

---

# Short shrift to long lists

## An alternative approach to the development of performance standards for school principals

Standards for  
school principals

William Louden and Helen Wildy  
*Edith Cowan University, Churchlands, Australia*

99

---

Received September  
1997

Revised March 1998  
Accepted March 1998

**Keywords** *Case studies, Education, Performance, Standards*

**Abstract** *Many educational systems are currently establishing standards frameworks for school principals' work. This paper critiques three current examples and describes an alternative approach. The authors argue that by combining qualitative case studies with probabilistic measurement techniques, the alternative approach provides contextually rich descriptions of the growth in performance on a series of dimensions.*

### Introduction

The notion of "standards" has become a central metaphor of educational reform. There has been widespread attention to the specification of standards expected of students through the development of the national curriculum in England and Wales, the National Statements and Profiles and all of their local variants in Australia (See, 1994), the Ontario Provincial Standards and the Toronto Benchmark Program in Canada (Ministry of Education, 1993), and the New Standards Project in the USA (Ravitch, 1995). Similarly, in the context of teachers' performance there have been many attempts to develop statements of professional standards. These include the activities of the National Board for Professional Teaching Standards (National Board for Professional Teaching Standards, 1989), the Professional Standards for Teaching Mathematics (National Council of Teachers of Mathematics, 1991), and the National Science Education Standards in the USA (National Research Council, 1994), and the National Competency Standards for Beginning Teaching in Australia (National Project on the Quality of Teaching and Learning, 1996).

In the case of school principals, the notion of standards is also at the heart of new performance assessment frameworks in the USA, England and Wales, and Australia. This paper begins with a description of current work in setting standards, considers the strengths and weaknesses of several approaches to standard setting, and describes an alternative strategy for the development of professional standards for school principals. The strategy described in the paper is the first of three phases of a research project designed to investigate and field test a standards framework, describe the framework, and then apply

---

This research is supported by grants from the Australian Research Commission and the Education Department of Western Australia. It is part of a larger project led by William Louden and Max Angus of Edith Cowan University. This paper does not necessarily reflect the position or policy of the funding agencies and no official endorsement of them should be inferred.

Journal of Educational  
Administration, Vol. 37 No. 2, 1999,  
pp. 99-120. © MCB University  
Press, 0957-8234

the framework as both a self assessment process for school principals and a method for setting performance benchmarks. Studies of the effectiveness of this framework for assessing performance are the subjects of future papers.

### Three standards frameworks

In Australia, several state education departments are using competency frameworks to describe standards of principals' work. One set, currently in use in Queensland and Western Australia (in a slightly modified form), follows a conventional competency framework approach. Seven key responsibility areas are identified, including "educational leadership", "curriculum management", and "people management" (Queensland Department of Education, 1996). Each of these key areas is subdivided into a further six or eight key competencies such as "develop an effective performance management process for all staff". The standard of performance in each competency is to be judged by indicators such as "[a] performance management process is in place which is clearly understood by all staff and is consistent with the regulatory framework" or "staff are encouraged to reflect on their performance and identify strengths and areas for development". Table I contains an extract from this material showing the competencies in one key responsibility area and the indicators relating to these competencies.

A somewhat similar approach has been used in England and Wales in development of the National Professional Qualification for Headship (The National Professional Qualification for Headship, 1996). This standard's framework identifies a core purpose, a set of key areas, a set of skills and abilities, and a statement of knowledge and understanding required. Under the key area "People and relationships" appears a set of tasks such as "Ensuring sound management practices throughout the school through which all are clear of their roles and responsibilities". Candidates for this qualification need to be able to demonstrate, among other things, that they are "able to develop

---

Key responsibility area  
and key competencies

Some indicators

---

4 People management  
4.2 Develop an effective  
performance management  
process for all staff

Performance management process is in place which is clearly understood by all staff and is consistent with the regulatory framework  
Staff are encouraged to reflect on their performance and identify strengths and areas for development  
Negotiates explicit expectations of staff roles and responsibilities  
Provides and receives regular and constructive feedback  
Staff take responsibility for their own performance

**Table I.**  
The professional  
development  
framework for  
principalship  
(Australia)

---

effective relationships with parents, governors and others in the local community". The following is an excerpt from Key Area 2 of the National Professional Qualification for Headship.

(iii) People and relationships

Tasks comprise:

T.12 Building teams to meet the school's values, aims and objectives;

T.13 Ensuring sound management practices throughout the school which all are clear of their roles and responsibilities

T.14 Ensuring professional development activities for staff and self are linked to individual and school needs through appraisal and the school development plan;

T.17 Working effectively with the governing body.

Candidates will be expected to show knowledge, understanding, skills and abilities in relation to these tasks at an appropriate level for new headteachers. For example they should:

- have successfully built and managed teams;
- be able to develop effective relationships with parents, governors and others in the local community.

A third example, from the USA, is the Interstate School Leaders Licensure Consortium standards framework (Council of Chief State School Officers, 1996a). There are six standards in the framework. Standard 2, for example, is "A school administrator is an educational leader who promotes the success of all students by advocating, nurturing, and sustaining a school culture and instructional program conducive to student learning and staff professional growth". Evidence of reaching this standard includes knowledge and understanding such as "adult learning and professional development models", dispositions such as being committed to "lifelong learning for self and others", and performances such as "professional development promotes a focus on student learning consistent with the school vision and goals". Table II contains an excerpt from Standard 2 of the Interstate School Leaders Licensure Consortium standards.

Despite their differences in layout and detail, these three approaches to developing standards frameworks for school principals have much in common. The Australian standards framework follows a competency-based model, describing observable behaviours based on a careful job analysis. The English standards framework follows a slightly different course, providing lists of leadership tasks, lists of skills and abilities, and lists of knowledge and understandings. The American standards framework is similar to the English framework, providing lists of knowledge, dispositions and performances, grouped under a set of leadership dimensions. Three common characteristics of these standards frameworks may be noted: hierarchical lists, decontextualised performances, and false dichotomies, features of what Masters (1992) calls "frameworks of precision".

