

Conceptualising and Evaluating Teacher Quality: Substantive and methodological issues

Lawrence Ingvarson and Ken Rowe¹

Australian Council for Educational Research

Paper presented at the *Economics of Teacher Quality* conference
Australian National University, 5 February 2007

Abstract: Whereas findings from recent research highlight the importance of *teacher quality* in improving students' academic performances and experiences of schooling, substantive and methodological issues surrounding the conceptualisation and evaluation of *teacher quality* are not well-understood. Such deficiencies are particularly evident in claims for 'findings' derived from econometric research – especially from those studies that merely employ conceptualisations and proxy 'measures' of *quality* in terms of teachers' qualifications, experience, and students' academic outcomes. Moreover, the econometric models fitted to the available, mostly aggregated data, typically fail to conceptualise and 'measure' *teacher quality* in terms of what teachers should *know* (subject-matter knowledge) and *be able to do* (pedagogical skill). Nor do such models account for the measurement, distributional and structural properties of the data for response and explanatory variables – failings that all too frequently yield misleading interpretations of findings for both policy and practice.

Following brief introductory comments related to current contexts, the paper focuses on two approaches towards the resolution current deficiencies – both of which have important implications for conceptualising and evaluating *teacher quality*, namely: (a) capacity building in teacher professionalism grounded in evidence-based pre-service teacher education content and subsequent in-service professional development, and (b) the specification and evaluation of teaching standards. The paper concludes by arguing that since the most valuable resource available to any school are its teachers, there is a crucial need for both a substantive and methodological refocus of the prevailing economic *teacher-quality/student-performance/merit-pay* research and policy agenda to one that focuses on the need for capacity building in teacher professionalism (and its evaluation) in terms of *teaching standards* related to what teachers should *know* and *be able to do*.

Introductory comments

Consistent with the adoption of corporate management models in educational governance and the prevailing climate of outcomes-driven economic rationalism in which such models operate, policy activity related to issues of: *accountability, assessment, standards monitoring and benchmarking, performance indicators, quality assurance, teacher quality, school and teacher effectiveness*, are widespread.² However, political, economic and industrial issues surrounding *educational effectiveness* are sensitive, despite the level of non-partisan political consensus (at least in Australia) regarding the macro and micro economic importance of *teacher quality* and *quality teaching* for equipping students adequately to meet the constantly changing demands the modern workplace (e.g., Bishop, 2007; Macklin, 2006; Nelson, 2002, 2004).

The global economic, technological and social changes under way, requiring responses from an increasingly skilled workforce, make high quality educational provision an imperative – especially high quality teaching. Although OECD education ministers have committed their

¹ Correspondence related to this paper should be directed to: Dr Lawrence Ingvarson, Principal Research Fellow, ACER, Private Bag 55, Camberwell, VIC 3124 (Email: ingvarson@acer.edu.au); or Dr Ken Rowe, Research Director, *Learning Processes* research program, ACER, Private Bag 55, Camberwell, VIC 3124 (Email: rowek@acer.edu.au).

² For example, see: Access Economics (2005); Alton-Lee (2002, 2005); Curtis and Keeves (2000); Fenstermacher and Richardson (2005); Hanushek (1971, 1986, 2004); Ingvarson and Kleinhenz (2006a-c); Kleinhenz and Ingvarson (2004); Marsh, Rowe and Martin (2002); OECD (2005, 2006); Rowe (2001, 2004a); Rowe and Stephanou (2003).

countries to the goal of raising the quality of learning for all, this ambitious goal will not be achieved unless all learners, irrespective of their characteristics, backgrounds and locations, receive high-quality teaching (OECD, 2001, 2005). Since teachers are the most valuable resource available to both schools and higher education institutions in the realisation of this goal, an investment in *teacher quality* and on-going professionalism is vital. In our view, this goal can only be realised by ensuring that teachers are equipped with subject-matter knowledge and an evidence- and standards-based repertoire of pedagogical skills that are demonstrably effective in meeting the developmental and learning needs of all students for whom they have responsibility – regardless of students' backgrounds and intake characteristics, and whether or not they experience learning difficulties.³

