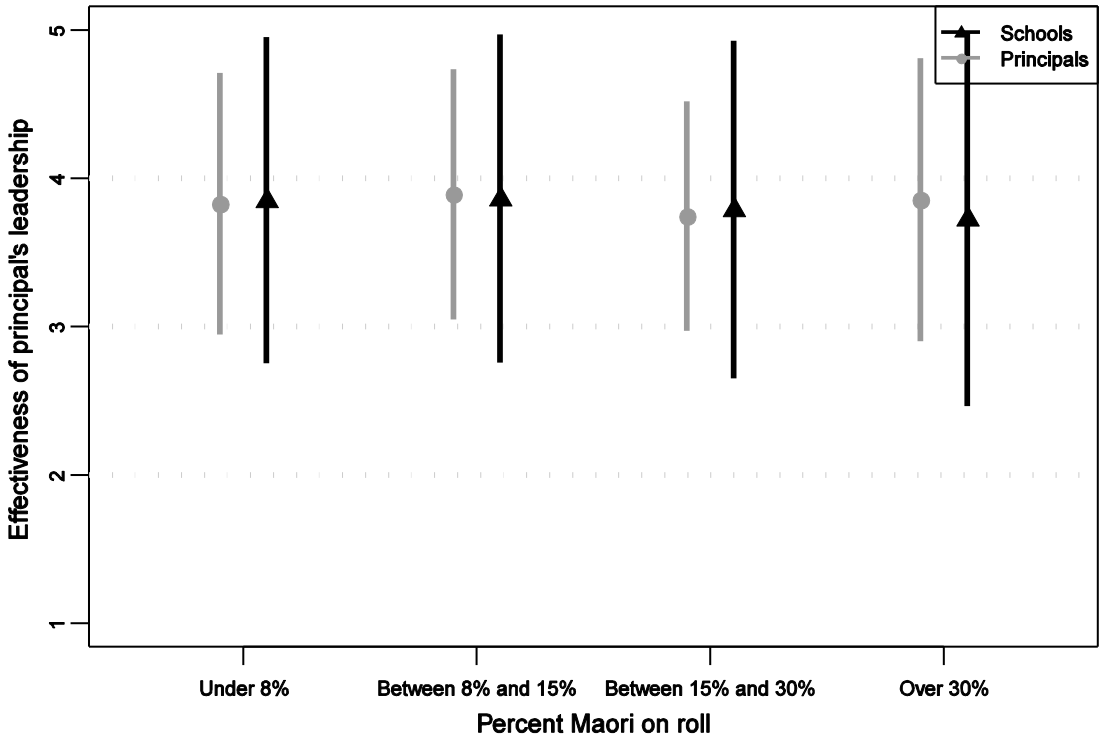
Figure 61 **Principal Leadership according to school decile**



Percentage of Māori students on the school roll

Principal and school ratings of Principal Leadership effectiveness showed no clear patterns of difference in relation to percentage of Māori students on the school roll (see Figure 62 below).

Figure 62 **Principal Leadership according to percentage of Māori on the school roll**



Principal Pedagogical Leadership Context scale

The ELP items focused on pedagogical leadership context answered by principals can be categorised as either support or barriers to their ability to exercise pedagogical leadership. In this section, to highlight the "schools like us" context, we have showed overall trends for support, and for barriers, according to school characteristics.

School type

In Figures 63 and 64 we see slightly lower ratings on the support items for principals of composite and secondary schools, as well as a slightly higher identification of barriers for these principals. The areas that showed the difference were in *teacher recruitment* and *retention*, *managing staff* and *attending to school paperwork for external agencie*s eroding time for *pedagogical leadership*, and having *affordable access to external expertise to increase student achievement. Manageable workloads* were also a challenge for these principals. Secondary principals gave lower ratings to their *performance appraisals giving them insight into strengthening leadership*.

Figure 63 **Support for pedagogical leadership according to school type**

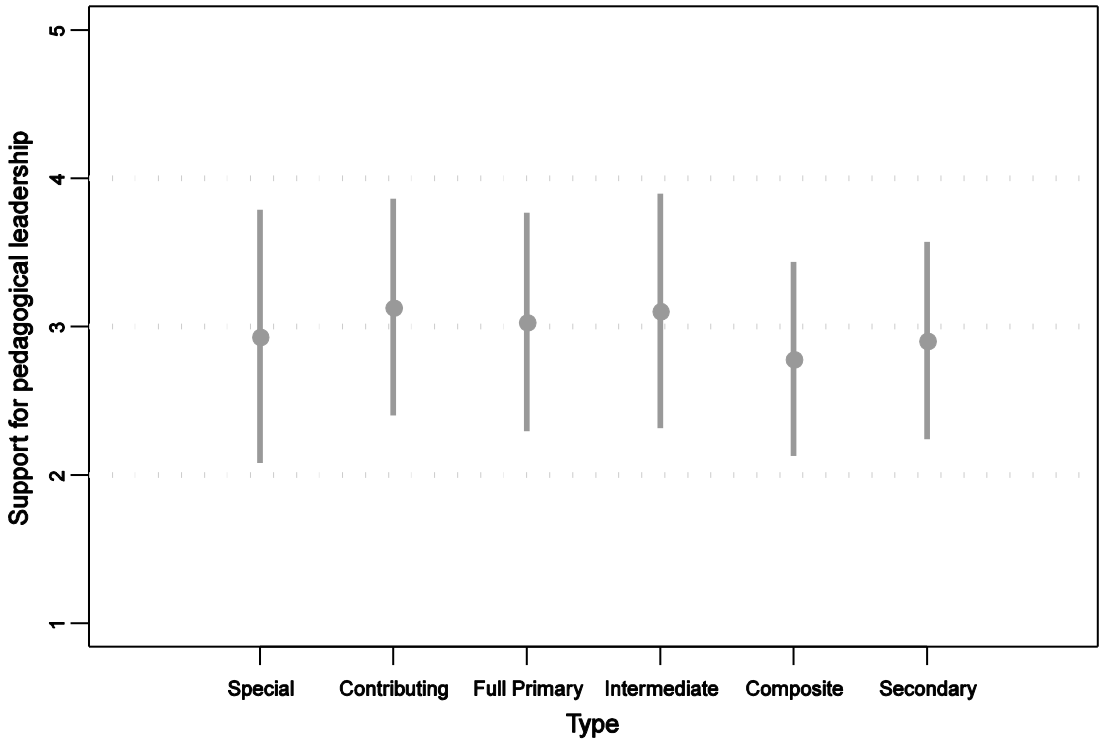
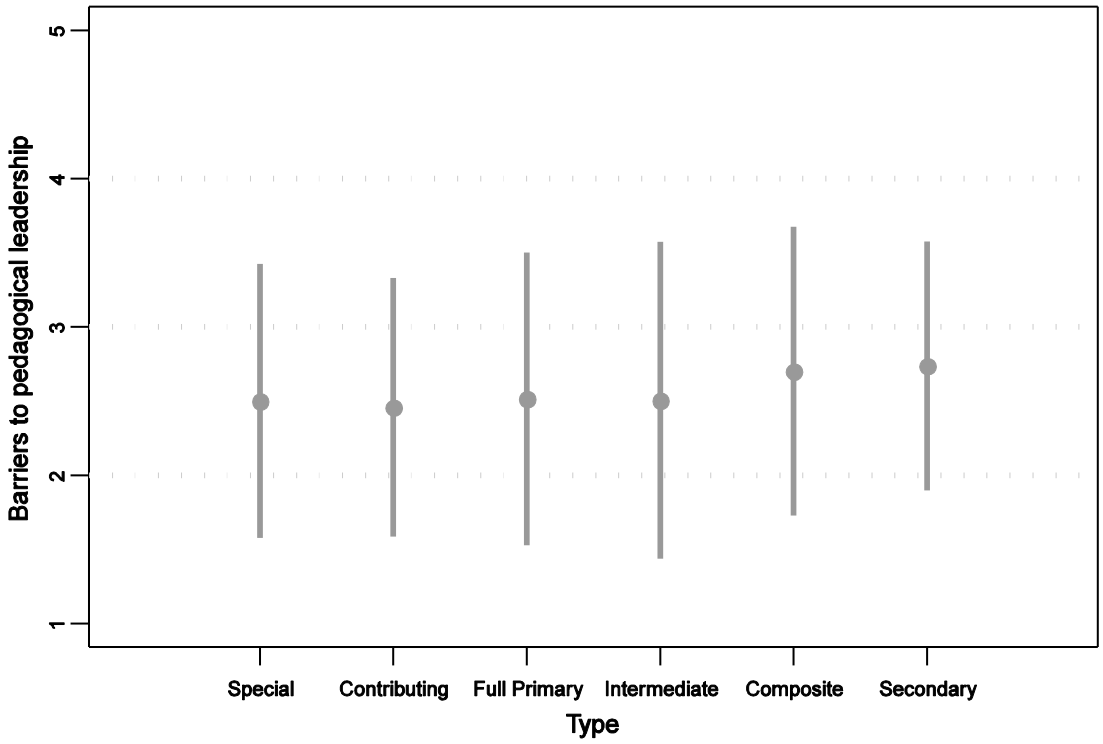


Figure 64 **Barriers to pedagogical leadership according to school type**



School size (U-grade) differences

School size was unrelated to ratings of support for pedagogical leadership. In relation to barriers, the differences were minimal, but principals in the largest schools were likely to agree with the statement that *managing staff significantly erodes time for pedagogical leadership* and that *ensuring the school paperwork meets external agency standards erodes time for pedagogical leadership.* Principals in the smallest schools were more likely to disagree that they had *access to data management expertise it needs to generate analyses of our student assessment data* (see Figures 65 and 66).

Figure 65 **Support for pedagogical leadership according to school size**

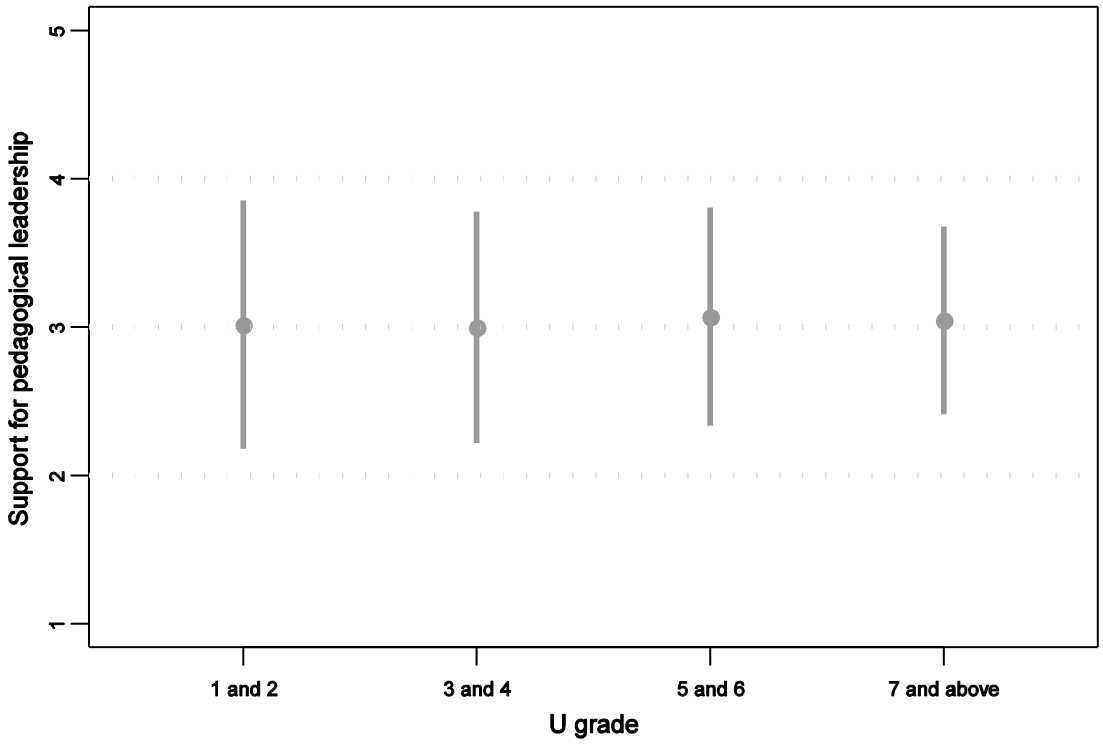
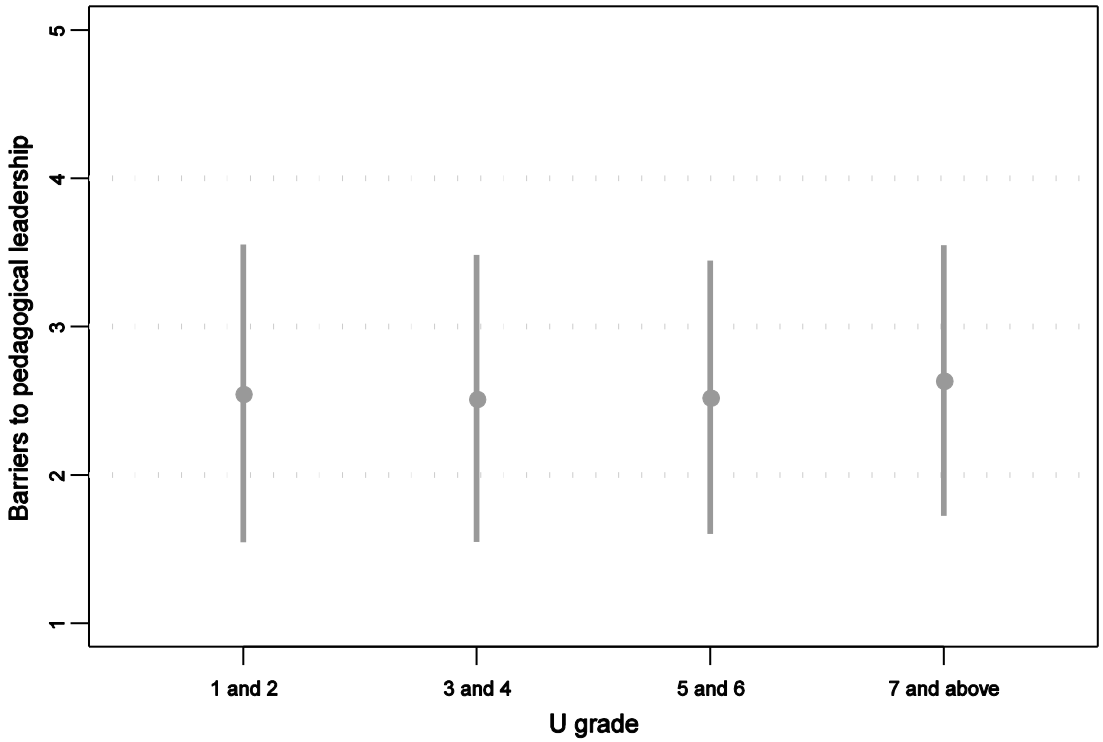