**Table II.**  
Interstate School  
Leaders Licensure  
Consortium (USA)

---

Knowledge	Dispositions	Performances
<i>The administrator has a knowledge and understanding of:</i> Student growth and development	<i>The administrator believes in, values, and is committed to:</i> Student learning as the fundamental purpose of schooling	<i>The administrator facilitates process and engages in activities, ensuring that:</i> All individuals are treated with fairness, dignity and respect
Adult learning and professional development models	The proposition that all students can learn	There is a culture of high expectations for self, students and staff performance
School cultures	Preparing students to be contributing members of society	

---

### *Hierarchical lists*

All three standards frameworks attempt to divide complex professional performances into hierarchical lists of dispositions, knowledge, or duties. Each of the lists is organised around a set of major headings. In the Australian material there are seven key responsibility areas. In the English material there are four parts: core purpose, key areas for development and assessment, skills and abilities, and knowledge and understanding. In the American material there are six standards. Under each major heading there appears another set of sub-headings: five key areas for development and assessment in the English standards, three to eight competencies in the Australian standards, and three kinds of indicators in the American standards. Beneath these subheadings, there is a further subdivision in each case: three to eight tasks in the English example, and four or more indicators in the Australian and American examples. At the lowest level of the hierarchy, there are some 41 separate items in the English lists, 134 items on the Australian lists and 173 on the American lists.

Despite the caveats which accompany each set of lists, warning that “skills and abilities should not be viewed in isolation” (The National Professional Qualification for Headship, 1996), that individual “competencies do not stand alone” (Queensland Department of Education, 1996), and that “indicators here are designed to address the big categories of understanding” (Council of Chief State School Officers, 1996b), the consequence of long hierarchical lists is to fragment professional performance. Careful warnings not to see the lists as fragmentary do not overcome the problem of fragmentation. Within a single incident, principals may demonstrate a whole range of competencies which appear as separate items on separate lists. To use the Australian standards as an example, principals exercise people management skills at the same time as they exercise curriculum management skills at the same time as they exercise educational leadership.

---

*Decontextualised performances*

A second objection that may be made to professional standards which follow the form of these examples is that they separate the performance from the context within which it occurs. For any one competency, the context in which it is required will vary the nature of the performance required of a principal. For example, quite different knowledge and skills would be required to demonstrate the competency “provides and receives regular and constructive feedback” in the context of well established collegial relationship than the knowledge and skills required in giving feedback in the context of an alleged moral impropriety by a teacher. Within most readers’ experience there would be people who were much more effective at providing feedback in the intimate and personal environment of a collaborative work team, or who were much better at providing feedback in the quasi-legal context of an allegation of misconduct. A performance that appears to be an obvious and separate competency when stated in a general form may require a range of different knowledge and skills in different contexts.

*False dichotomies*

A third objection which may be made to standards frameworks based on long lists of duties or dispositions is that the wording of items on the lists implies a degree of precision difficult to realise in real professional contexts. For example, to use the competency quoted in the previous paragraph, the use of words such as “regular” and “constructive” is presumably intended to imply a level or standard of performance. Having the competency requires more than just being seen to give and receive feedback, the feedback must be of a particular kind (regular) and of a particular type (constructive). What counts as regular enough to demonstrate the possession of the competency, or what counts as constructive enough to demonstrate the competency remains unspoken. Rather than supporting a clear judgment of competent or not competent, the indicator begs questions which can only be answered by more minute and detailed specification of the conditions under which it may be performed.

*Probabilistic standards frameworks*

More careful refinement of the format, wording or content of lists of principals’ duties and dispositions cannot overcome these three problems. The long lists strategy is based on an approach to the conceptualisation of standards that is inappropriate to professional performance. In order to avoid problems such as hierarchical lists, decontextualised performances and false dichotomies, it has been argued that an alternative measurement strategy should be used. Masters (1992) has argued that standards should be based, not on “frameworks of precision”, but on “probabilistic frameworks”. Writing in the context of student standards, Masters has identified four key ideas underlying probabilistic standards frameworks: progression, estimation, tasks as examples, and probabilistic interpretation.

*Progression*

The concept of progression may be contrasted with the dichotomous judgments which underlie competency-based standards. Instead of presuming that competence is a quality a principal either has or does not have, it may be presumed that competence occurs on a continuum. For example, rather than asking whether a principal does or does not create “a culture of high expectations for self, student and staff performance”, probabilistic standards would locate the level of performance on a continuum of performances in the same domain.

*Estimation*

The purpose of assessment within a probabilistic framework is to provide an estimate, not a final determination about the individual’s current level of attainment. Instead of expecting that a standards framework could be used to determine whether or not a principal “provides and receives regular and constructive feedback”, assessment in a probabilistic framework indicates what evidence there might be that would support an estimate of high, medium or low performance on a particular continuum.

*Tasks as examples*

Within a probabilistic framework exemplars or items are interchangeable and expendable, useful only to the extent that the items help to estimate a performance on a continuum. This may be contrasted with the fixed items on a competency-based list, each thought to describe an essential element or aspect of competent performance. Whereas an exemplar or item may be dropped from a probabilistic standards framework if it fails to assist in locating a level of performance, the knowledge, skills and dispositions listed in a competency-based standards framework are chosen because they are each thought to identify an essential aspect of competent performance.

*Probabilistic interpretation*

Finally, within a probabilistic standards framework, the goal is to be able to describe what can normally be expected of people with a level of performance, rather than to determine finally whether a person can or cannot achieve mastery of a particular competency. For example, as some of the material in the next section will explain, a probabilistic assessment framework can identify aspects of performance which other principals regard as associated with poor or good or very good performance.