Despite the emphasis placed on the importance of *teacher quality* and *quality teaching* in recent OECD publications, as well as similar emphases underlying the *2001 No Child Left Behind Act* in the USA (see: Center on Education Policy, 2003; LaTrice-Hill, 2002; US Department of Education, 2002), the bulk of international scholarly discourse concerned with *educational effectiveness* has largely ignored the importance of specifying standards of *instructional effectiveness* and their evaluation for teacher registration, accreditation, and on-going professional development. With few exceptions, especially from the related school effectiveness research literature (e.g., Mortimore, 1991; Reynolds, Creemers *et al.*, 2002), discussions that focus on the constituent elements of *teacher quality* in terms of what teachers should *know* and be able to *do* (i.e., *instructional effectiveness*, or the *what* and *how* of quality teaching), are conspicuous by their absence.⁴ Rather, the dominant emphasis continues to be characterized by offerings advocating *structural* changes for systemic reform, including curriculum reconstruction, single-sex schooling, class size (see Hattie, 2005b) etc., that have a long and not-so-distinguished history of rarely penetrating the classroom door.

A note about methodological limitations endemic to econometric research focussing on the link between *teacher quality* and student academic performance is appropriate here (e.g., Hanushek, 1971, 2004; Monk, 1992; Podgursky, Monroe & Watson, 2004; Rivkin, Hanushek & Kain, 2001). Since these limitations are well established, they need not be reiterated here in detail.⁵ In brief, however, an extensive body of work indicates that the single-level econometric models typically fitted to the available data employing *general linear model* (GLM) techniques under ordinary-least-squares estimation, are inappropriate on at least two counts. First, they fail to conceptualise, measure and evaluate *teacher quality* in terms of *what teachers know* (subject-matter knowledge) and *can do* (pedagogical competence). Second, such models rarely account for the measurement, distributional and structural properties of the data for response and explanatory variables – oversights that all too frequently yield misleading interpretations of findings for both policy and practice.

Failures to account for the inherent hierarchical structure of the data are particularly problematic. Findings from fitting explanatory multilevel models to relevant data (at the student, class/teacher, and school levels) consistently indicate that in excess of 40 percent of the residual variance in measures of student performance (adjusted for students' background and intake characteristics) is at the class/teacher-level (see citations given in footnote 5; and for key findings from meta-analytic syntheses of more than 500,000 evidence-based studies, see Hattie,

³ See: Coltheart and Prior (2007); Darling-Hammond and Bransford (2005); Farkota (2005); Hattie (1987, 2003, 2005a); Hoad, Munro *et al.* (2005); Purdie and Ellis (2005); Rowe (2005a,b, 2006a); Slavin (2005); Stronge (2002); Westwood (2006); Wheldall (2006).

⁴ For examples of exceptions, see: Bond, Smith *et al.* (2000); Darling-Hammond and Baratz-Snowden (2005); Darling-Hammond and Bransford (2005); Fenstermacher and Richardson (2005); Fullan, Hill and Crévola (2006); Ingvarson (2001); Ingvarson and Kleinhenz (2006a-c); Rowe (2002; in press a,b).

⁵ For relevant examples, see: Goldstein (1997, 2003); Goldstein and Spiegelhalter (1996); Hill and Rowe (1996, 1998); Millmann (1997); Raudenbush and Bryk. (1988); Raudenbush and Willms (1991, 1995); Rowe (2000, 2004b, 2006b, 2007); Rowe and Hill (1998).

2003, 2005a). These findings are especially useful. By identifying that the major sources of residual variation in students' learning and achievement progress are at the class/teacher level, they assist in specifying and evaluating *teacher quality* in terms of what *quality teachers* should *know* and be able to *do*. Moreover, such findings constitute invaluable data for informed, evidence-base content of pre-service teacher education and subsequent in-service professional development (Ingvarson, 2003; Rowe KS, Pollard & Rowe KJ, 2005), as well as for the specification and evaluation of *teaching standards* (Ingvarson & Kleinhenz, 2006a-c).