Figure 66 **Barriers to pedagogical leadership according to school size**



Decile differences

While ratings of supports for pedagogical leadership were similar across deciles, there were slightly less supports experienced by principals in decile 1 schools, and slightly less barriers in high-decile schools. There were three aspects where principals of deciles 1 and 2 schools felt they experienced more challenge to their role as pedagogical leaders in their schools: *recruiting and retaining effective teachers; student welfare eroding time for pedagogical leadership*; and *the board bringing good understanding and insight about academic and social achievement of our students.*

Figure 67 **Support for pedagogical leadership according to school decile**

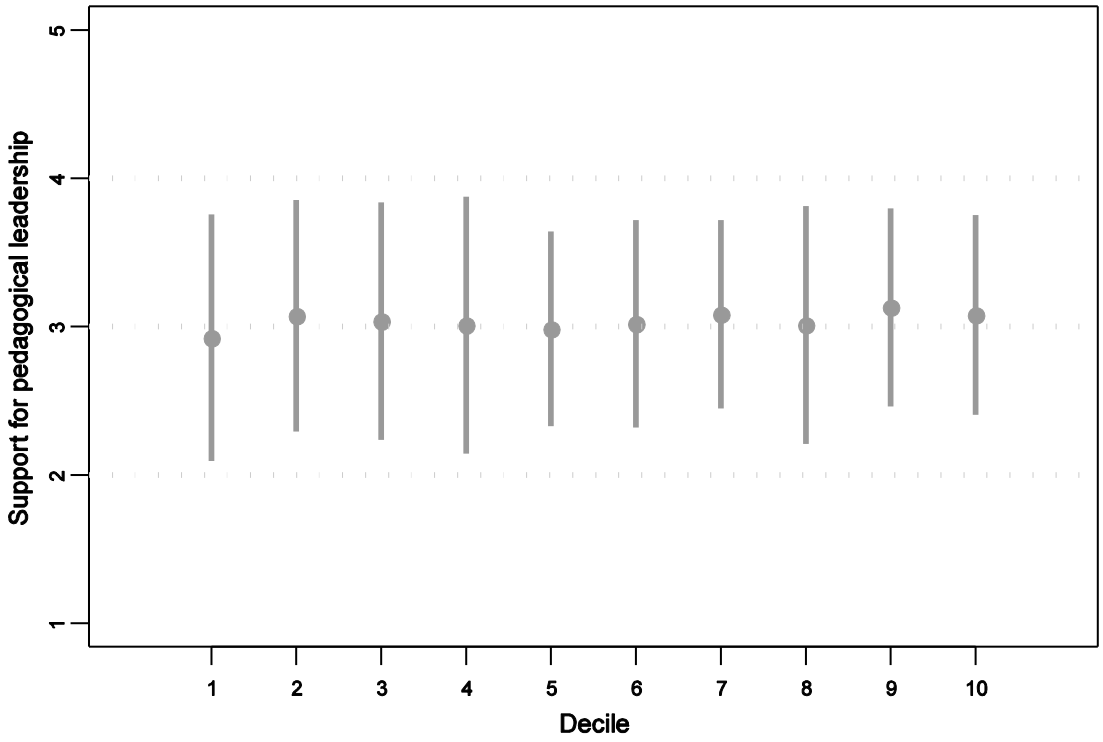
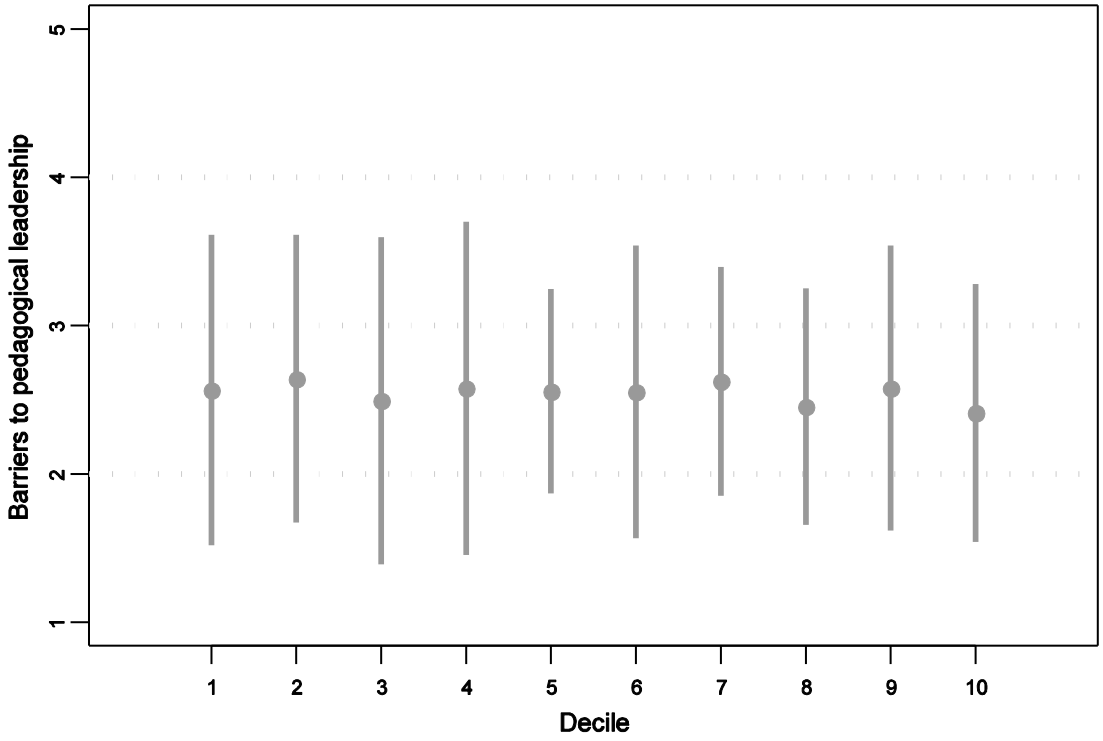


Figure 68 **Barriers to pedagogical leadership according to school decile**



Percentage of Māori students on the school roll

No differences related to the proportion of Māori students on the roll were evident in terms of support and barriers for pedagogical leadership. However, two items showed significant differences in principals' responses. The higher the percentage of Māori on the school roll, the more principals disagreed with the statement *the school has no difficulty recruiting and retaining effective teachers* and the more they agreed with the statement that *student welfare eroded time available for pedagogical leadership*.

Figure 69 **Support for pedagogical leadership according to percentage of Māori on the school roll**

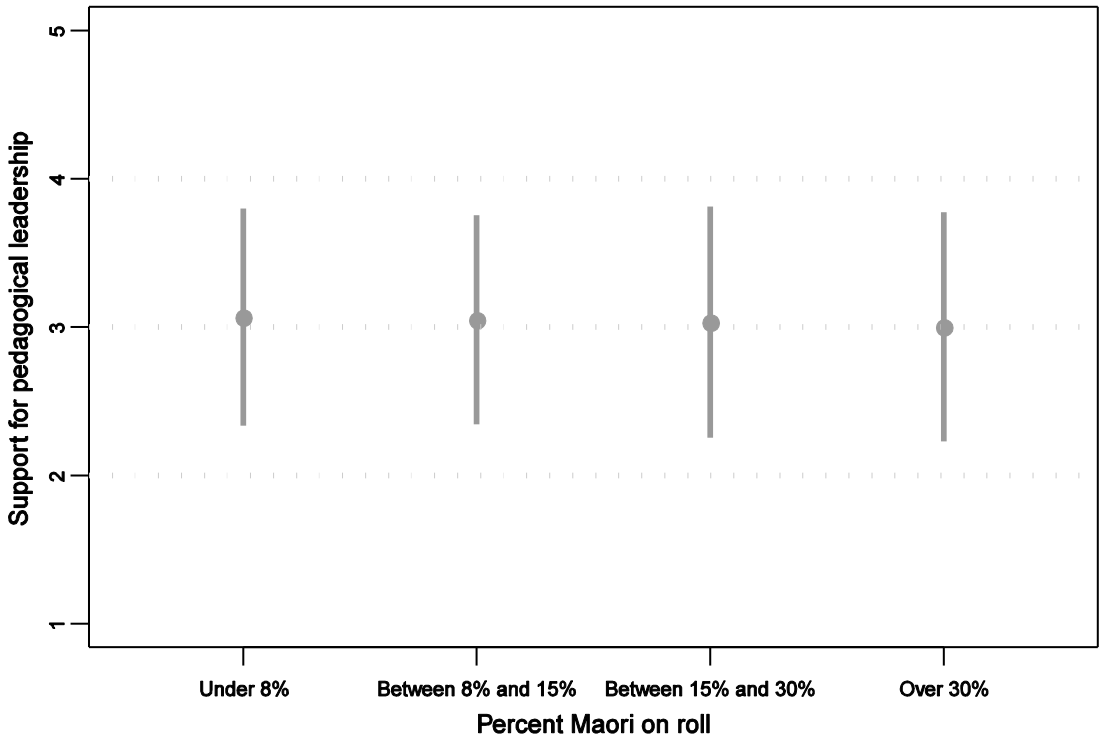
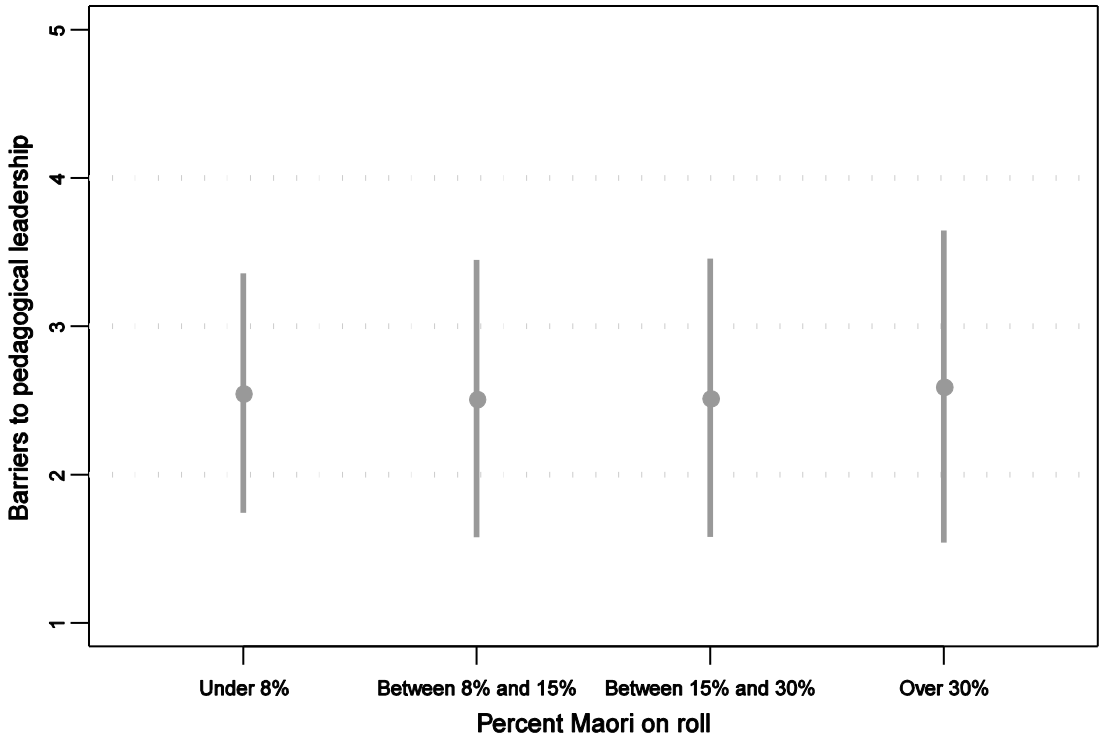


Figure 70 **Barriers to pedagogical leadership according to the percentage of Māori on the school roll**



Summary of patterns across the scales in relation to school characteristics

School type

Overall, there are about 10 scaled score point differences between the means for primary and secondary schools across the seven scales that link together. Intermediate and composite schools sit about halfway between primary and secondary schools. For the Ensuring a Safe and Orderly Environment scale, intermediate and composite school educational leadership means are closer to primary school leadership means. On the Principal Leadership scale, special schools and secondary schools had the lowest means. Differences were much less marked for the Māori Success scale.

School size

Again, the spread of scale score means is about 10 scaled score points—between U1 and U2, and U7 plus schools, with U3 to U6 schools sitting midway between the means. The pattern of mean averages is repeated across the seven scales that link together. Although a different metric was

used, the same pattern of means across school size was also repeated with Principal Leadership. Differences were marked (with higher means) at U1 and U2 schools for the Māori Success scale.