**An alternative approach**

There are several important strengths of the long list approach to defining performance standards. For one thing, the development processes used in compiling and validating the lists are often hugely consultative. There have been, for example, more than 10,000 school administrators involved in the development of the American ISLLC Standards (Council of Chief State School

---

Officers, 1996c). A second strength of these approaches is that they provide comprehensive descriptions of the elements of principals' work. The remainder of this paper describes a process of standards development which attempts to be both consultative and respectful of the range of territory in principals' work, without accepting the weaknesses associated with "frameworks of precision".

The Performance Standards for School Principals project is a collaborative undertaking with the Education Department of Western Australia. The Department is currently introducing a new performance management process for its 770 school principals (Education Department of Western Australia, 1996). The research project is exploring the possibility of using a probabilistic standards framework to support the judgments made by principals and their superintendents. The framework is not intended to describe in detail the territory of principals' work; the purpose is to specify and illustrate the range of performance within the territory. The project has proceeded in three phases. The first phase was a pilot project, which explored a series of methodological questions. The second phase involved development of the dimensions of a standards framework. The third phase, which is currently under way, involves preparation of progress maps which describe progression on each of the dimensions of the standards framework. Once the standards framework is developed, further work will be needed to develop a process of assessing individual principal's performance in relation to the framework.

### *Phase 1: pilot study*

The first step in the research process was to select a trial dimension, a potential continuum of performance. Rather than make a firm commitment to a full set of dimensions, the project began with one of the dimensions commonly found in principals' standards frameworks. The dimension chosen was "staff management", called "people management" in the Australian materials and "people and relationships" in the English materials.

The second step was to approach a wide group of principals and ask each of them to talk about the some of the dilemmas they encountered in working with staff. On each occasion the researchers spent an hour or more interviewing each principal in their school. The opening questions for the interviews asked principals to give an example of recent, practical problems related to a set of dilemmas (Berlack and Berlack, 1981; Louden, 1991; Tripp, 1993). The following sets out the dilemmas and a sample interview question related to each dilemma:

#### *How to back up all staff and be honest about standards*

"How do you handle it when a parent or a student claims that a teacher has done the wrong thing?"

#### *How to be the boss and also be a member of the team*

"How do you get staff to agree with you about something?"

#### *How to be methodical without being inflexible about methods*

How important is it for you to be methodical? For others to be methodical?

*How to balance looking in and looking out*

What do you think you have to do in order to have other people think you are a good principal? Parents? Colleagues? Students?

*Keeping a clear personal view and accommodating others' views*

"When you were new to the school, what sorts of things did you want to do? How did you go about this?"

*Responding to regulation and acting autonomously*

"How do you manage the temporary teacher returns to the Department?"

A total of 40 principals were visited and interviewed at this stage. The contexts of their work ranged from large city secondary schools with more than 100 on the teaching staff to three-teacher primary schools serving remote Aboriginal communities. Principals were male and female, experienced and inexperienced, and working in urban, rural and remote contexts. Based on these interviews, a total of 65 short narrative case studies were prepared. Following methods developed in the context of professional standards for teachers (Louden, 1994), the case studies were written by the researchers using audio-tape and verbatim written records. The case studies are about 500 words long, long enough to capture some of the contextual detail of the dilemma; short enough to be manageable for readers and raters. First drafts of these stories were returned to principals for amendment and improvement, and to ensure that the accounts contained no breach of confidentiality. The following provides an example of one such case study:

*Supervision*

A parent phoned me with a general complaint about a teacher. I said "Tell me specifically what the problem is." So the parent listed: no Unit outline, only two pieces of work collected and then not handed back, work not completed in class and then no instructions for finishing it, and finally, "My son says the teacher is boring". I thanked the parent for the details of the complaint and asked if she had contacted the teacher. "No", she said "I didn't because it wouldn't do any good."

I suggested she approach the teacher with the specifics and give the teacher time to respond. Actually, we spent quite a bit of time planning the conversation. I coached her in asking questions rather than making judgements, to avoid a defensive response from the teacher.

I'm new to this school so I needed to do my homework. I made enquiries about the student and found he was one of those whose views were to be taken seriously: excellent academic record, no behaviour problems. I also made some enquiries about the teacher. The deputies claimed that they had had many behaviour problems from his classes over the year. Others painted a picture of a teacher with a long history in the school: he coasts along, then there are complaints, then some improvement, but then the cycle is repeated.

As soon as I could I met with the Head of Department. I saw the responsibility for the teacher's performance rested with the Head of Department, particularly in curriculum matters such as assessment practices, marking guidelines, course overview and planning. That a student had not received a Unit outline was clearly the responsibility of the Head of Department. I asked him about his department's assessment policy, Unit outlines and so on. It became clear that he was not carrying out some of his basic duties. I asked him how much time he spent supervising his teachers and he admitted to very little. "It's often rewarding to look at students' work" I told him.

The Head of Department agreed to visit the teacher in his classroom and to let me know when he had spoken to the teacher. From his comments later, I realised that what had taken

---

place between the Head of Department and the teacher was general, superficial and not likely to lead to any changes on the part of the teacher. It was obvious that I hadn't given the Head of Department enough guidance. My focus is working not as much with the teacher as with the Head of Department to develop his skills of supervision.

As long-time observers of principals, the researchers had strong opinions about whether the performance described in a case study belonged at the high or low end of a performance continuum. The next step in the research process, however, was to find out what sort of estimates a large group of principals might make about the case studies. Twenty-four case studies were chosen from the pool of 65 available at the time. These case studies were then mailed to another set of 150 principals. More than 100 of these principals accepted the researchers' invitation to each of them rating a group of five stories on the dimension "staff management". Specifically, principals were asked to "read this account of a principal working with staff" and then to "rate the quality of the principal's performance" on a four point scale (poor, adequate, good, very good, unable to rate).

The next step in the research process was to analyse the ratings using the extended logistic model of Rasch (Andrich, 1991; Masters, 1991; Wright and Masters, 1982; Wright and Stone, 1979) using the computer program Quest (Adams and Koo, 1992). The Rasch method provides an estimate of the position of each case study on a continuum, and confirmation of whether items that are grouped at the bottom, middle and top of the continua represent approximately equivalent levels of competence.