Rather than focussing on the *economics* of *teacher quality, per se* (well documented by other contributors to this conference), the present paper stresses the need for policies and processes designed to improve *teacher quality* through building teacher capacity, including the need for valid methods of specifying and evaluating *teacher quality*, as well as teaching standards. Whereas such policies and processes have universal applicability, this paper focuses on the urgent need for the adoption of these policies and procedures throughout Australian education systems.

The need for valid methods of assessing *teacher quality*

Pronouncements on the importance of *teacher quality* to student learning outcomes usually recognise the need to place greater value on teaching if the profession is to attract and retain high quality graduates from schools and universities (DEST, 2003; Ramsey, 2000). The major argument of this paper is that we will find it difficult to place greater value on teaching in substantive ways, such as better salaries and career paths for accomplished teachers, unless we greatly improve the capacity of the profession to define, evaluate and certify high quality teaching. (For a detailed review of national and international approaches to evaluating and rewarding accomplished teaching, see Ingvarson, Elliot *et al.*, 2006a).

Policies with respect to *teacher quality* fall into two main groups – policies designed to affect the *composition* of the teacher workforce, and policies designed to improve the *capacity* of individual teachers. Strategies in both areas are obviously important. Australia shares the problem of attracting and retaining a necessary share of the best graduates from schools and universities (OECD, 2001, 2005a). A recent synthesis of research on attitudes to teaching as a career found that extrinsic factors such as remuneration, workload, employment conditions and status were the most significant factors influencing able graduates not to choose teaching, and to leave the profession (DEST, 2006). If the ability of the teaching profession to compete with other occupations for the best graduates is to increase, research indicates that teaching salaries relative to those in related professions is the most importance factor (e.g., Dolton *et al.*, 2001), especially relative salaries after ten to fifteen years in the job.

This paper focuses mainly on policies related to improving *teacher quality* through building capacity, though it is recognised that these two strategies overlap. Policies designed to improve career paths and rewards for good teaching, for example, may aim to affect both *composition* and *capacity* if rewards are linked to evidence of appropriate professional development. Measures of *composition* typically focus on demographic data such as SES, TER scores, and GPAs. Measures of *capacity* focus on data about what teachers *know* and *do* in schools and classrooms.

Why do we need better methods for measuring *teacher quality*? The 2006 edition of the OECD's report, *Education at a Glance* (OECD, 2006), indicates that whereas the average ratio of the salary at the top of the incremental scale is 1.70, it is only 1.47 in Australia. It is nearly 3 in Korea and Japan. The typical salary scale for teachers in Australia does not place high value on evidence of teacher quality. Consequently, it is a weak instrument for improving student achievement. It does not provide incentives for professional development nor reward evidence of attaining high standards of performance. This ratio seems unlikely to improve unless further salary increments are linked to evidence of enhanced teacher knowledge and skill. Thirteen of 32 OECD countries report that they adjust the base salary of teachers on the basis of outstanding

performance in teaching, or successful completion of professional development activities. Australia is not one of them.

While progression to the top of the salary ladder is rapid in Australia – it takes only 9 years for most Australian teachers to reach the top of the scale compared with 24 years on average in OECD countries – there are no further career stages based on evidence of attaining higher levels of teaching standards. The implicit message in most Australian salary scales is that teachers are not expected to improve their performance after nine years. We suggest that the profession needs clearer guidelines as to what it expects its members to get better at with experience. Indeed, the salary scale provides few incentives for continued development of expertise in teaching. The relationship between evidence of professional development and salary progression is weak in teaching.

A recent survey of public opinion about teacher quality in the USA found that all groups recognised the importance of teacher quality and strongly support reforms that will lead to significant increases in teacher salaries, if those reforms will also provide better guarantees that these increases reward evidence of professional development and quality teaching (Hart & Teeter, 2002). Public attitudes in Australia are probably not much different. Guarantees of quality teaching, however, will be meaningless without valid methods of measuring teacher performance. Nonetheless, there has been renewed discussion about performance-based pay in Australia as a means of placing greater value on teaching. A forthcoming review of research in this area by the Australian Council for Educational Research (ACER), commissioned by the Australian Government Department of Education, Science and Training (DEST), indicates that the reason for so many failed merit pay schemes over the past thirty years has been the lack of understanding about the complexity of developing valid and professionally credible methods for gathering data about teaching and assessing teacher performance (Ingvarson & Chadbourne, forthcoming).