School decile

Differences are a little smaller when analysed according to decile—around 5 to 6 scale score points difference between the means of low-decile schools and high-decile schools. Means appear clumped in three groups: low-, middle and high-decile schools. There were only slight patterns of differences for school decile on the Principal Leadership scale. Deciles 1 and 2 schools had slightly higher mean scores on the Māori Success scale.

Percentage of Māori on the school roll

Differences are smaller again here; around 5 to 6 scaled score points difference between the means of schools with less than 8 percent Māori students on the roll, and schools with more than 30 percent of Māori students on the roll. Means appeared to fall into two groupings: less than 15 percent Māori on the school roll, and more than 15 percent Māori on the school roll. There were no clear patterns of difference for Principal Leadership according to percentage of Māori on the school roll. Schools with over 30 percent Māori on the school roll had higher means in the Māori Success scale.

4. What factors account most for differences in school leadership scores?

To see which factors seem to account most for differences between schools in their school leadership scores, we used multilevel models. These multilevel models are not based on the average teacher scores we have used so far, but use individual teachers' actual scores, "nested" within their school. This means we can use the full range of scores in a school. The statistical software we used for this modelling was MLwiN version 2.23. In a series of exploratory models, we included the school characteristics, teacher variables and principal variables included in the ELP survey to see which of these appeared to account for some of the difference (variability) in school scores. This exploratory work enabled us to identify a set of variables that appeared to account for some of the differences in school scores, that we could include in the models reported here. We tested two-level models (teachers nested in schools). The length of principal experience (either in total as a principal, and as principal of their current school) and previous experience as a teacher and as a senior manager did not show statistically significant associations with the overall leadership scale score, and so were not included in the models.

We focused on just the overall leadership scale (as we did in Wylie & Hodgen, 2010).

We start with the null model, which simply includes the range of school leadership scores for each school, and estimates the proportion of the variance between school scores that remains unaccounted for. We needed first to check the assumptions underlying our model.

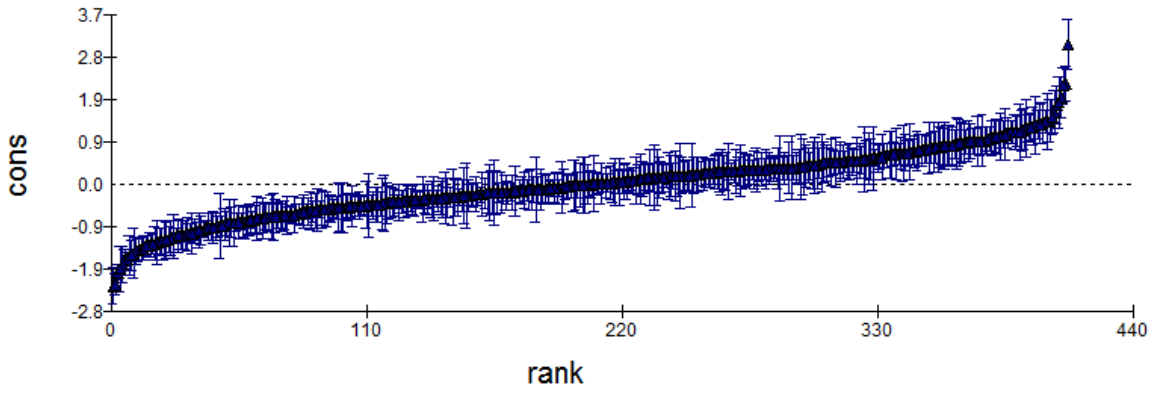
A common way to check that the assumptions underlying the models are met is to calculate and then plot "residuals", or the difference between the measured value and what the model would have predicted. Plots of residuals at both levels (teacher and school)[4] showed some departures from what would be expected, and a few problems with one or more data points with too much "leverage" (or influence on the model, in the sense that the fitted values would change markedly if the individual teacher or school were excluded from the model). This was unlike the situation in the previously fitted model. Findings from the new models should be interpreted with some care, and some differences between the parameter estimates presented in this report and those presented in Wylie and Hodgen (2010) may be at least in part attributable to these issues. On the whole, the parameter estimates presented in this report and the 2010 report are quite similar.

[4] At the teacher level, the residual is the difference between the scale score and the score predicted or estimated by the model. At the school level, the residual is based on the mean of the residuals for the teachers in the school, but takes into account the structure (clustering of teachers within schools) of the data.

A commonly used plot of residuals to get an image of the differences between schools is the "caterpillar plot" where the residuals are arranged in ascending order, and then plotted with 95 percent confidence intervals. In Figure 71 the residuals for the overall ELP scale with the null model fitted are shown. The horizontal dotted line through 0 indicates a perfect match between the scale score for the school and the predicted value. The schools represented on the left of the graph, where the residual is below 0, are those where the observed value was lower than the predicted value (the approximately 100 schools that had less effective leadership), and those represented on the right of the graph (residuals above 0) are those where the observed value was higher than the predicted value (the approximately 100 schools that had more effective leadership). Those significantly below or above 0 have confidence intervals that do not cross the horizontal dotted line.

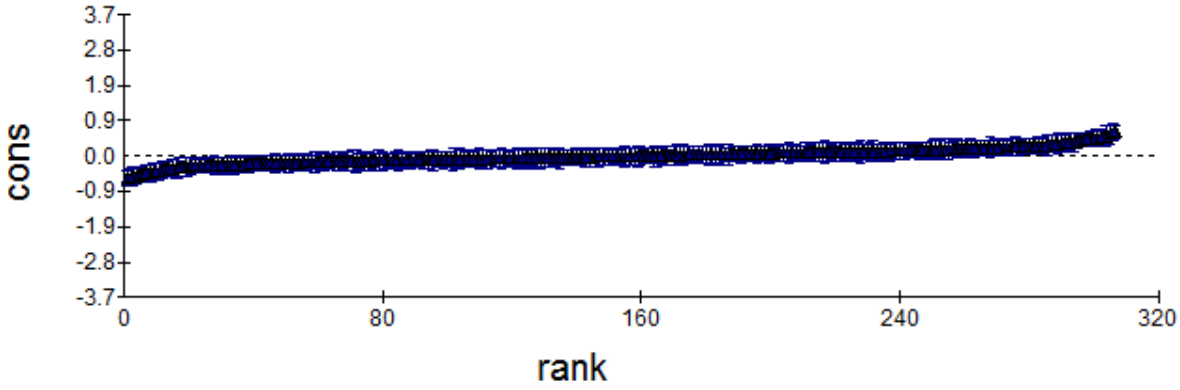
If we take no school or individual teacher information into account, around 100 schools (25 percent) had scores below the average, and 100 schools (25 percent) scores above the average predicted by the model. The extremes of the plot show relatively steep increases in residual between adjacently ranked schools, compared to the schools in the middle of the plot, where there is very little difference between adjacently ranked schools. This suggests that, taking a superficial look at the results, there are marked differences in leadership scale scores between some schools at either end of the spectrum. This model leaves 31 percent of the variance between schools unaccounted for.

Figure 71 **School-level residuals for the overall school leadership scale, null model**



After we added relevant school characteristics—teacher characteristics, teacher morale, workplace experience and view of their principal's leadership, and the principal's views of their context for pedagogical leadership to the model—the residual plot for the same scale shows much less variability between schools (Figure 72). In this plot there are around 40 schools in each of the extremes, doing better or worse than expected, and the slope in the two tails is much less marked. The height of the error bars is also decreased, as the information used in the model is able to explain quite a lot of the variation in scale score between individual teachers. The proportion of variance between schools left unaccounted for has now gone down to 8.6 percent.

Figure 72 **School-level residuals for the overall school leadership scale, full model**



What these graphs and the underlying models tell us is that differences in school leadership among the schools do exist. These differences are on a continuum, with only around 10 percent of the schools in this national sample showing very different scores from most (either very low or very high) once we have taken relevant factors into account.

Do some factors carry more "weight" than others in accounting for differences between schools?

Table 5 shows how the factors that showed associations in the exploratory models appear to be related to differences between the schools in this national sample, in terms of changing the proportion of variance left unaccounted for (at the bottom of the table), and in terms of the difference they make to the mean score for the reference group in each model. Statistically significant differences are given in bold; and the standard error of each change in score is given alongside.[5]

These reference groups are the groups that showed the lowest average scores. In the teacher characteristics model, it is for a Pākehā teacher with more than 15 years' experience. In the teacher workplace model, it is a teacher with low morale, who strongly disagrees they have a positive workplace experience, and who does not think the principal's leadership is effective. In the school characteristics model, it is a deciles 1–2 school, and a secondary school. In the principal leadership context model, the reference category is of teachers in a school where the principal strongly disagrees that they have sufficient time for the pedagogical leadership part of their job and who disagrees that they can recruit and retain effective teachers, and can access

[5] The models reported here have been kept as simple as possible. When developing these models we initially added several other variables and later removed them as they did not add significantly to the model. These variables include: length of time the teacher has taught under the principal; gender of teacher; whether they are part-time or full-time; permanent or relieving; their highest (most responsible) role in the school; the size of the school; whether it is in a rural area or not; the principal's previous experience; gender of principal; and age of principal.

advice, support and expertise needed. In the full model, it is for a teacher having all the characteristics described for each of the separate models.

The teacher model actually increases the proportion of unexplained variance to 32.0. This is to be expected, as the model reduces variance at the teacher level, with the result that the *proportion* of variance at the school level is increased. It shows that the overall school leadership score is likely to be higher if judged by a probationary teacher (with less than 2 years' experience), or by Pasifika, Māori or Asian teachers.[6]

The teacher workplace model decreases the unaccounted variance between schools from 31 to 21 percent, and shows that overall school leadership scores were indeed higher if judged by teachers who have good morale, and positive or not strongly negative workplace experiences; and the scores were higher again if judged by teachers who rated their principal's individual leadership as outstandingly effective.

The school characteristics model decreases the proportion of unaccounted variance between school leadership scores from 31 percent to 23 percent, showing increases in overall scale score related to increases in school decile (in two "lumps" rather than a steady rise), and for primary teachers.

The context for the pedagogical leadership model drops the between-school variance left unaccounted for from 31 to 27 percent; scores decrease for those in schools whose principals had high scores for their perceived barriers to pedagogical leadership, and increase (but not statistically significantly) for those in schools whose principals thought they had high levels of support.

The full model brings all these together, and the overlap between some of the variables means that the size of their contribution may change, and some are no longer significant. Teacher views of the principal's individual leadership, school decile, school type, teacher morale and workplace experience, and teacher ethnicity remain significant. The principal's leadership context variables are no longer statistically significant. Just under 11 percent of the variability between schools is not accounted for by the full model.

A school that is most likely to have a higher-than-average overall scale score is a decile 5 or higher primary school with a principal who is judged to be a very effective leader, and one in which teacher morale is high, and teachers perceive they have a positive workplace experience.

---

[6] Perhaps new teachers are most positive because they have less to compare with; we have some caveats around the ethnic differences, which may simply reflect differences in the size of these groups, cf. the majority Pākehā, or differences in assumptions about school practices (we do not know if there are in fact such differences).

Table 5 **Multilevel models of relevant factors accounting for differences in schools' overall school leadership score in this national sample**

Variables		Null model	Teacher model	Teacher work-place model	School characteristics model	Principal pedagogical context model	Full model
Reference category mean		58.5(05)	58.9(0.5)	41.9(0.5)	63.0(0.9)	58.7(3.4)	39.0(2.0)
Years' teaching experience	0–2 years		**1.6(0.6)**				-0.3(0.5)
	3–5 years		**-2.2(0.5)**				-0.8(0.5)
	6–10 years		**-1.7(0.5)**				-0.7(0.4)
	11–15 years		**-2.0(0.5)**				**-1.3(0.5)**
Ethnicity	Māori		**1.3(0.6)**				**1.7(0.6)**
	Pasifika		**4.8(0.9)**				**3.2(0.8)**
	Asian		**3.8(0.8)**				**3.3(0.8)**
	Other		-0.1(1.2)				1.4(1.1)
Morale is good	Agree			**2.8(0.5)**			**2.8(0.6)**
	Strongly agree			**3.7(0.6)**			**4.1(0.7)**
Positive workplace experience	Disagree			**1.9(0.5)**			**1.4(0.6)**
	Agree			**2.8(0.4)**			**2.2(0.5)**
	Strongly agree			**6.3(0.5)**			**5.4(0.6)**
Effectiveness of principal's leadership	Satisfactorily effective			**6.4(0.4)**			**6.1(0.5)**
	Highly effective			**11.3(0.4)**			**11.0(0.5)**
	Outstandingly effective			**20.4(0.5)**			**20.2(0.5)**
*School decile	3 & 4				**2.3(1.3)**		**1.4(0.9)**
	5 & 6				**3.7(1.3)**		**3.8(0.9)**
	7 & 8				**5.3(1.3)**		**4.1(0.9)**
	9 & 10				**4.9(1.3)**		**3.8(0.9)**
*School type	Primary				**10.3(1.0)**		**6.4(0.7)**
	Intermediate				**5.8(1.5)**		**3.1(1.0)**
	Composite				4.0(2.1)		0.7(1.3)
	Other				**9.1(2.2)**		**6.7(1.4)**
Barriers to pedagogical leadership	Disagree					-1.2(2.3)	-1.6(1.2)
	Agree					-2.4(2.4)	-2.1(1.2)
	Strongly agree					**-5.9(2.9)**	-3.2(1.5)
Support for pedagogical leadership	Agree					1.9(2.6)	-2.1(1.4)
	Strongly agree					4.9(2.7)	-0.5(1.5)
PVC (%)		30.7	32.0	20.9	22.7	26.9	10.6
n		6,348	5,817	5,829	6,348	4,690	4,282

* School-level variables. Statistically significant differences from the reference category are given in bold.