The analysis of the pilot data located the case study called "Supervision" above the middle of the continuum "working with staff". The following describes a principal's performance rated at the lower end of the same continuum:

*New school: old problems*

I had been looking forward to my transfer but the first part of the year has been a real struggle. I want to build relationships among the staff to develop confidence in the administration but I keep coming across blockages.

It's difficult for me to find out what's going on in the school. I know a lot of good things are happening but it's hard to track them down because people seem to be working separately. Most of the staff have been here for many years – quite a few for more than 25 years. A number of teachers are not far off retirement.

I am dealing with two staff problems.

One is a teacher, with over 25 years of experience, confident, lots to say for himself, who is new to the school. He got off side with staff pretty quickly by talking non-stop in the staff room when people were trying to work. In the first two weeks I had letters of complaint from parents claiming the curriculum was not being followed and of his use of racist language. Worse, he had told his students that exams were not important. You can imagine how well that went down in this school with its reputation for outstanding success in the university entrance examinations.

I documented everything – letters from parents, complaints from teachers, comments from students. When I confronted him with the evidence he was shocked. He didn't agree with any of the comments. I spent a great deal of time with him, listening to his attempts to justify himself. By the end of term 1 he took leave. He'll be away from school for some time on long service leave. I don't know what I'll do if he returns.

My other problem is less easily dealt with. The teacher is a Head of Department who has been here for many years and developed a reputation for high standards. The department is efficiently organised and the university entrance examination results are excellent. I am concerned about her methods. She demands a lot of students, uses fear, put downs, sarcasm and the like. Parents have complained to me about her manner.

My concern is that, in pushing for results, she is destroying students' love of learning. Whenever I tackle her she always has a sound philosophical base for what she is doing. She can talk to you for hours about standards. She takes a pride in being out of step with current approaches to teaching and learning. What really annoys me is seeing her walk off every afternoon at 3.30 empty handed – no marking, no preparation, no reading to do out of hours. I'm at a loss to know how deal with her. All I can do is chip away at members of her department. That's hard though because they are very loyal to her.

*Phase 2: dimensions of the standards framework*

A probabilistic standards framework requires a description of the dimensions of performance in a domain of learning and an account of the nature of progression on each of these dimensions. As the earlier discussion of English, American and Australian standards frameworks for principals showed, there may be many plausible dimensions to describe a particular domain of learning.

The pilot phase of this project began with the assumption that the standards framework could be built around duties-based dimensions such as “staff management” or “school planning” which appear in previously published standards frameworks and which had been assumed by our collaborative partner, the Education Department of Western Australia. These duties were:

- school planning;
- staff management;
- curriculum management;
- student management;
- school and community interaction;
- management of school finances, resources and operations.

Several aspects of the pilot study disturbed this preconception. The first of these was that as the case studies were collected it became clear that the dilemmas faced by principals could not easily be allocated to a single duty or key responsibility area. For example, many of the case studies collected with the intention of illustrating the dimension “staff management” seemed equally to illustrate dilemmas faced by principals in their contact with parents and community members.

The second observation which disturbed our preconceptions about duties-based dimensions was the principals' qualitative comments about the case studies mentioned. In addition to rating the case studies, principals were asked to describe briefly the characteristics of the principal's performance in the case study description. For example, in relation to the case study “Supervision”(see p. 106), principals made the following comments:

- 
- “Is interested in standards of teaching and learning in the school.”
  - “Listens well; makes inquiries before coming to conclusions.”
  - “Works with the parent to support her in giving appropriate feedback to the teacher.”
  - “Delegates responsibility appropriately, but also provides support to the head of department who has the responsibility.”
  - “Has a clear focus on staff development.”
  - “Consultative, well informed, systematic.”
  - “Clear understanding of what is required to make feedback contribute to improvement.”
  - “Helpful and supportive to all stakeholders.”
  - “Constructive approach: looks for a solution to the problem, rather than for a person to blame for poor performance.”

In contrast, the comments made about the work of the principal described in “New school, old problems” (p. 107) included:

- “Does not seem to confront problems as soon as they appear. Seems to have ‘saved up’ the problems until they are very large.”
- “Does not seem to place these examples of ‘problem’ staff into an established system or pattern of regular professional conversations with staff.”
- “Gives the impression that the principal is working in isolation. What use might have been made of other members of the administrative team in the school?”
- “Understands that the school works as a set of inter-connected systems, but in the beginning (at least) is not quite clear how the parts of this school system relate to the whole.”

Such comments suggested that the dimensions of the standards framework might also need to include interpersonal skills such as listening and negotiating, or moral dispositions such as fairness and courage. In order to explore this possibility, a content analysis was conducted of the qualitative data provided by the 100 or so principals who responded in the pilot study. The words they used were grouped and classified into a set of subheadings. A series of interpersonal skills and moral dispositions were identified, based on the written comments made by principals. Through several iterations these skills and dispositions were then used to classify the case studies that had been collected. At the end of this process, two sets of dimensions were thought to be adequate to account for the range of qualitative comments made by principals responding to the pilot case studies, for the researchers’ own readings of the case studies, and for the understanding of duties current in the school system. They appear in the list below:

- (1) *Interpersonal skills:*
  - listening;
  - confronting complexity;
  - negotiating;
  - collaborating;
  - leading.
- (2) *Moral dispositions:*
  - fairness and consistency;
  - patience and persistence;
  - courage and decisiveness;
  - sensitivity and tact;
  - flexibility and creativity.

Rather than abandon the six duties-based dimensions identified by the school system, or choose between the set of interpersonal dimensions and the set of moral dimensions, it was decided to use all three sets of dimensions. Each of them promises to have some explanatory power; together they may help to capture the complexity that characterises principals' work.