Unlike most other professions, the teaching profession has found it difficult to create a strong market for highly accomplished practitioners. A major reason for this is that the profession has yet to develop a voluntary system for providing certification to teachers who attain high standards of performance, at least one that employing authorities find credible and useful (Ingvarson & Kleinhenz, 2006c). There are many highly accomplished teachers, but no profession-wide system by which they can gain a highly respected and portable certification of their accomplishments. Consequently, incentives for teachers to provide evidence of professional development through stages of increasing expertise are weak.

Despite the paucity of incentives, there are strong indications that many in the profession wish to move down this path. A stronger market for highly accomplished teachers may be critical in areas of teacher shortage. This is partly why the Australian Science Teachers Association and the Australian Association of Mathematics Teachers have developed their own standards for highly accomplished teachers in recent years (Brinkworth, 2006; Semple & Ingvarson, 2006). School systems within Australia are also looking for better ways to recognise and retain good teachers, such as Western Australia with its Level 3 Classroom Teacher scheme.

A recent ACER review on performance pay found evidence that there is a stronger demand – in the sense of a greater capacity to offer over-award payments – for highly accomplished teachers in independent schools. The NSW Association of Independent Schools is introducing a system of remuneration based on increasing levels of professional standards (Newcombe, 2007). This applies at the entry level as well. This year (2007), all graduates of the highly selective Graduate Diploma of Education for secondary teachers from the University of Western Australia (UWA) accepted positions in non-government schools.

Other major related challenges are to ensure greater equity in the distribution of highly accomplished teachers across schools and school systems. At present we know that out-of-field teaching is more likely to be found in rural, remote and disadvantaged schools, but we do not

know how equitable the distribution of quality teachers is across schools. Without valid measures of *teacher quality*, we cannot conduct research on the contribution that variation in *teacher quality* might make to Australia's comparatively high levels of variation in student learning outcomes in schools for students drawn from high to low socioeconomic status (SES) backgrounds.

Effective teacher education is essential to *teacher quality*. A recent ACER study conducted for Teaching Australia (Ingvarson, Elliot *et al.*, 2006b) examined current procedures for the assessment and accreditation of teacher education courses. The findings indicated that these procedures are generally weak as quality assurance mechanisms. None is based on outcome measures of the quality of graduates or their competencies. There are over 200 teacher education courses in Australia, but, apart from one ACER study (Ingvarson *et al.*, 2005), we know little about the relative effectiveness of these courses. Clearly, there is a need to develop much better measures of the outcomes of teacher education courses if we are to understand the characteristics of courses that are more effective in producing competent teachers. ACER is currently coordinating an international study in 15 countries comparing the effectiveness of programs for preparing teachers of mathematics. This study includes the development of survey instruments that include measures of mathematical and pedagogical knowledge, which may enhance our capacity to measure the outcomes of teacher education course outcomes. (Further details can be found at <http://teds.educ.msu.edu/default.asp>).

Registration of new teachers is another important mechanism for ensuring *teacher quality*. Ideally, registration provides an assurance that new teachers are not only qualified but competent, but this is not the case in most states and territories. In most Australian States and Territories, registration follows automatically from completing an approved university qualification, despite the fact that this qualification alone is an uncertain guide to a teacher's capacity to promote learning in real school contexts (Parliament of Victoria, Education and Training Committee, 2005). Most professions delay registration until a period of internship in workplace settings has been completed satisfactorily (Ingvarson *et al.*, 2006b).

The Victorian Institute of Teaching has introduced new standards-based assessment procedures for provisional registration, which means that registration now depends on successful completion of a period of provisional registration supported by a mentor. By the end of this period, graduate teachers are expected to provide evidence that their practice has met standards of performance established by the VIT before gaining full entry to the profession. These new procedures are perceived as valid assessments against the VIT standards (Ingvarson *et al.*, 2007). Other states such as NSW are developing similar procedures. The success of these new procedures in promoting better teacher education and professional learning during induction will depend on the development of valid standards and measures of teacher performance.