This model of the variables that account for a reasonable proportion of the variance in overall school leadership scores does not provide a causal account—and some of the variables would not operate in a unidirectional fashion. For example, teachers' view of their principal's individual leadership may colour their views of the wider school leadership practices—but those views of practices may also enter into their judgements of their principal. Morale and positive workplaces are likely to be the result of good school leadership practices—but they also provide a fertile ground in which to build and sustain those practices.

The variables in this final model do provide some evidence of contextual factors that appear to have a bearing on school perceptions of the quality of educational leadership practice—particularly school decile, type and, to a lesser extent, the support for pedagogical leadership (and barriers to its exercise). They also provide some indicators that the school leadership practices covered in the ELP have positive links with teacher morale, good workplace practices and judgements of principal quality.

5. Benchmarking

Introduction

The national sample provides a clear picture of patterns of school and principal views on the effectiveness of school leadership in relation to the occurrence of a range of educational practices across schools. The patterns reported for the overall scale and the individual scales of the ELP, and the patterns shown for different school contexts allow schools to compare their own ELP survey results with other schools. That comparison will see some schools feeling affirmed in what they are doing, and others challenged. Beyond comparison with national norms, there is an evaluative question to be asked: **How** effective does a school's educational leadership need to be to function effectively? Ideally, we would be able to shed light on this question by analysing ELP scores in relation to other information about a school, particularly levels of student engagement and achievement. Such studies do need to be carried out in future. In the meantime, we enlisted a group of experts to participate in a day's benchmarking workshop to provide an initial answer to the questions: How good is good enough? When is a school at a critical level in terms of the effectiveness of their school leadership; and when is a school doing so well that they could be considered as an exemplar school for other schools? How do we set levels on the ELP that give useful indications to schools?

The benchmarking workshop focused solely on the seven scales that make up the overall ELP leadership scale (Goal Setting, Strategic Resourcing, Curriculum Quality, Quality of Teaching, Promoting and Participating in Teacher Learning and Development, Ensuring a Safe and Orderly Environment and Ensuring Educationally Powerful Connections with Family, Whānau and Community). For this workshop we stayed with the seven scales that make up the overall leadership scale. Further work could also be done with appropriate experts around the Māori Success scale, and the Principal Leadership scale.

Methodology

In level setting, no one method is right or wrong. The task is always to select the most appropriate method for a given situation. Level setting (whatever the method) is a robust process that takes a group of subject matter experts through a series of systematic steps.

There are a number of well-used methods established in the measurement literature for item-centred level-setting exercises such as this (refer Brennan, 2006, Chapter 12). The Angoff, Ebel and the Bookmark methods are the most relevant examples. The Angoff method requires a panel of experts to make judgements about probabilities of certain responses being given to each item

by respondents with particular characteristics. With the ELP, this would be a very complex exercise, and therefore is not really suitable for our purposes.

The Ebel method requires judgements about item difficulty—which is irrelevant here as the items are already located on the ELP scale—and additionally requires a panel of subject matter experts to make judgements about how relevant, or how important, it is to give certain responses to each item. The latter aspect is a central consideration in the ELP benchmarking exercise.

Bookmarking requires items to be presented in order of difficulty. Subsequently, experts gradually move up the items and agree on cut-off points that indicate the next level has been reached. Although the ELP survey items are ranked on the ELP scale, making a bookmarking-type method look attractive, this particular exercise is complicated by Likert-type item responses. Responses are given on a 5-point scale which would increase the number of required judgements four-fold compared to dichotomous (yes/no) items. This plethora of scale locations would make a traditional bookmarking exercise unwieldy and difficult for experts to apply effectively.

We chose to combine some aspects of each of these three methods. Overall, we needed to achieve the following with the experts in the workshop:

- the experts becoming very familiar with the survey, and what it measures
- defining what it means to be "basic" or "sound", for example, and coming to a consensus about what these definitions mean in terms of the survey items and operationalising these definitions by determining the responses to survey items that would indicate a school has demonstrated they warrant the level or classification under discussion ("basic"/"sound" etc.).

Process followed in ELP benchmarking

We convened a panel of experts with collective expertise in school leadership and the research related to effective school leadership and effective school practices. A group approach promotes discussion, and the sharing of ideas, to gain a common understanding about the ELP scale overall, and about individual ELP items.

The group included:

- practising principals
- leaders working in professional development roles in educational leadership
- key Ministry of Education personnel.

Before coming together, the panel had been asked to individually read the items carefully, consider the relevance of each and think of examples of how each item would be visible in a school—that is, what does an item "look like"? Panel members also learned about the item scaling process in place (Rasch scaling).

The group worked through a series of tasks during the day to provide context and support for the final actual benchmarking exercise. (We ran a "practice" session the day before to ensure the tasks were appropriate and possible for the group.) Below we describe these tasks.

Task: To gain familiarity with and common understanding of items

Experts were organised into small groups (two or three) to discuss what each item means, what it is "getting at" and coming up with examples of how the items may be exemplified in a school context. Some of this work was done individually before the benchmarking day, but this was an opportunity to continue the familiarisation, challenge viewpoints and possibly reassess individual thinking.

Task: To gain a concept of "key" items

Some items may be regarded as critical items in that they ask about aspects of critical importance with respect to educational leadership. These items were noted. The noting of key items also helped in familiarisation with the ELP as a whole. It required panel members to think deeply about each item, what it meant and what aspect of leadership it was really asking about.

Task: To explore the notion of benchmarks on the ELP scale

How many benchmarks do we want? What sort of categories can we discriminate between? Can we tell the difference between a school that is doing fairly well in its leadership practices and one that is doing even better? It was agreed to try to identify four distinct categories, using "satisfactory" as a stem—yielding four levels of less than satisfactory, just satisfactory, better than satisfactory and outstanding. To disassociate from the actual ELP item ratings, four scale descriptors were generated to describe ranges on the ELP scale to be benchmarked:

- "Invisible"
- "Basic"
- "Sound"
- "Exemplary"

Task: To explore the concept of Minimal Acceptable Competence (MAC-Basic)

This concept (used in the Angoff and Ebel methods), describes a borderline location on the ELP scale where a school may be considered **just** good enough to be placed in a particular category. For example, how high does a school have to score on the ELP scale to be considered at a basic level? Or, alternatively, how **low** could a school score on the ELP scale and still be considered "basic" rather than "invisible"?

Task: Complete the ELP survey for borderline cases

After a general discussion the experts were reassigned to small groups and asked to consider and discuss the MAC concept with respect to each item. What response would a respondent have to give to each item to locate a school at this borderline point on the ELP scale? Experts were asked to discuss and challenge each other's viewpoints, but subsequently individually complete the ELP ratings for a school considered to be at the minimal level for the "basic" category. The idea of

"key" items came into these judgements. In general, those items seen as "key" would have warranted a higher rating than items asking about practices considered less important.

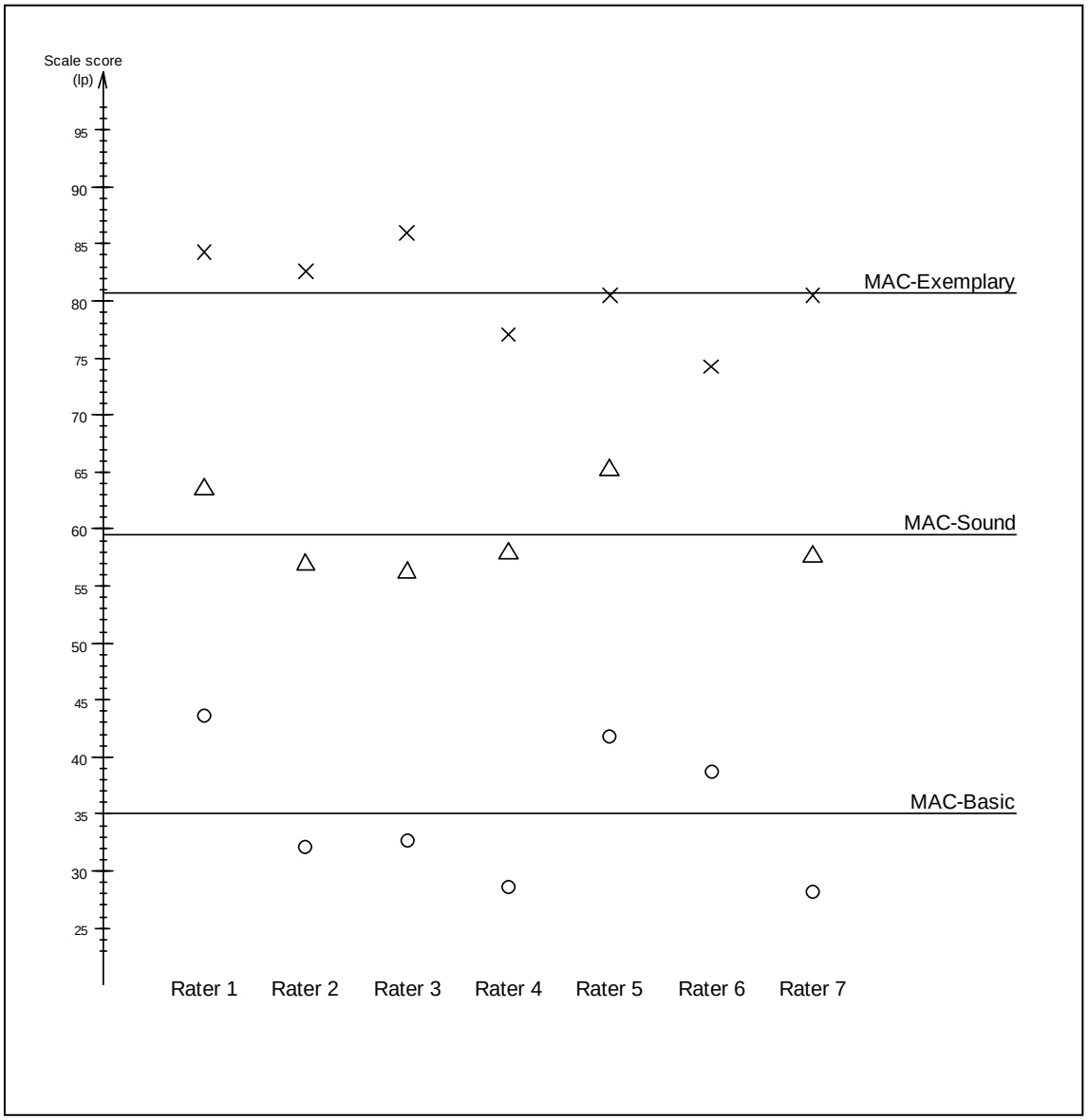
After filling in the ELP, the experts were encouraged to compare their completed forms and make any changes they wished. Consensus was not required.

Each expert filled in three survey forms, one for each of MAC-basic, MAC-sound and MAC-exemplary.

Creating the benchmarks

The experts' responses were collected and transformed into ELP scale scores to give a collective view on where each benchmark should be set. Figure 73 gives the experts' ratings for each of the levels. There does appear to be good separation between each of the benchmarks, which means that the panel of experts was able to distinguish well between schools in the different categories discussed at the benchmarking workshop. Figure 73 (p.105) also shows a consistent distinction between benchmarks for experts overall. This is shown by there being no overlap in expert views on where these benchmarks should lie, and individual ratings are relatively close to the benchmarks.

Figure 73 **Expert panel's views on where ELP benchmarks should be set** [7]



To what extent did the experts agree?

Inter-rater reliability is an important issue in benchmarking exercises such as this. Given we had seven experts and only five points on the rating scale for each item, a fairly crude measure using percentage agreement or near-agreement between raters on items is the best way to assess the inter-rater reliability. Correlation coefficients are very unreliable with small numbers. Tables 6, 7 and 8 show percentage agreement between raters for each of the benchmarks. "Agreement" is regularly gauged as being within one category of each other. Experts were not held to agreeing on

[7] Rater 6 had to leave the session early and therefore there is not a complete set of ratings for rater 6.

their ratings as they completed the survey forms. They were asked to discuss, share and challenge, but there was never a requirement for total consensus. In this way we were able to incorporate a variety of views and backgrounds into benchmark setting. It is interesting to note that, in the end, items were rated in very similar ways for the different MAC levels. So without forcing consensus we managed to reach a high level of rater consistency overall.

Table 6 **Rater agreement for MAC-basic benchmark**

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7
Rater 1	1.00	0.79	0.81	0.69	1.00	1.00	0.69
Rater 2	0.79	1.00	1.00	0.97	0.85	0.88	0.95
Rater 3	0.81	1.00	1.00	0.93	0.86	0.90	0.93
Rater 4	0.69	0.97	0.93	1.00	0.75	0.83	0.98
Rater 5	1.00	0.85	0.86	0.75	1.00	1.00	0.75
Rater 6	1.00	0.88	0.90	0.83	1.00	1.00	0.83
Rater 7	0.69	0.95	0.93	0.98	0.75	0.83	1.00

Table 7 **Rater agreement for MAC-sound benchmark**

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7
Rater 1	1.00	0.97	0.97	0.88	1.00	-	0.86
Rater 2	0.97	1.00	1.00	0.97	0.95	-	0.95
Rater 3	0.97	1.00	1.00	0.97	0.95	-	0.95
Rater 4	0.88	0.97	0.97	1.00	0.90	-	1.00
Rater 5	1.00	0.95	0.95	0.90	1.00	-	0.90
Rater 6	-	-	-	-	-	-	-
Rater 7	0.86	0.95	0.95	1.00	0.90	-	1.00

Table 8 **Rater agreement for MAC-exemplary benchmark**

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7
Rater 1	1.00	0.98	0.98	0.95	1.00	0.95	0.98
Rater 2	0.98	1.00	1.00	0.98	0.98	0.95	1.00
Rater 3	0.98	1.00	1.00	1.00	0.98	0.95	1.00
Rater 4	0.95	0.98	1.00	1.00	0.95	0.95	1.00
Rater 5	1.00	0.98	0.98	0.95	1.00	1.00	0.97
Rater 6	0.95	0.95	0.95	0.95	1.00	1.00	0.98
Rater 7	0.98	1.00	1.00	1.00	0.97	0.98	1.00

Results of the benchmarking exercise when compared with the national sample results

Overall comparisons with national sample

Figure 74 below shoes that benchmarking levels with the mean/standard deviation of the national sample marked as a grey block horizontally. It suggests that a good proportion of schools rate their educational leadership as at least “basic”, if not “sound”. Figures 75, 76, 77 and 78 describe the benchmarking results, using box plots, when compared with the national sample according to school characteristics. Secondary schools, larger schools and lower decile schools are more likely to be operating at the “basic” level according to our benchmarking exercise.