The final stage in the development of the dimensions of the standards framework was to revisit each of the 65 pilot case studies and to re-classify them according to all three sets of dimensions: the duty being carried out, the interpersonal skill most salient in the principal's performance, and the moral disposition most salient in the principal's performance. Figure 1 provides a schematic overview of the standards framework.

Figure 1 begins with the six duties-based dimensions of the standards framework. On each of these dimensions, a set of case studies is mapped, ranging from cases judged to portray lower levels of performance to cases judged to portray higher levels of performance. Location of the case on the dimension is determined by the quantitative ratings; the indicators associated with each case are determined by the qualitative comments raters make on each case study.

Similarly, the same set of case studies are mapped with respect to the five interpersonal skills dimensions and the five moral disposition dimensions. In principle, a case study mapped at the lower end of performance with respect to a particular duty, say school planning, may be mapped at the higher end with respect to, say, listening or sensitivity and tact.

### *Phase 3: progression on the dimensions*

The set of dimensions developed in the second phase of the project provided an account of the content of principals' work, the duties to be undertaken and the skills and dispositions required to perform the duties at a high level of performance. The third phase of this project was designed to develop progress

Duties	Lower performance			...	Higher performance				
School planning	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Staff management	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Curriculum management	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Student management	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
School & community interaction	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Management of school finances, resources & operations	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators

Interpersonal Skills	Lower performance			...	Higher performance				
Listening	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Confronting complexity	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Negotiating	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Collaborating	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Leading	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators

Moral Dispositions	Lower performance			...	Higher performance				
Fairness and consistency	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Patience and persistence	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Courage and decisiveness	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Sensitivity and tact	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators
Flexibility and creativity	Case study Indicators	→	Case study Indicators	→	Case study Indicators	→	...	→	Case study Indicators

**Figure 1.**  
Schematic standards framework

maps which describe the progression in performance on each of the dimensions in the standards framework.

In order to complete the progress maps, a further 25 case studies were collected, with a view to providing 15-20 case studies for each dimension of duties, skills and dispositions. A total of 67 principals have been interviewed, including almost all of the principals in one country school district (24 principals) and almost all of the principals in one city school district (29 principals).

Two further rounds of rating have been conducted, using 24 case studies allocated to the “staff management” dimension and 20 case studies allocated to

the “school and community interaction” dimension. A total of 197 respondents rated the staff case studies and 148 rated the community case studies. Table III provides an analysis of ratings for the 24 staff management case studies. The data were analysed using the Rasch method and the Quest computer program. In this analysis, the four-point scale was collapsed to a two-point scale, with ratings “poor” and “adequate” grouped and the ratings “good” and “very good” grouped. Approximately 10 percent of the least consistent raters, as identified by the analysis, have been deleted.

The statistics included in Table III are threshold and infit mean square. The threshold is the parameter estimate that describes the difficulty of the item. It indicates the estimated location of the case study on the staff management dimension of the performance map. Scores for the 24 staff management case studies in Table III varied from  $-9.30$ , indicating that raters found the case study “The hangover” was easiest to agree represented good or very good performance, to  $10.26$ , indicating that the raters found the case study “Confronting” hardest to agree represented good or very good performance. The threshold values between  $-9.30$  and  $10.26$  locate the order and intervals of the other cases studies on this dimension of the performance map. The zero score associated with “No, it’s not favouritism” indicates no range of responses among raters; all raters found it very difficult to agree that this case study represented good or very good performance. Similarly, the perfect score for “More staff meetings and shorter” indicates that all raters found it very easy to agree that this case study represented good or very good performance.

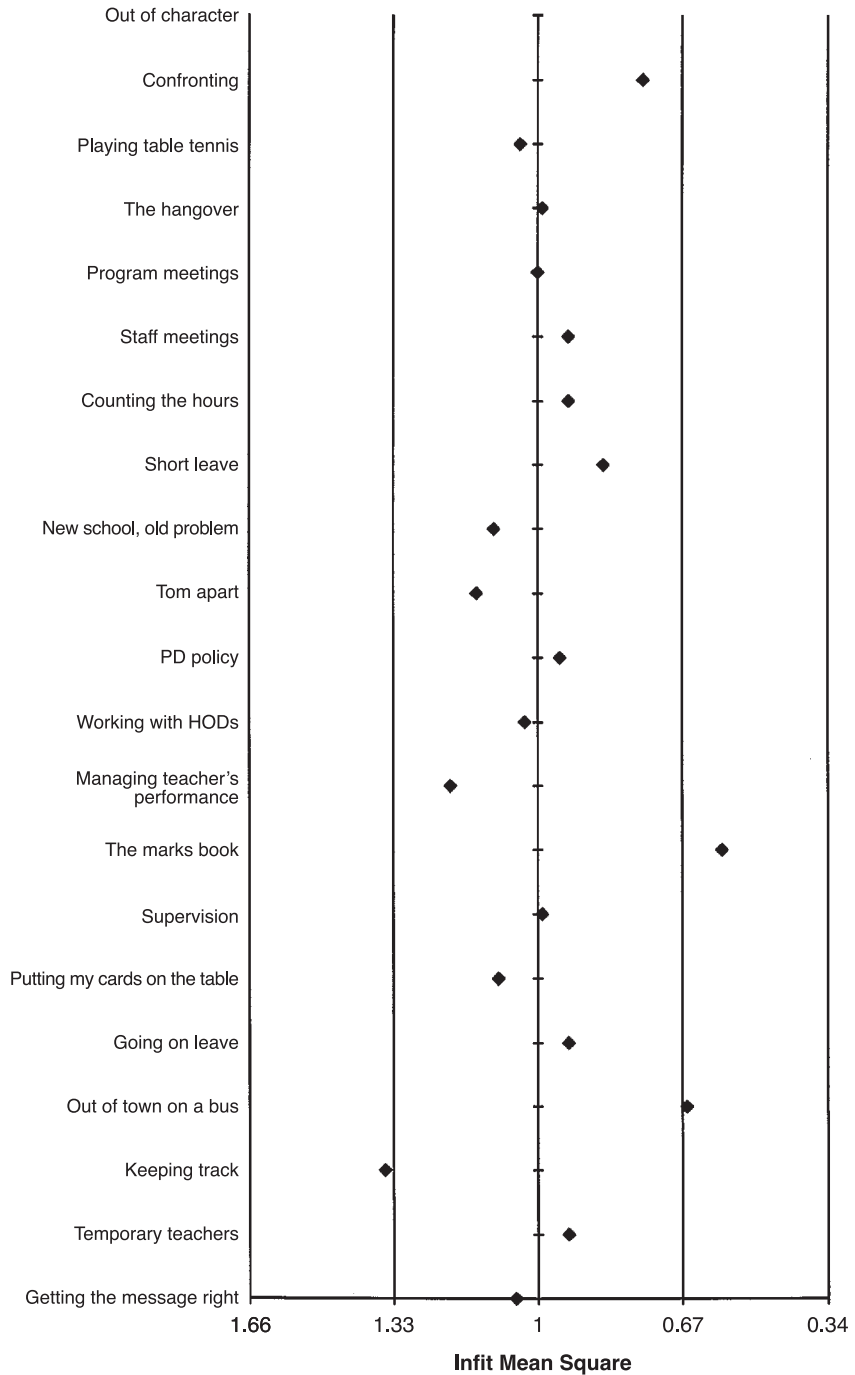
The infit mean square statistic indicates the degree of fit between the data and the model. For each item, an infit mean square of  $1 + x$  indicates  $100x$  percent more variation between the observed and predicted response patterns than would be expected if the data and the model were compatible; an infit mean square of  $1 - x$  indicates  $100x$  percent less variation between the observed and predicted response patterns than would be expected if the data and the model were compatible (Adams and Koo, 1992).