The foregoing indicates several reasons why it is important to improve our capacity to measure teacher quality in ways that are valid, reliable and fair. The focus of this paper is on recent developments in standards-based approaches to measuring teacher performance designed to address these purposes. In summary, these purposes include:

- Accreditation of teacher education courses;
- Registration of new teachers; and
- Certification of accomplished and highly accomplished teachers.

These purposes constitute the three key quality assurance mechanisms in any profession. They provide the answers to the following questions: 'Who gets the right to train teachers?' 'Who gets to enter the profession?' and, 'Who gains recognition for attaining high standards of practice?' If the rhetoric about improving and valuing *teacher quality* is to become a reality, these three fundamental quality assurance functions need to be operating effectively – functions that are best carried out at the national or profession-wide level.

With some rare exceptions, however, there is little recent or current evidence to suggest that these mechanisms are operating effectively in Australia. This should be taken as a description of the current situation, not a criticism of any particular group. This paper is based on the proposition that, to carry out these functions more effectively, we need to develop more rigorous methods of assessing teacher quality. Paradoxical though it may seem, more rigorous methods of summative assessment lead to better planning and formative assessment in teacher education and professional development (Ingvarson *et al.*, 2006a).

If we are to develop methods for evaluating *teacher quality* for purposes such as outlined above, we need strong conceptual foundations for what we mean by *teacher quality*. The remainder of this paper focuses on methods for evaluating *teacher quality* for the purposes of developing a profession-wide system for identifying and recognising highly accomplished teachers.

Conceptualising *quality* in teaching

The guiding questions for this section of the paper are: “How do we develop valid indicators of teacher quality for purposes such as those above?” and “How do we decide what teachers should know and be able to do?” A closely related question is, “On what bases should teachers be evaluated?” Another is, “for what is it fair to hold teachers accountable?” These are questions that apply to all professions, as we have seen with respect to medicine in the television series MDA.

On what foundations should teachers be evaluated?

If measures of *teacher quality* are to be used in making decisions that are critical to teachers’ lives and careers, they should be based on valid criteria or defensible foundations. There is a long tradition of research on teacher evaluation issues in the USA. Millman and Darling-Hammond (1990) provide one of the most comprehensive reviews in their *Handbook of Research on Teacher Evaluation*. Based on the work of Michael Scriven (e.g. Scriven, 1994), Wheeler (1994) provides a helpful classification of foundations or sources that have been used in the US for developing criteria for evaluating teachers, together with comments on their relative validity. These include:

- Government regulations and requirements;
- Professional standards;
- Outcomes of teaching;
- Theories grounded in practice;
- What teachers are doing;
- What others would like teachers to be doing; and
- What teachers should be doing.

The Appendix to this paper provides an elaboration of each of these sources. Each provides a way of answering the question, ‘how will we determine what teachers should know and be able to do?’ Each aims to provide a source for criteria to be used in determining the domains of performance and attributes that should be included in a system for evaluating *teacher quality*.

Scriven (1994) and Wheeler (1994) weigh the arguments for and against, using each of these sources as a basis for evaluating teachers. They argue that, for employer purposes, such as performance management and decisions about retaining employment, the appropriate basis for evaluation of teachers is the last item, namely, what teachers should be doing, based on the duties and responsibilities of a teacher as should be delineated in an employment contract. However, for professional purposes beyond a single employer, such as registration and advanced certification, a more appropriate basis for assessing teacher quality is what the profession says teachers should *know* and be able to do – as specified in a set of professional standards.

Quality teaching

It is important to note that the purposes for defining and measuring *teacher quality* above all relate to “high stakes” decisions. As in other professions, legal issues will arise when teachers believe that measures of their professional performance do not have a sound basis (Hopkins, 2007). Methods of defining *teacher quality* need to have a sound and defensible conceptual basis, especially if they are used in quality assurance decisions such as registration, employment, promotion and professional certification.