Figure 74 **Benchmarking results compared to means (and one standard deviation either side) for national sample results**

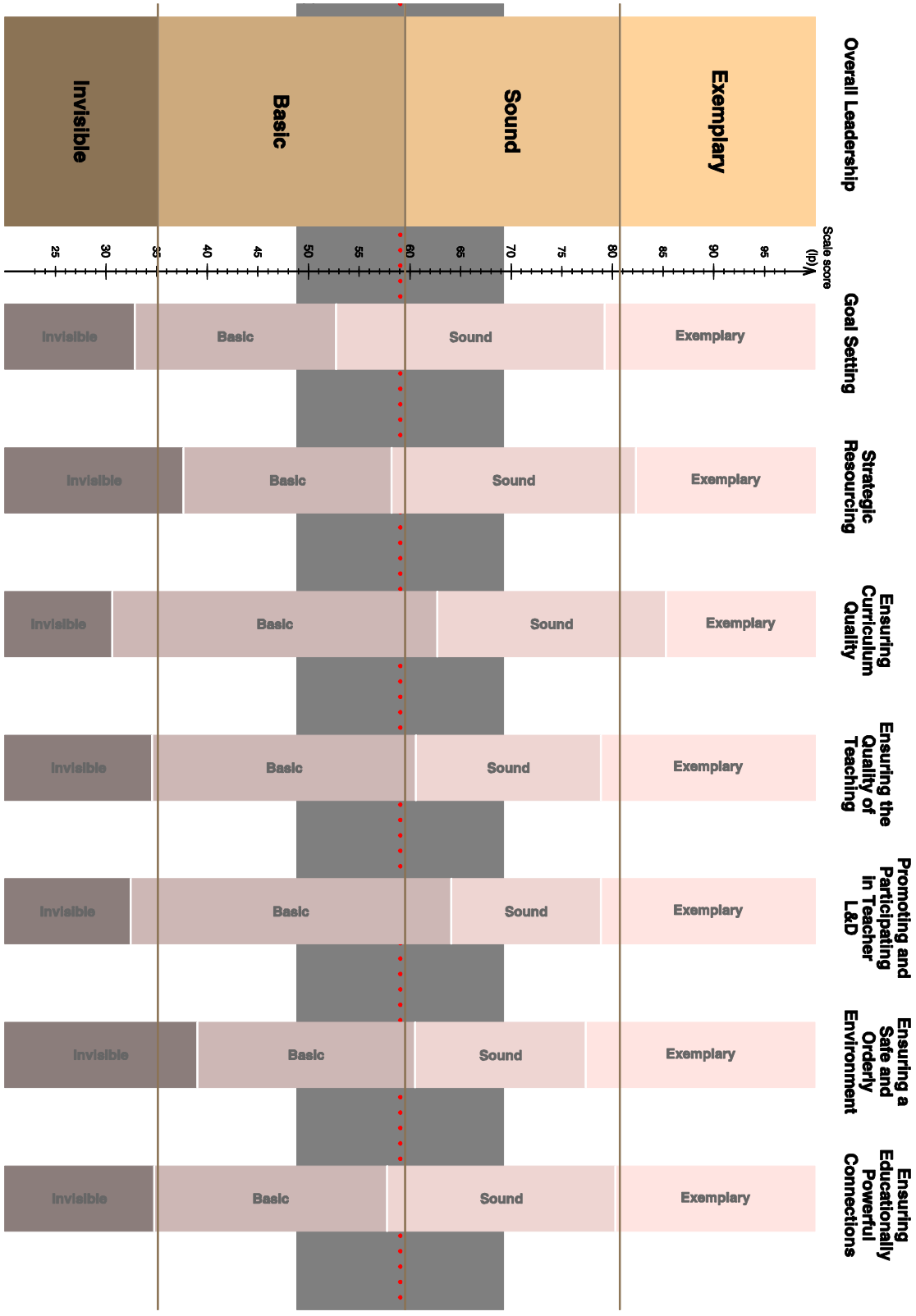


Figure 75 **Benchmarking results compared with national sample according to school type**

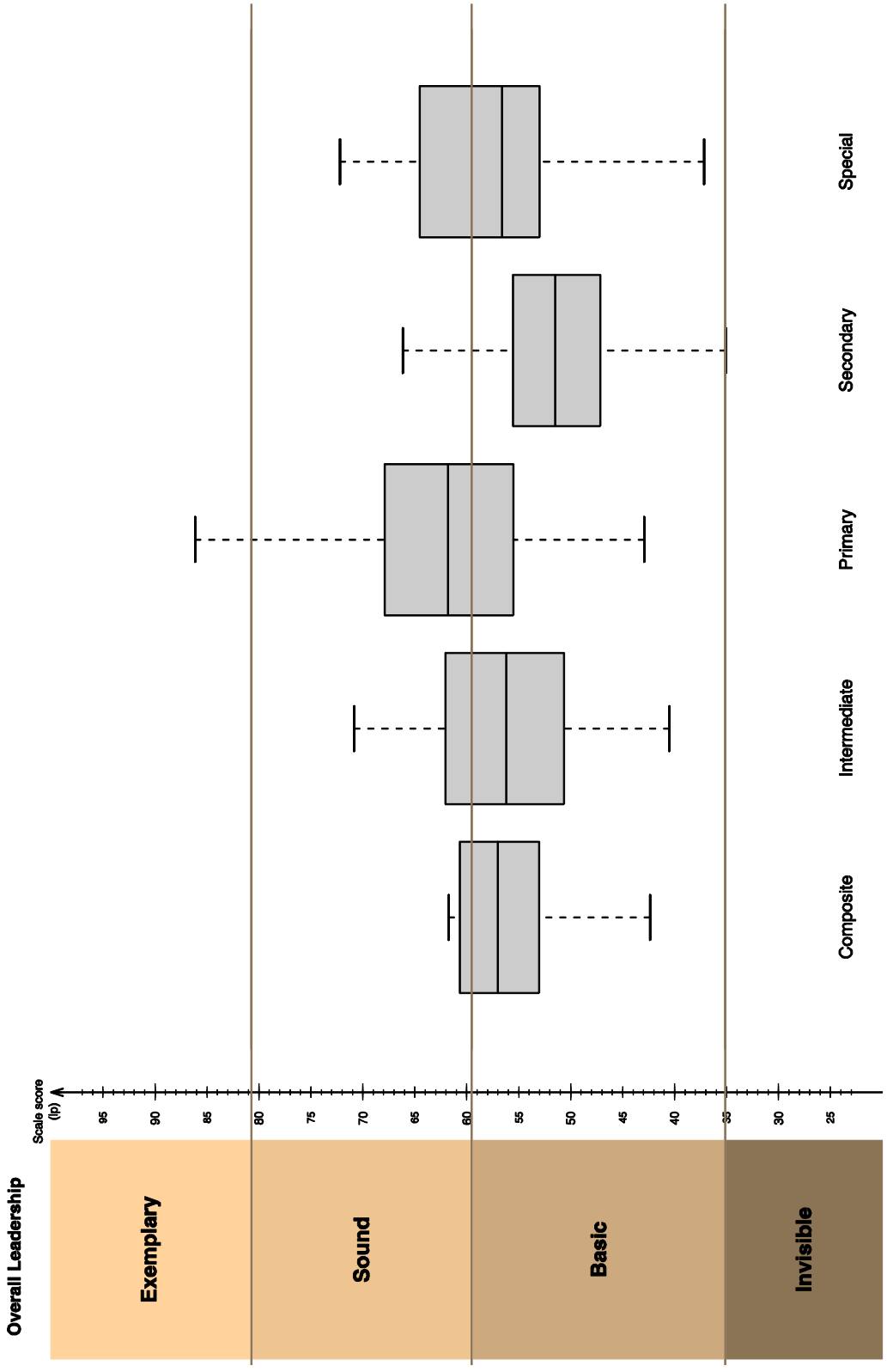


Figure 76 **Benchmarking results compared with national sample according to school size**

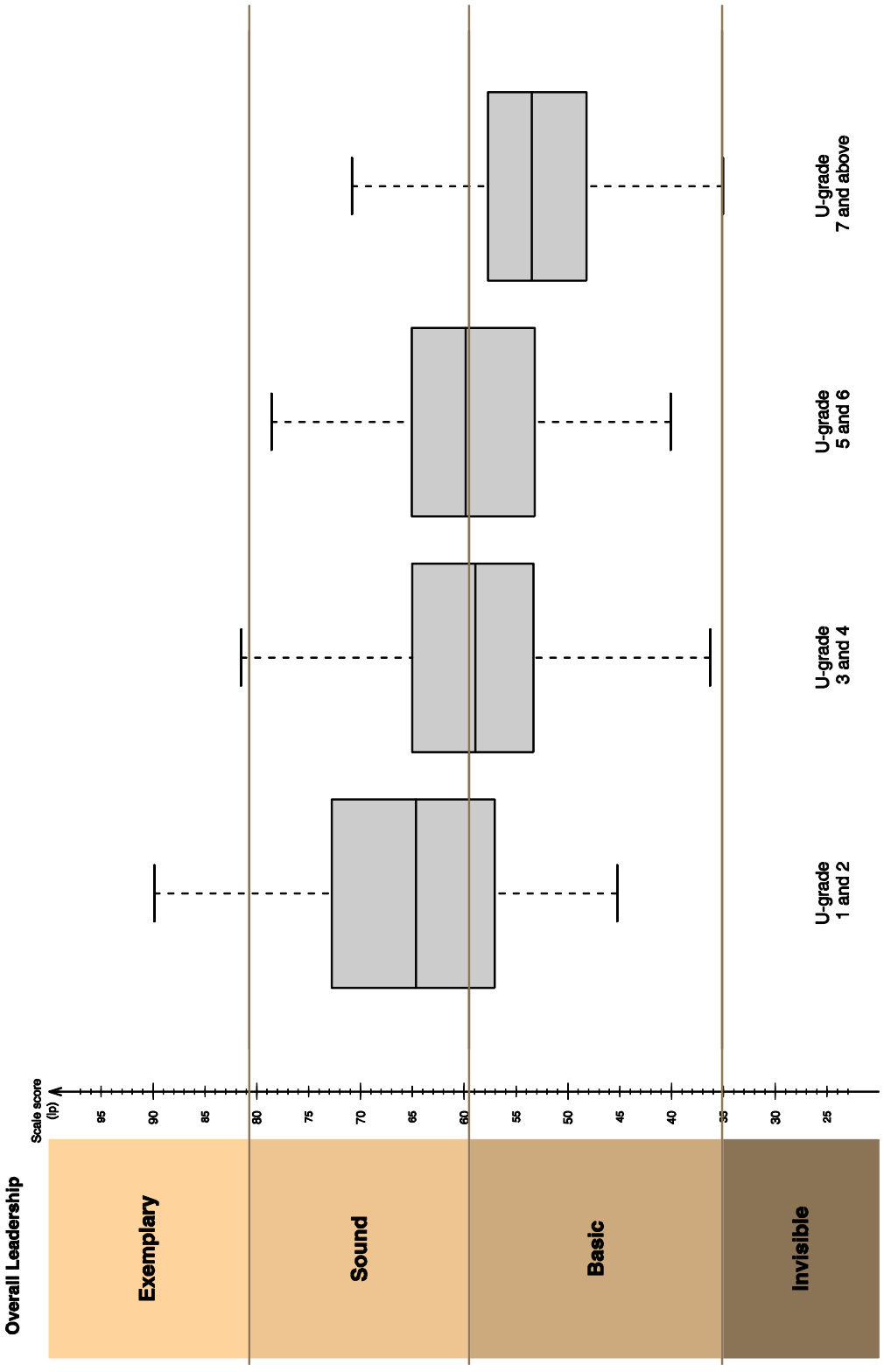


Figure 77 **Benchmarking results compared with national sample according to school decile**