Figure 2 provides an item fit map for the 24 case studies on the duties-based dimension “Staff management”. The closer the infit statistic is to the value 1, the more closely the data fit the model. Items which are not close to the value 1 are normally removed from the scale. How close to the value 1 counts as close enough is not determined by the Rasch method or the software; it requires a judgment based on the size of the sample and on how high the stakes are in the use of the scale. For the purposes of this study, 33 percent variation between the observed and expected response patterns would be adequate to maintain an item in the scale. At this level the items “Keeping track”, “Out of town on a bus” and “The marks book” would be removed from the scale.

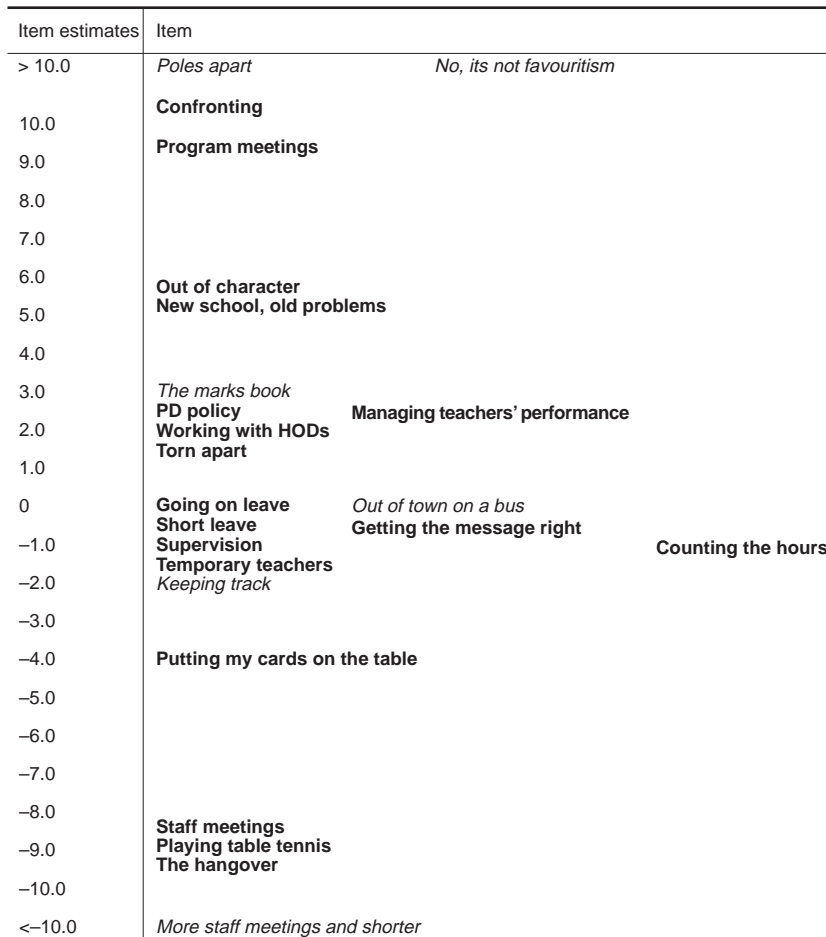
Figure 3 provides a progress maps for “Staff management”. Items with infit mean squares with a variation of more than 33 percent between the observed and expected response, and items with zero and perfect scores are shown in italics; items with adequate fit statistics are shown in bold.