Many have tackled these complex questions over the years. There is insufficient space here for a thorough review of the extensive literature on the various approaches to conceptualising *teacher quality*. Research on the characteristics of effective teachers and teaching has been conducted over the past 100 years and is well documented in a series of Handbooks of Research on Teaching and on Teacher Education (e.g., Richardson, 2005). Researchers have conceptualised *teacher quality* in diverse ways over this time, including personality characteristics, teacher behaviours (as in process-product research) and more recently in terms of what effective teachers know and do, where the guiding research questions include, ‘What knowledge is essential for teaching? (e.g., Shulman, 1987), and, ‘What is the nature of expertise in teaching? (e.g., Berliner, 1992). Recent research programs such as Shulman’s Teacher Assessment Program (Shulman, 1991) have paved the way for new approaches to defining quality teaching and developing teaching standards. These have drawn attention to the complexity of what effective teachers know about what they teach and how they help students to learn. As a consequence of this research, standards are emerging as a sound basis for defining levels of expertise in teaching and assessing teacher performance.

Fenstermacher and Richardson (2005) make a distinction between quality teaching and successful teaching that is useful to the present discussion, especially if measures of *teacher quality* are to be used for high-stakes decisions affecting teachers’ careers or salaries. They remind us that quality teaching is about more than whether something is taught. It is also about ‘how it is taught’ (p. 189). Successful teaching in the former sense may not be good teaching in the latter sense. Teaching is undeniably a moral enterprise. Similarly, what counts as “performance” varies. For some, the main indicators of performance should be measures of student outcomes, based on standardised tests of student achievement. This is what Fenstermacher and Richardson (2005) refer to as “successful teaching”, as follows:

By successful teaching we mean that the learner actually acquires, to some reasonable and acceptable level of proficiency, what the teacher is engaged in teaching (p. 191).

For others, the evidence of a teacher’s performance should be based on observations of the quality of opportunities they provide for student learning in their classrooms in relation to teaching standards. This is what Fenstermacher and Richardson call “good teaching”:

By good teaching we mean that the content taught accords with disciplinary standards of adequacy and completeness, and that the methods employed are age appropriate, morally defensible, and undertaken with the intention of enhancing the learner’s competence with respect to the content studied (p. 191).

This distinction points to two different approaches to conceptualising *teacher quality* – and two different views on what teachers should be held accountable for: one in terms of student achievement on standardised tests, the other in terms of the quality of opportunities for learning that teachers establish in their classrooms. The purpose of teaching standards, as we shall see below, is to capture what is meant by *good teaching* and to explicate what teachers need to *know* and be able to *do*, to establish quality opportunities for student learning.

Conceptualising teacher quality in terms of student achievement

Although it seems plausible to use student learning outcomes as a measure of “good teaching” and a basis for measuring teacher quality, the direct relationship between good teaching and learning outcomes is uncertain. The relationship between the two is far from a simple 1:1 causal

relationship. Fenstermacher and Richardson (2005, p. 190) point out that successful teaching, as defined above, depends not only on good teaching, but on three other conditions as well:

1. willingness and effort by the learner;
2. a social surround supportive of teaching and learning; and
3. opportunity to teach and learn.

Good teaching is only one of the ingredients necessary for successful teaching: a teacher may be “good” while being “unsuccessful” in certain contexts. While it may be reasonable to hold teachers accountable for good teaching in the sense above, there will be problems in evaluating teachers and holding them accountable using measures of successful teaching, since the latter depends also on conditions being in place for which others are accountable.

There have been significant developments in attempts to use student achievement as a measure of *teacher quality*. Millman (1997) reviewed four of these schemes in the USA, each using different kinds of student assessment. Two of them used “value-added” models for isolating and estimating school and teacher effects; the Tennessee Value Added Assessment System (TVAAS) and the Dallas Value-Added Accountability System. Proponents of these schemes claim that they are able to separate the effects of teachers and schools from the effects of other important factors such as family background. These two schemes are used, along with a range of other sources of information, to examine patterns of performance and to provide, for example, an indication of teachers who require professional development. While, these two schemes are not linked to salaries or bonuses, Pennsylvania has recently drafted a bill that proposes to use student achievement results to evaluate and reward administrators and teachers under the Federal Government’s Teacher Incentive Program.