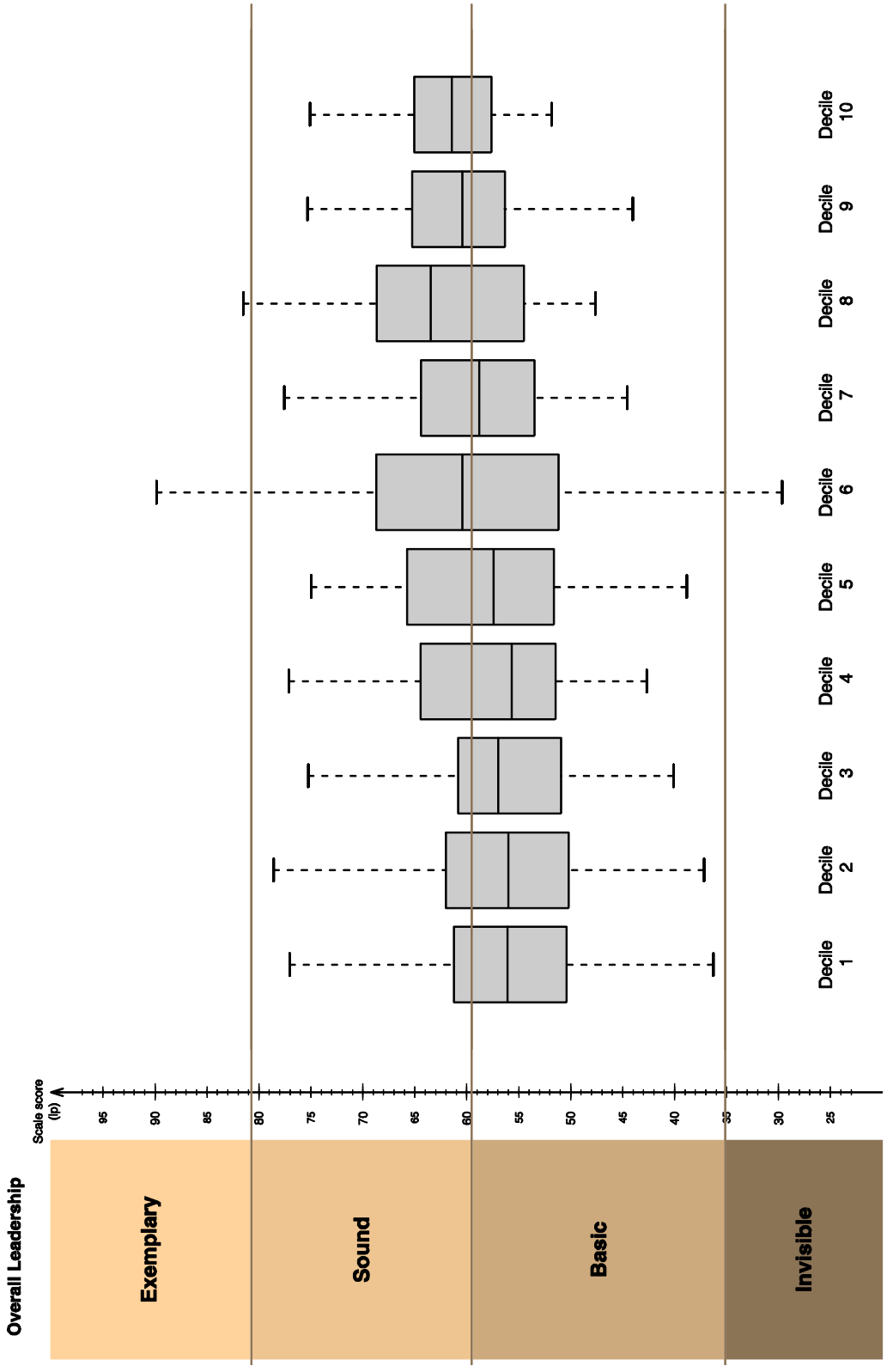
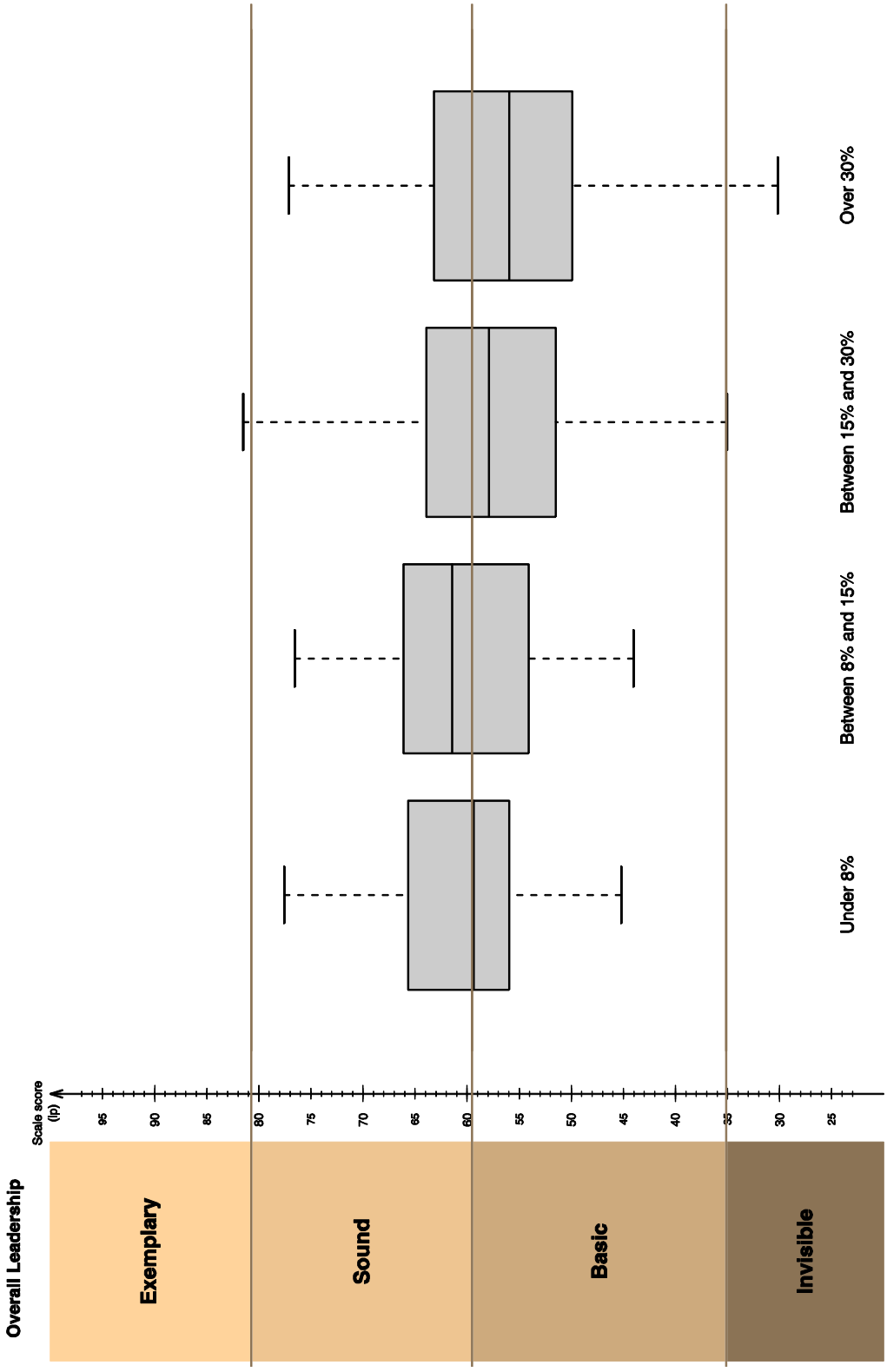


Figure 78 **Benchmarking results compared with national sample according to percentage of Māori on the school roll**



Concluding comments

It is our view that the benchmarking exercise was a robust process that yielded useful results. The outcomes, when compared with the ELP national sample, show good correlation between the national sample results and the benchmarking results. It suggests that most New Zealand school leadership is operating at basic or sound level, with a few at an exemplary level. It also suggests that there is plenty of room for "stretch" in school leadership, which was always an aspect of the ELP scale development. These benchmarking results would serve as a sound basis to begin professional development conversation about school scores on the ELP, and to develop an action plan for next steps.

6. Conclusions and recommendations

In this conclusion, we focus on the implications of the patterns found for this national sample of schools and the associated benchmarking exercise, in relation to the Ministry of Education's goal of developing strong educational leadership in every school. This section includes:

- a general discussion of the patterns and trends for schools overall and for particular school characteristics (type, size, decile and percentage of Māori on the school roll), relating to the national sample and benchmarking results
- some recommendations about the implications for next steps in professional learning and development in educational leadership, given the above discussion
- recommendations about the future role of the ELP in both supporting professional learning and development, and in providing data on the impact of professional learning and development at a system-wide level. These recommendations cover administration, use, analysis and scoring for the ELP.

Summary of overall main patterns relating to educational leadership levels as indicated by the ELP

There was a wide range of school scores on the overall leadership scale, but using our benchmarking approach, almost all schools were functioning at least at a basic level, with many at a sound and some at an "exemplary" level for some scales. There was a high level of inter-correlation between the scores on each separate aspect and the overall ELP score.

Goal Setting, Ensuring a Safe and Orderly Environment, and Principal Leadership were the scales that had the highest proportions of schools rating their school leadership as highly or outstandingly effective (50 percent or more). Promoting and Participating in Teacher Learning and Development and Māori Success were the scales that had the lowest proportions of schools rating their school leadership as highly or outstandingly effective (37 percent and 17 percent).

Some key trends in each scale:

Goal Setting: There appears to be more confidence about the role of leadership in relation to schools' guiding frameworks than about the embedding of the goals into ongoing use and evaluation—linking the goals back to individual teacher roles and individual student learning.

Strategic Resourcing: The schools gave highest ratings to the effectiveness of their school leadership in ensuring that the timetable reflected the school's priorities for teaching and learning, and lowest to items related to working with families and communities. In between come items related to teaching resource relevance and availability.

Curriculum Quality: School leadership was seen as most effective in ensuring the systematic monitoring of each student's progress and the existence of assessment plans to collect the information needed to monitor progress on priority learning goals, and least effective in ensuring that rigorous feedback was given to teachers about the quality of their schemes or unit plans, that all students experience challenging programmes and that all curriculum included content relevant to diverse learners.

Quality of Teaching: More than half the schools in this national sample thought that their school leadership was highly or outstandingly effective in ensuring that everyone shared responsibility for student learning (56 percent), that assessment data were used to improve teaching (61 percent) and that those teachers with particular expertise were used in the school to help other teachers' development (56 percent). The lowest rating item was students providing feedback to teachers on the effectiveness of their teaching, followed by support provided for teachers having difficulty helping students reach important academic and social goals, and challenge and support to improve teaching for teachers whose students remain disengaged.

Promoting and Participating in Teacher Learning and Development: Open discussion of student results and teachers helping each other develop more effective teaching strategies, serious discussions of how to improve teaching and learning in staff meetings and analysis and use of student achievement patterns to plan professional learning priorities were the items most likely to attract highly or outstandingly effective ratings of school leadership. Schools were much less likely to give such ratings to the provision of systematic opportunities to improve teaching through observing effective colleagues at work, teachers' use of a range of evidence sources to evaluate the effectiveness of their teaching, and adequate opportunities are provided for teachers to discuss why they might need to change their practice.

Ensuring a Safe and Orderly Environment: Most schools gave their school leadership highly or outstandingly effective ratings for ensuring they had positive environments for learning in which student learning is the central focus, and that the school was a positive environment for everyone, irrespective of culture. The gathering and use of student views in relation to school safety and culture were the two items on which the school leadership was least likely to be rated as highly or outstandingly effective.

Ensuring Educationally Powerful Connections with Family, Whānau and Community: Two-thirds of the schools thought their leadership was highly or outstandingly effective in ensuring that staff were responsive to families' views about their child's learning needs. Only about one-third of schools thought that their school leadership was highly or outstandingly effective in ensuring that parents understand the achievement levels of their children in relation to national benchmarks.

Māori Success: Schools were markedly less sanguine in this area. Schools were most likely to rate their school leadership as highly or outstandingly effective in relation to having clear school-wide targets for the academic achievement of Māori students (just over one-third of schools in the national sample), and least likely to rate them so for ensuring that there were professional

development opportunities that enabled teachers to develop the knowledge and skills needed to provide quality teaching to Māori learners (only around 15 percent of schools).

Principal Leadership: The top items in this scale were mostly related to the principal promoting the values of the school, having integrity and gaining and showing others respect; it also included making tough decisions when necessary. Identifying and resolving conflict quickly and fairly was the item with the lowest proportion of schools rating their principal as showing highly or outstandingly effective leadership.

Contexts for pedagogical leadership

Principals were asked questions about their pedagogical contexts. We divided the items in that set of questions into items related to "support", and items related to "barriers". There were slightly lower ratings for support factors in composite schools and secondary schools, and no other difference with respect to supports.

While supports seem for the most part similar across different school characteristics, there are some differences when looking at barriers. Overall, we do see some marked constraints experienced by a significant minority of principals in the ELP national sample in relation to being able to focus on the pedagogical leadership aspect of their role. These constraints are most evident in relation to time taken in staff management and student welfare issues, teacher recruitment and retention, school governance capability and expertise related to analysis of student achievement data and paperwork issues.

Time constraints and student issues were more likely to occur for principals at deciles 1–2 schools. Secondary principals and principals of larger schools were more likely than others to experience staff management as an erosion of their time for pedagogical leadership, and secondary principals were somewhat less likely to think their workload was manageable or sustainable.

Characteristics related to differences between school scores for educational leadership practices

High-scoring schools on the ELP scale are most likely to be primary schools, small schools, high-decile, with a Māori roll of less than 15 percent. These differences in school characteristics suggest that teacher views of school leadership effectiveness are likely to be lower where the school organisation is more complex—as it is in secondary and larger schools; or where the challenges of the student population are greater—as they are in deciles 1–2 schools, and in secondary schools.

Relatively higher scores on the Māori Success scale were likely to occur in small schools, lower decile schools and those with over 30 percent Māori enrolment.

Principal Leadership ratings showed less difference related to school characteristics than the other ELP scales. However, secondary and special school principal effectiveness had lower ratings, as did principals of larger and/or lower decile schools.

A note of caution re secondary school ratings

In noting the lower ratings for secondary schools, we want to remind the reader of the notes of caution in Robinson et al. (2009) and in associated commentaries on Robinson et al. (Notman, 2010; Youngs, 2011) about the paucity of statistical research evidence thus far available on secondary school leadership and its relation to student outcomes. We believe it is important that further construct validity research on the ELP at secondary level (i.e., looking at correlations between various aspects on the ELP and various measures of student achievement and engagement) is carried out.

Multilevel modelling showed that some variables do seem to account for much of the difference between schools in their overall school leadership scores. After accounting for these variables, only around 10 percent of the schools in the sample showed distinctly different scores (either very low or very high).

The variables that the multilevel modelling found to be associated with differences in school perceptions of the quality of educational leadership practice included contextual factors—particularly school decile, school type and, to a lesser extent, the support for pedagogical leadership (and barriers to its exercise). The modelling also provides some indicators that the school leadership practices covered in the ELP have positive links with teacher morale, good workplace practices and judgements of principal quality.