Item number	Item name	Threshold 1	Infit mean square
1	Getting the message right	-1.06 0.36	1.05
2	Temporary teachers	-1.34 0.38	0.93
3	Keeping track	-1.87 0.39	1.35
4	Out of town on a bus	-0.03 0.45	0.66
5	Going on leave	-0.35 0.32	0.93
6	Putting my cards on the table	-3.91 0.77	1.09
7	Supervision	-0.90 0.50	0.99
8	The marks book	2.28 0.73	0.58
9	Managing teachers' performance	1.94 0.46	1.20
10	Working with HODs	1.31 0.51	1.03
11	PD policy	2.15 6.40	0.95
12	Torn apart	1.07 0.49	1.14
13	New school, old problem	4.27 1.03	1.10
14	Poles apart	Zero score	
15	Short leave	-0.53 0.53	0.85
16	Counting the hours	-1.25 0.56	0.93
17	More staff meetings and shorter	Perfect score	
18	Staff meetings	-8.35 0.49	0.93
19	Program meetings	9.53 1.41	1.00
20	The hangover	-9.30 0.52	0.99
21	Playing table tennis	-8.93 0.52	1.04
22	Confronting	10.26 1.14	0.76
23	No, it's not favouritism	Zero score	Zero score
24	Out of character	5.02 1.21	0.05

**Table III.**  
Staff management item  
estimates (thresholds)  
( $N = 197, L = 24$   
Probability level =  
0.50)



**Figure 2.**  
Staff management item  
fit map. Worse fitting  
raters deleted (~10 per  
cent) ( $N = 197, L = 20$ )



**Figure 3.**  
Staff management  
progress map. Worst  
fitting raters deleted  
(~10 percent) ( $N = 197$ ,  
 $L = 24$ )

## Discussion

As the number of readers and raters has grown, several issues have come into sharper focus. One issue concerns the range of reactions given to some stories. Although most of the case studies have fit statistics that allow the case studies to be located on the performance maps, some case studies seem to draw different responses from some kinds of raters. In professional development workshops using the case studies, some differences between primary school and secondary school principals' judgments have emerged. Principals in very small primary schools, for example, may be inclined to underestimate the political complexity of managing change in large secondary school contexts. Anecdotally, it seems that this leads them to give relatively lower ratings to case studies in which the principal manages by indirect means. Similarly, there may be differences between experienced and inexperienced principals, or between male and female principals. These differences of emphasis which may

distinguish different sub-groups of principals have not yet been tested empirically.

### *Sampling*

A second issue concerns sampling. How representative of the whole population are the principals who were interviewed during the collection of case study material, or the principals who have rated the case studies?

Case studies were collected from interviews with a stratified random sample of 24 rural and 29 metropolitan principals, augmented by a top-up convenience sample of an additional 14 metropolitan principals. The total of 67 principals interviewed included 43 metropolitan and 24 rural principals. This constitutes a slight over-representation of metropolitan principals (64 percent of principals interviewed located in metropolitan schools, compared with 52 percent of the population of principals located in metropolitan schools). Although only 22 percent of the principals interviewed were female, this is almost exactly the proportion of female principals in the population (23 percent of all principals are female). Primary school principals were under-represented among the group interviewed. Whereas for the whole population 71 percent of principals are in primary schools, only 5 percent of the case study principals were in primary schools (Education Department of Western Australia, 1995).

Since the end of the pilot study, ratings have been collected from an additional 350 principals. Initially the ratings were collected through a stratified random sampling strategy. The response rate of 60 percent was considered adequate, but was also supplemented by collecting ratings at district meetings of principals, in the context of professional development activities offered by the authors. This has proven to be a more efficient data collection strategy. A State-wide schedule of district meetings has been established; by the time all of the ratings data have been collected, it is expected that almost all of the 770 Education Department principals will have been involved in the project as raters. Among the current set of raters, female principals are substantially over-represented, primary principals are slightly over-represented and rural principals are slightly over-represented. Whereas only 23 percent of principals in the population are female, 43 percent of raters were female; 79 percent of principals are in primary schools and 71 percent of raters were primary principals; 45 percent of principals are in rural school and 48 percent of raters were in rural schools.

Of all of these differences in the sampling of cases and raters, the over-representation of female raters is largest. By following the new sampling strategy of attending principals' meetings in all of the State's school districts, the proportion of female raters will gradually grow closer to the proportion of female principals in the whole population.

### *Expert judgments*

Behind the issue of sampling stands an important conceptual issue. If the sampling were perfect, the performance maps would represent the opinions of

---

the whole range of principals, according to gender, experience and location. The conceptual issue, however, is whether performance standards should be based on the opinion of the whole range of principals or on those best qualified to make judgments about performance. For example, should raters include principals whom others regard as incompetent? Although no attempt has been made to judge the performance of raters, the current progress maps must represent the view of both more competent and less competent principals. The Rasch model fit statistics, however, suggest that for most of the cases the range of opinions about the performance is relatively small, notwithstanding the range of gender, location, school type and competence of principals.

An alternative approach to the question of rater competence is to compare the ratings on the progress maps with ratings of a small group of acknowledged expert principals. It is not yet known whether “experts” would rank the cases in a different order. Presuming that they did rank some cases differently, it may add further richness to the discussions about the skills and dispositions required for successful principalship to be able to contrast cases rated differently by the whole range of current principals and by a subset of acknowledged experts.

### *Complexity*

The fourth and final issue concerns Occam’s razor. This paper began with a critique of the long lists which characterised other attempts to develop performance frameworks for school principals. It might be argued that these long and complex lists have been replaced by a short set of lists and a complex web of case studies. The decision to collect three separate ratings for each case study, and to represent the ratings on three complementary sets of dimensions, adds to the complexity. When there has been an opportunity to test the use of the framework in a variety of contexts, it may be fruitful to apply William of Occam’s medieval prescription about theoretical complexity — “Entities should not be multiplied unnecessarily”.

### **Conclusion**

This paper has described an alternative method of developing a standards framework. It combines two well-established methodologies: the narrative method of written case studies, and the mathematical technique of Rasch modelling. The goal of the research project is to explore the development of probabilistic standards, in contrast to the more familiar competency-based method of setting professional standards. Several weaknesses of competency-based standards have been identified: they require long lists of items – sometimes hundreds – to cover the range of principals’ work. They separate professional performance from the contexts in which it occurs, in direct conflict with a lesson of everyday experience in school leadership, which is that the details of the context of each event are crucial in judging the relative success or failure of a particular performance. Third, competency-based standards offer

false dichotomies, promising to separate principals into those who reach a standard and those who fail to reach a standard.