The consensus among those who are closely familiar with these schemes is that they do not provide, and are unlikely to provide, a valid basis for high-stakes decision-making about the quality of teaching, such as those involved in performance-related pay (Braun, 2005; Kupermintz, 2002; McCaffrey *et al.*, 2003; Raudenbush, 2004). Some experts in educational measurement regard schemes such as the TVAAS as flawed because they use national norm-referenced tests that are usually insensitive to detecting the effects of teachers “instructional efforts” (Popham, 1997, p. 270). A danger with such schemes is that they may use student assessment data for a purpose that was not initially intended. That is, they may use students' scores on a nationally standardized test to assess the performance of a teacher when the test scores have not been validated for the latter purpose. Such tests are usually designed to discriminate between students, not teachers. In a recent review of the literature on the use of value-added modeling (VAM) in estimating teacher effects, McCaffrey *et al.* (2006) conclude:

... VAM-based rankings of teachers are highly unstable, and that only large differences in estimated impact are likely to be detectable given the effects of sampling error and other sources of uncertainty. Interpretations of differences among teachers based on VAM estimates should be made with extreme caution (p. 113).

The reliability of value-added estimates depends on the quality of the student achievement measures that underpin them, and the margins of error in most existing measures need to be understood. In addition, measures available so far are limited mainly to reading and numeracy in the primary years. For most subjects in the primary and secondary curriculum there are no measures to which value-added modelling could be applied.

There are two further reasons why state-wide measures of student outcomes are inappropriate as measures of individual teacher quality for high-stakes decision-making. First, they do not measure all that teachers are trying to achieve (Bond *et al.*, 2000; pp. 60, 63). Second, they do not provide useful information for teachers about what they need to know and be able to do to teach more effectively (Darling-Hammond, 1992).

Standards as a basis for measures of teacher quality

Teacher quality, for purposes such as those outlined in the introduction, is more appropriately conceived in terms of Fenstermacher and Richardson's concept of "good" teaching:

Quality teaching ... is about more than whether something is taught. It is also about how it is taught. Not only must the content be appropriate, proper, and aimed at some worthy purpose, the methods employed have to be morally defensible and grounded in shared conceptions of reasonableness. To sharpen the contrast with successful teaching, we will call teaching that accords with high standards for subject matter content and methods of practice 'good teaching. Good teaching is teaching that comports with morally defensible and rationally sound principles of instructional practice. Successful teaching is teaching that yields the intended learning. (Fenstermacher & Richardson, 2005; p. 189).

It would be tempting, say these writers, to conclude that 'quality teaching' is some kind of simple combination of 'good' and 'successful' teaching. But that argument is 'fraught with complexities':

There is currently a considerable focus on quality teaching, much of it rooted in the presumption that the improvement of teaching is a key element in improving student learning. We believe that this policy focus rests on a naïve conception of the relationship between teaching and learning. This conception treats the relationship as a straightforward causal connection, such that if it could be perfected, it could then be sustained under almost any conditions, including poverty, vast linguistic, racial or cultural differences, and massive differences in the opportunity factors of time, facilities, and resources. Our analysis suggests that this presumption of simple causality is more than naïve; it is wrong (Fenstermacher & Richardson, 2005; p. 205).

The writers of this paper conclude that appraisal of quality teaching is strongly interpretative and requires high levels of discernment on the part of the evaluators:

The vital insight is that when making a judgement of quality, one is always engaged in an interpretation – in a selection of one set of factors or indices over another, in attention to some dimensions of the phenomenon over other possible dimensions, in desiring and valuing some features of the task or the achievement more than other features (Fenstermacher & Richardson, 2005; p. 206).

The major implication of this discussion for the measurement of teaching quality is that measures of quality should focus on the quality of the opportunities for learning that teachers are providing for their students. One of the main aims of developers of teaching standards is to articulate 'sound principles of instructional practice' and what teachers should *know* and be able to *do*, to provide quality of the opportunities for learning.

Developing standards-based measures of teacher quality

Defining teaching standards

Dictionaries give two inter-related uses of the term "standard": to *rally*, as around the banner, or flag (standard); and to *measure*. Both definitions apply to the development of standards for teaching. In the first sense, standards articulate professional principles and values. Like the flag on ancient battlefields, they can provide a rallying point. A full set of teaching standards should provide a vision of good teaching and quality learning to guide the development of standards in the second sense.