Benchmarking

The benchmarking exercise cast some very useful light on the question "How good is good enough?" The exercise results reflect good inter-rater reliability and some sound starting points for discussion with schools about where to next in their further development of educational leadership practices. The results provide schools or groups of schools with an opportunity to look at their ELP patterns, comparing their own results with the normative "schools like us" information and the benchmarking information.

Implications of ELP patterns in relation to the need for focused professional development and support for school leadership

The analysis shows that the length of principal experience, either in total or at their current school, is not associated with school scores for the effectiveness of either the principal leadership, or the school leadership as a whole. So the New Zealand education system cannot rely on the individual accumulation of experience to either maintain the current levels of educational leadership

practices, or develop them further. (The caution here is that the sample reported on in this report had few principals with 2 or less years of experience. Therefore this sample cannot be used to make recommendations about early induction programmes for new principals.)

Our benchmarking process conducted in this study provides some preliminary information on "How good is good enough?" and suggests most schools have a basic or sound level of educational leadership. This overall national picture provides grounds for both confidence, and some challenges. The current levels of educational leadership practices do indicate that there is certainly room to develop further, given that the existing research shows associations between most of the practices asked about on the ELP and student achievement.

The associations between ELP scores and levels of principal perception of support for their pedagogical leadership also raise the policy questions of ensuring that such support is available, such as continuing to address issues of teacher supply, and providing guidance for the most effective and affordable ways to raise student achievement.

Should professional development and support for educational leadership be targeted?

The associations with school size, decile and type with the ELP scores, and in relation to principal perceptions of support or barriers to their pedagogical leadership bring up the very real questions of factors beyond individual school control. They also pose real policy issues, given that there is little likelihood of ensuring that we have a more even social mix in our schools (a factor which certainly helps lift overall student performance of children from low-income homes, who are overrepresented among the low performers in education),[8] though we may be more ready now to start tackling the complex nature of secondary school organisation. Given this real constraint, if there is any need for prioritisation, deciles 1–2 schools and secondary and composite schools stand out.

Since there are no clear associations with principal years of experience, this is probably not a strong criterion to use in any prioritisation of professional development or support when it comes to the educational leadership component of the principal's role.

There appears to be most scope for further development in relation to the Promoting and Participating in Teacher Learning and Development scale, and Māori Success; and in terms of practices related to feedback on performance and effectiveness, providing timely challenge and support to both teachers and students, including student voice, and supporting parent understanding of student achievement.

---

[8] Causa, O., & Johansson, A. (2009). *Intergenerational social mobility*. Paris: OECD, Economics Department Working Papers No. 707; Field, S., Kuczera, M., & Pont, B. (2007). *No more failures. Ten steps to equity in education*. Paris: OECD; Webber, R., & Butler, T. (2006). *Classifying pupils by where they live: How well does this predict variations in their GSCE results?* London: Kings College, CASA Working Paper No. 99.

In relation to Māori achieving success as Māori at school we know that, while there is evidence of improvement in Māori achievement, there is still wide disparity evident. For example, the percentage of Māori school leavers with NCEA Level 2 or above was 35 percent in 2006, and rose to around 54 percent in 2010. In the same period, non-Māori figures rose from around 67 percent to about 78 percent (see Ministry of Education reports against Ka Hikitia goals at http://www.educationcounts.govt.nz/topics/31351/36805#figure4). Generally, there are lower ratings for ensuring effectiveness in relation to securing appropriate professional development related to Māori success at school. Also of concern are the lower ratings for this scale in higher decile schools. While there are more Māori students in lower decile schools, there are about 16,000 Māori students (almost 10 percent of Māori students in New Zealand schools) in higher decile schools where the ratings of ensuring effectiveness for Māori students' success are significantly lower. Of particular concern are the very low ratings around availability of good professional development in relation to Māori achieving success as Māori. Recent changes in the delivery of Ministry of Education-funded professional learning and development contracts have had this as a focus area; over time the ELP might be one way of monitoring outcome changes in this area, as well as enriching needs analyses at school level.

Robinson et al. (2009) describe the effect size for school leadership promotion and participation in teacher learning and development as being twice as large as any other effect size documented in their synthesis. Given this scale was one of the two with the lowest ratings on the ELP, it is important that this area is given further consideration. Additionally, as already noted, we need to determine if the research reflected in the ELP fits well within a model of effective secondary school leadership.

Implications for the ongoing development and use of the ELP

The ELP does provide a useful way of gauging and describing school leadership practices that are linked to teaching and learning. The original intention was to provide a tool with a range of items, including aspirational items that allowed "stretch", and the capacity to show change over time. The national sample results and the benchmarking results both show that the ELP is such a tool.

We cannot tell from the ELP levels alone whether they are high enough to make a real difference to student engagement and performance, or whether there is a minimal level that is necessary to ensure a given level of student engagement and performance. The benchmarking exercise in this report does help us in this latter regard. However, to further investigate the benchmarking results and the relationship of the ELP results to student achievement trends/changes, we would need to link patterns in ELP scores over time to patterns in student engagement and learning, and to the benchmarking results we have outlined in this report. A further national sample analysis will provide information on this.

ELP development

Evidence-based tools need to stay current with the evidence. It is anticipated that over time there might be a need for a revision of the tool to reflect the most recent evidence. One example raised in this report is in the area of effective leadership in the secondary school environment as it related to teacher learning and development. Some further investigation may be warranted in this area. We would suggest a revision of the ELP within the next 5 years (possibly the time of the next national sample) to ensure the tool remains reflective of best practice.

In previous reports we have referred to "sources of evidence" as one of the aspects of the results. In this part of the ELP, teachers and principals were asked to indicate the number of different sources they used to make their ratings. We have not used sources of evidence information in this report as our analyses show that it may have become a distracter to teachers completing the questionnaire (a possible respondent fatigue evident in some teachers' responses—this component was not completed towards the end of the survey items). We suggest removing this option from the ELP survey.

As already noted, there is still a need to look at the relation between ELP changes and changes in student achievement over time to ensure the ELP has strong concurrent validity.

Sharing the results of this national sample, and benchmarking exercise, with sector groups is an important step in continuing to collect data to confirm or refine the benchmarking exercise as well as an opportunity to provide current and potential users of the ELP with in-depth information about the tool and the associated information in this report. This in-depth knowledge has the potential to add to ELP interpretation and implementation of subsequent change at school/cluster level, and trigger opportunities for validity studies (in relation to a link with ELP scores and student achievement levels).

Marking services and data base analysis

There is currently a single marking service for the ELP which means all results can be readily aggregated. The availability of a single database with ELP scores provides useful information to national sector groups and to the Ministry of Education about trends over time, and trends in response to major policy initiatives (e.g., professional learning and development contract changes). In determining next steps with marking services the ready access possible with use of a single ELP database needs to be considered.

Recommendations

Given the conclusions and comments above, our recommendations are as follows:

Professional development and learning

- Consideration given to ensuring that educational leadership can access professional development opportunities that enable teachers to develop the knowledge and skills necessary to provide quality teaching for Māori learners.
- Consideration of targeted support for secondary and composite school educational leadership.
- Consideration of targeted support for educational leadership in the area of teacher learning and development.
- Consideration of support for principals and educational leadership generally in the area of identifying and resolving conflict quickly and fairly.
- Consideration of targeted support and policy for lower decile schools in the area of educational leadership in a way that reflects their current supports and barriers in relation to pedagogical leadership.
- Consideration of targeted support for educational leaders to have systems and routines in place to ensure the high-level planning implementation is actually reflected in day-to-day systems and processes within schools.
- Opportunities for professional development facilitators in the area of Leadership and Assessment to hear about and respond to the national sample results, and to the benchmarking exercise described in this report.

Where next with the ELP?

- Promulgation of use in a range of contexts (e.g. through Ministry-funded Leadership and Assessment contracts, and First Time Principal contracts).
- Retention of a single scoring system to maintain a single database capable of reporting on trends over time at a national level so that there is a mechanism for monitoring trends and major policy initiatives in this area.
- Consideration of use in conjunction with student achievement outcomes to ensure the ELP is a valid tool in relation to predictors of improved student outcomes (perhaps by matching NCEA patterns of achievement with ELP score patterns, using the “schools like us” models and benchmarking models provided in this report, and at primary level by perhaps including in the National Monitoring cycles in some way).
- Exploration of the ELP’s specific validity in relation to secondary school educational leadership.
- Ensuring the ELP continues to be reflective of current evidence by undertaking a revision within the next 5 years.

References

Brennan, R. L. (2006). *Educational measurement* (4th ed.). American Council on Education. Westport CT: Praeger Publishers.

Ministry of Education. (2007). *The New Zealand curriculum.* Wellington: Learning Media.

Ministry of Education (2008). *Kiwi Leadership for Principals: Principals as Educational Leaders*. Wellington: Ministry of Education.

Notman, R. (2010). Joining leadership and classroom pedagogies: A review of the best evidence synthesis iteration (BES). *Journal of Educational Leadership, Policy and Practice, 25*(1), 11–21.

Porter, A. C., Murphy, J., Goldring, E., Elliot, S. N., Polokoff, M. S., & May, H. (2008). *Vanderbilt assessment of leadership in education (Val-Ed)*. Nashville: Discovery Education.

Robinson, R., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why best evidence synthesis*. Wellington: Ministry of Education.

Wylie, C. (2010). Focusing leadership on adult learning: The secondary school challenge. *Journal of Educational Leadership, Policy and Practice*, *25*(1), 51–66.

Wylie, C., Brewerton, M., & Hodgen, E. (2011). *Shifts in Educational Leadership Practices Survey: Patterns in the Experienced Principals' Development programme 2009–10*.Wellington: Ministry of Education.

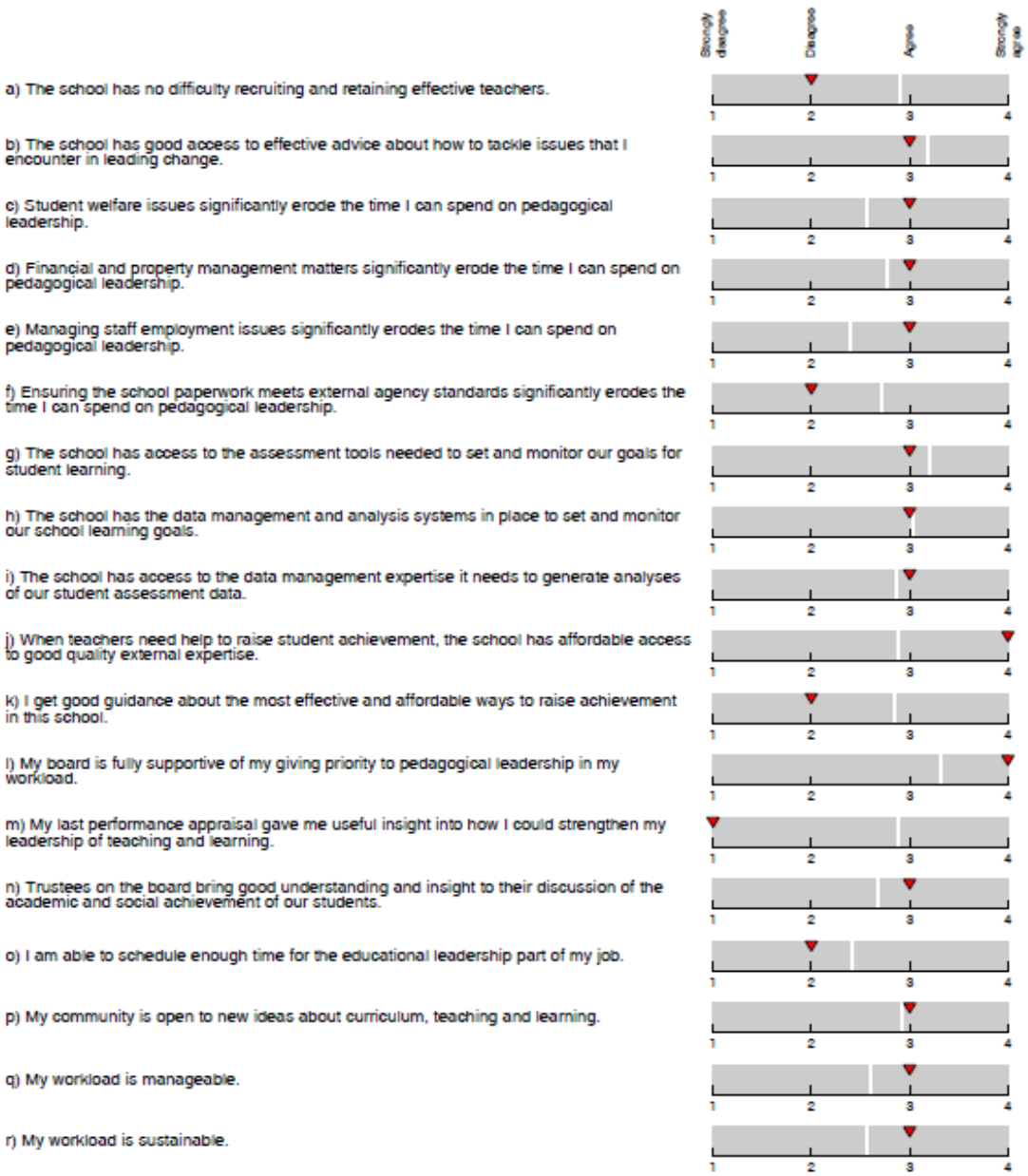
Wylie, C., & Hodgen, E. (2010). *Educational Leadership Practices Survey: Baseline 2009 overall profile of schools in the Experienced Principals' Development programme*. Wellington: Ministry of Education.

Youngs, H. (2011). The school leadership and student outcomes best evidence synthesis: Potential challenges for policy-makers, practitioners and researchers *Journal of Educational Leadership, Policy and Practice*, *26*(1), 16–27.