In contrast to competency-based standards, the probabilistic standards framework used in this research project locates performance on a set of continua, not on either side of a dichotomy; it offers an estimate of performance, rather than a final determination; it uses assessment items as expendable exemplars rather than as essential elements of principals' work; and it describes what can normally be expected of people at a given level of performance rather than whether an individual has attained mastery of a skill.

The research has demonstrated that it is possible to construct a probabilistic standards framework describing principals' performance. Dimensions representing duties, interpersonal skills and moral dispositions have been developed inductively, based on pilot study principals' reactions to reading and rating a series of case studies of principals' work. A performance map has been developed for two of the duty-based dimensions, "School and community interaction" and "Staff management". A total of 90 case studies have been prepared and will be used in the production of performance maps for the remaining four duties, five interpersonal skills and five moral dispositions.

Field testing of project materials with principals has demonstrated that this approach to representation of standards may be very useful as a self assessment process in performance management. The case studies provide recognisable contexts for principals' performance and capture the complex interconnectedness of elements of their work. Unlike the long lists of duties and dispositions contained in competency-based standards, the sets of case studies do not attempt to describe or cover the territory of principals' work. Instead they abstract from commonly shared experiences and familiar dilemmas those interpersonal skills and moral dispositions which appear to differentiate the quality of principals' performance. For these reasons, the rich reality of the case studies appears to engage principals at a deeper level than lists of duties and dispositions. At workshops and seminars based on this material, principals often remark, "exactly the same thing has happened to me". Equally, principals seem engaged by debate about whether or not a particular principal in a case study has acted appropriately, and what else the principal may have been wise to do.

As the project continues, several more forms of commentary will be available to enrich these discussions: ratings of the performance in each case, dimensions which locate each case on a performance continuum, and a set of alternative lenses for understanding each of the cases — duties, interpersonal skills and moral dispositions.

The next phase of this project explores the ratings of individual case studies in relation to rater characteristics. Specifically, responses will be analysed to establish whether there are differences between women and men in judging the quality of the performance represented in the cases; whether length of experience makes a difference to judgements; whether type of experience — primary, secondary — makes a difference; and whether the location of that

experience – metropolitan, rural – has an impact on judgements. In addition, samples of deputy principals, heads of department and teachers will be invited to rate the cases studies. The ratings of each sample will be compared with the ratings of principals. A small panel of acknowledged exemplary principals, identified using a range of techniques, will rate all case studies and their ratings will be compared with those of the larger sample of principals.

Subsequent phases of the research project will assess the effectiveness of the framework as a self assessment process and as a means of establishing performance benchmarks. A longitudinal study of a small number of principals is planned to monitor changes in practice over time and links between improved performance and a range of school outcomes.

### References

- Adams, R.J. and Koo, S.T. (1992), *Quest: The Interactive Test Analysis System*, Australian Council for Educational Research, Hawthorn.
- Andrich, D. (1991), “Applying modern social measurement theory for the general analysis of profiles of performance”, *Australian Journal of Education*, Vol. 35 No. 1, pp. 11-23.
- Berlack, A. and Berlack, H. (1981), *Dilemmas of Teaching*, Methuen, New York, NY.
- Council of Chief State School Officers (1996a), *Interstate School Leaders Licensure Consortium*, Washington, DC.
- Council of Chief State School Officers (1996b), *Interstate School Leaders Licensure Consortium. Standards for school leaders. Draft model standards*, Council of Chief State School Officers, Washington, DC.
- Council of Chief State School Officers (1996c), “Work in progress: The newsletter of the Interstate School Leaders Licensure Consortium of CCSSO (ISLLC)”, *The Newsletter of the Interstate School Leaders Licensure Consortium of CCSSO (ISLLC)*, Vol. 1 No. 1, p. 4.
- Education Department of Western Australia (1995), *Education Statistics Bulletin*.
- Education Department of Western Australia (1996), *Policy Framework of Performance Management*, Education Department of Western Australia, Perth.
- Louden, W. (1991), *Understanding Teaching*, Teachers College Press, New York, NY.
- Louden, W. (1994), “Setting standards in teaching”, in Ingvarson, L. and Chadbourne, R. (Eds), *Valuing Teachers’ Work: New Directions in Teacher Appraisal*, Australian Council for Educational Research, Melbourne, pp. 96-111.
- Masters, G. (1991), “Assessing achievement in Australian schools”, discussion paper commissioned by the Industry Education Forum, Australian Council for Education Research, Melbourne.
- Masters, G.N. (1992), “Towards a national framework for assessment and reporting”, *Unicorn*, Vol. 18 No. 1, pp. 66-77.
- Ministry of Education (1993), *Provincial Standards: Mathematics*, Ministry of Education and Training, Toronto.
- National Board for Professional Teaching Standards (1989), *Towards High and Rigorous Standards for Teaching Practice*, National Board for Professional Teaching Standards, Detroit.
- National Council of Teachers of Mathematics (1991), *Standards for Teaching Mathematics*, National Council of Teachers of Mathematics, Reston, VA.
- National Professional Qualification for Headship (1996), *Key Principles and Draft National Standards for New Headteachers*.

- National Project on the Quality of Teaching and Learning (1996), *National Competency Standards for Beginning Teaching*, Australian Teaching Council, Sydney.
- National Research Council (1994), *National Science Education Standards*, National Academy of Sciences, Washington DC.
- Queensland Department of Education (1996), *The Professional Development Framework for Principals*, Queensland Department of Education, Brisbane.
- Ravitch, D. (1995), *National Standards in American Education: A Citizen's Guide*, Brookings Institution, Washington, DC.
- See, M.C. (1994), *Producing a National Curriculum: Plans and Paranoia*, Allen & Unwin, Sydney.
- Tripp, D. (1993), *Critical Incidents in Teaching: Developing Professional Judgement*, Routledge, London.
- Wright, B.D. and Masters, G. (1982), *Rating Scale Analysis*, MESA Press, Chicago, IL.
- Wright, B.D. and Stone, M.H. (1979), *Best Test Design*, MESA Press, Chicago, IL.