Appendix 1: Sample individual schools' reports

Educational Leadership Practices

My Pedagogical Leadership Context



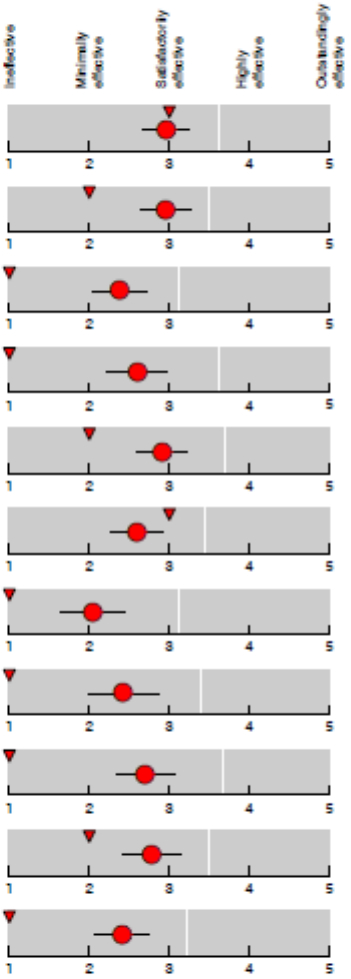
Educational Leadership Practices

Individual Item Report

1. Goal Setting

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view | Principal's view | National reference 2009

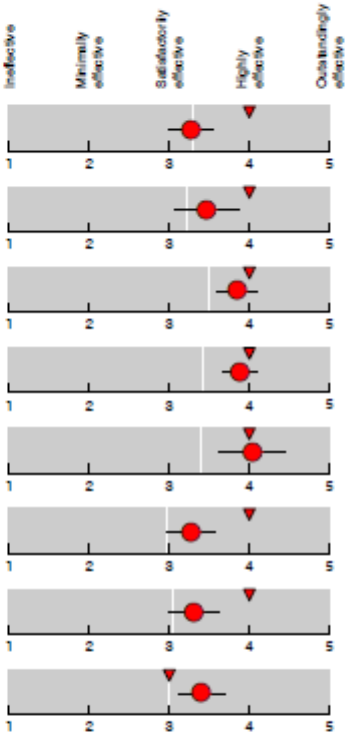
Educational Leadership Practices

Individual Item Report

2. Strategic Resourcing

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view | Principal's view | National reference 2009

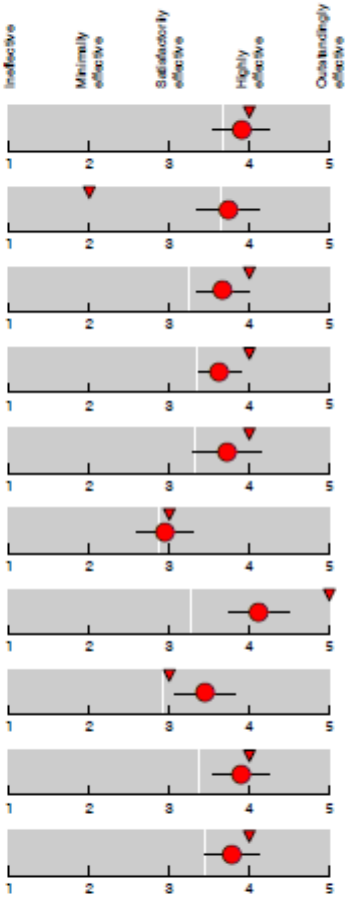
Educational Leadership Practices

Individual Item Report

3. Ensuring Curriculum Quality

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view ▼ Principal's view National reference 2009

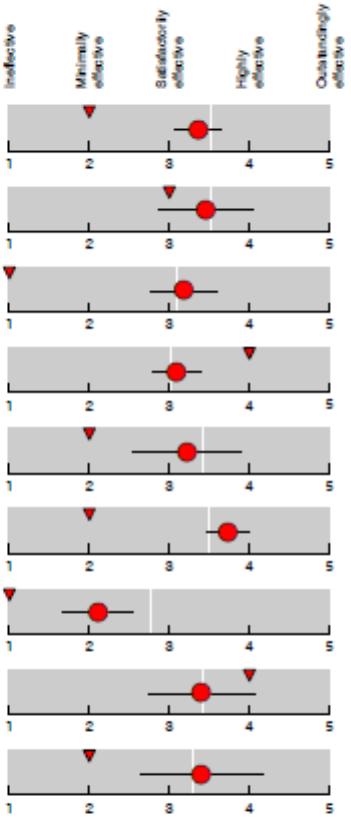
Educational Leadership Practices

Individual Item Report

4. Ensuring the Quality of Teaching

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view ▼ Principal's view National reference 2009

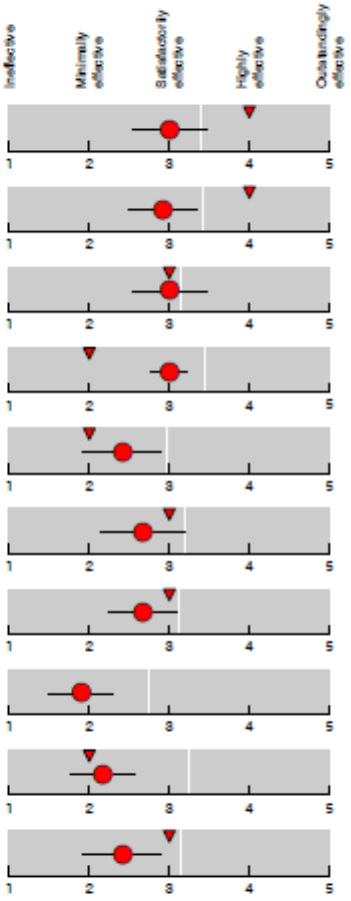
Educational Leadership Practices

Individual Item Report

5. Promoting and Participating in Teacher Learning and Development

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view | Principal's view | National reference 2009

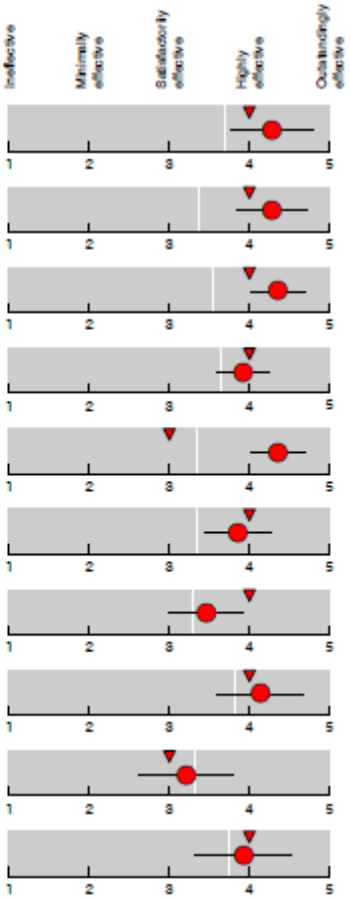
Educational Leadership Practices

Individual Item Report

6. Ensuring a Safe and Orderly Environment

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view | Principal's view | National reference 2009

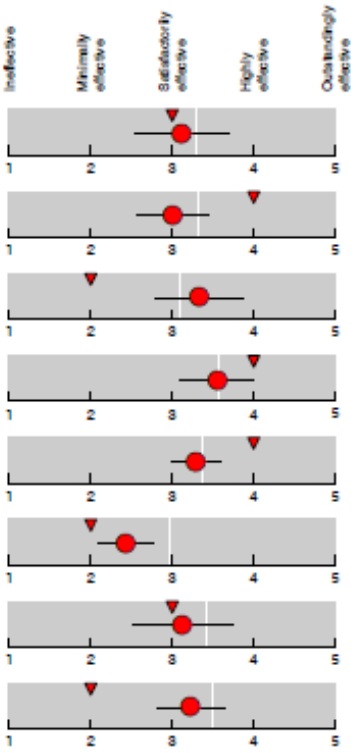
Educational Leadership Practices

Individual Item Report

7. Ensuring Educationally Powerful Connections with Families, Whānau & Community

Number of teachers participating:

How effective is the leadership of your school in ensuring that...



Key: Aggregated teachers' view | Principal's view | National reference 2009

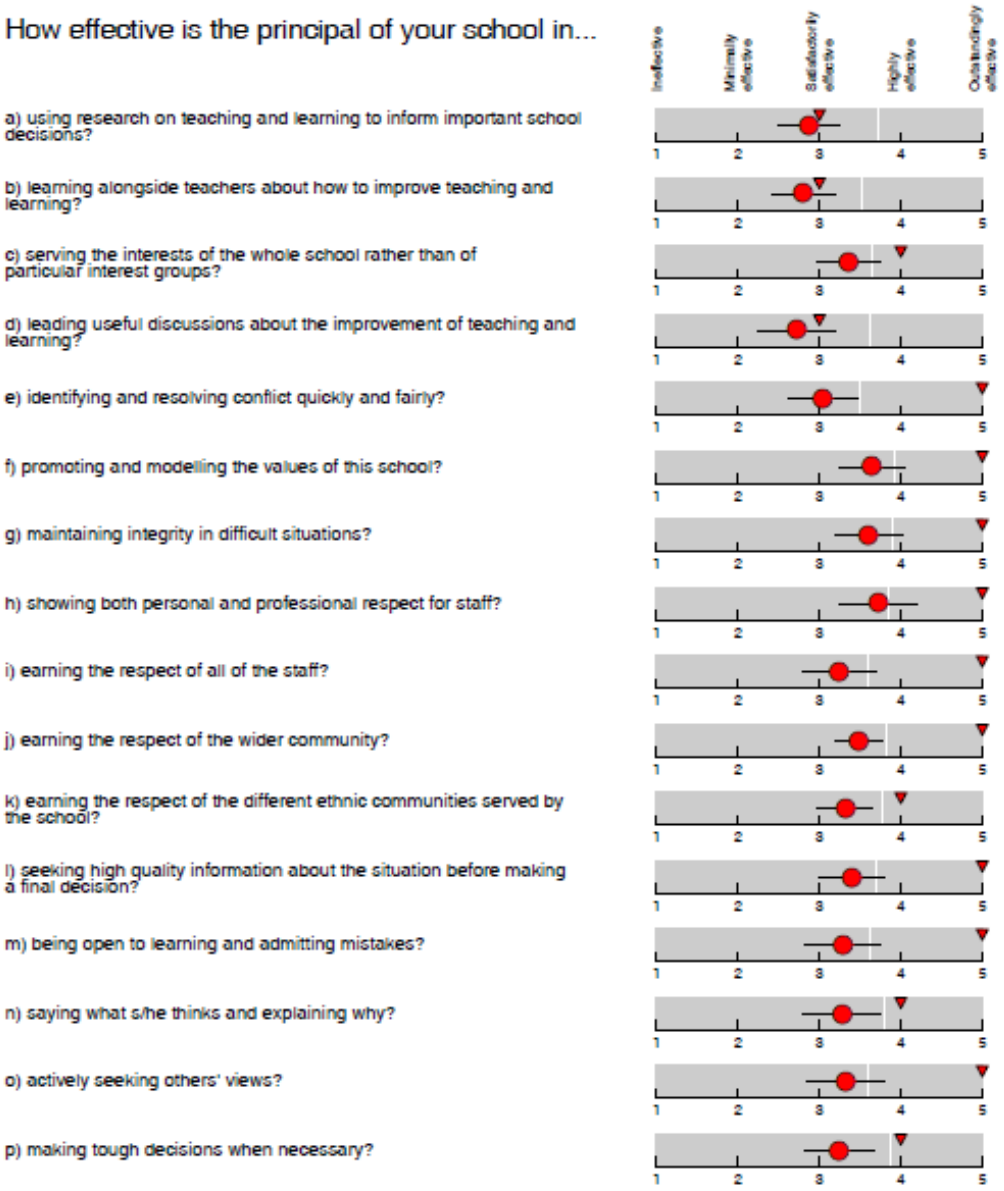
Educational Leadership Practices

Individual Item Report

Principal Leadership

Number of teachers participating:

How effective is the principal of your school in...



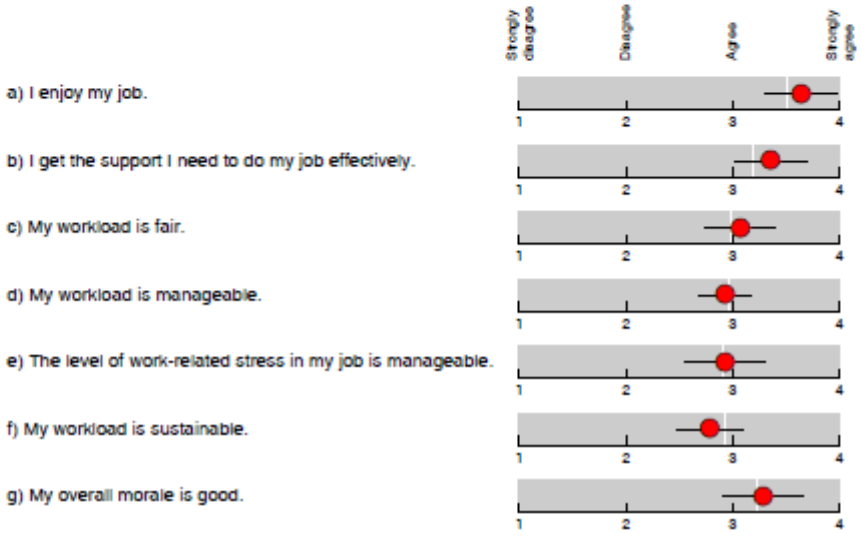
Key: Aggregated teachers' view | Principal's view | National reference 2009

Educational Leadership Practices

Individual Item Report

Teaching Work and Experience

Number of teachers participating:



Key: Aggregated teachers' view | National reference 2009