Standards are also measures, as indicated by the second definition. They are tools we use constantly in making judgements in many areas of life and work, whether measuring length, evaluating writing, critiquing restaurants, or measuring professional performance. Standards provide the necessary context of shared meanings and values for fair, reliable and useful judgements to be made. Measures are one of humankind's most powerful inventions and have been the basis for significant improvement in most areas of human endeavour.

Writers of teaching standards need to articulate a vision of quality learning that will guide their more detailed work of describing what teachers should *know*, believe and be able to *do*.

Reaching a consensus is a necessary part of standards development, but it is a consensus that must be justified in terms of research and the wisdom of expert practitioners. This means that teachers who develop teaching standards must reach agreement on the scope and the content of their work and the underlying principles.

Developing teaching standards

When standards are used in assessing teaching performance, for purposes such as registration, accountability, promotion or certification by a professional body, there are three essential steps in their development. These are:

- Defining what is to be assessed (often called *content* standards);
- Developing methods for gathering evidence about teaching for assessment; and
- Setting performance standards (evaluating teaching).

As illustrated in Figure 1 below, these standards need to be embedded in a set of core values and a guiding educational vision.

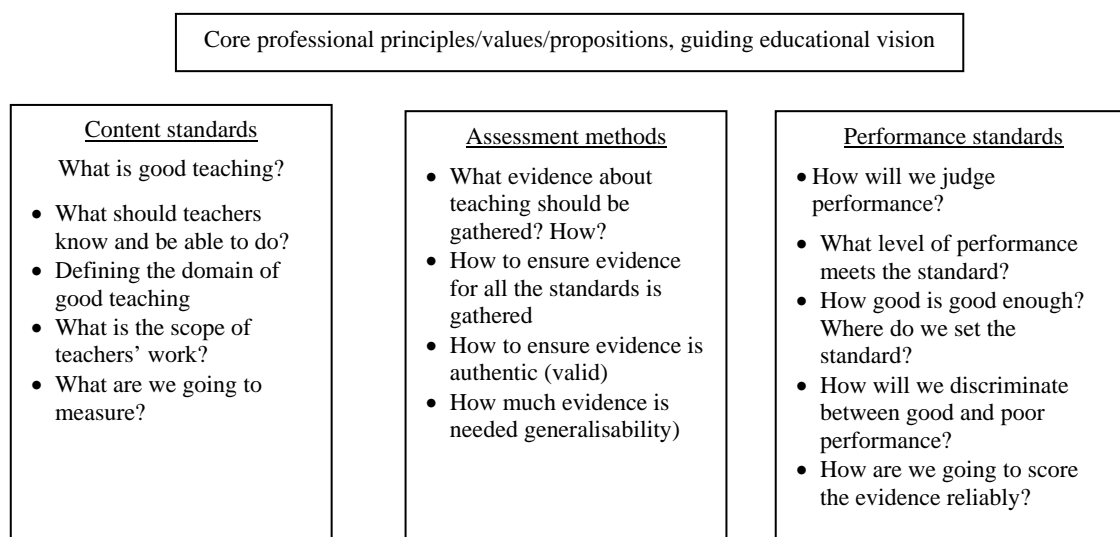


Figure 1. Performance-based teaching standards: Main components

The remainder of this paper follows the framework as set out in Figure 1, examining in turn content standards, assessment methods and the setting of performance standards in measuring teacher quality.

Trends in the development of teaching standards

1. They are developed by teachers themselves through their professional associations.
2. They aim to capture substantive knowledge about teaching and learning – what teachers really need to know and be able to do to promote learning of important subject matter.
3. They are performance-based. They describe what teachers should know and be able to do rather than listing courses that teachers should take in order to be awarded registration or certification.
4. They conceive of teachers' work as the application of expertise and values to non-routine tasks. Assessment strategies need to be capable of capturing teachers' reasoned judgements and what they actually do in authentic teaching situations.
5. Assessment of performance in the light of teaching standards is becoming one of the primary tools for on-going professional learning and development.

Sykes' and Plastrik (1993) define a standard as 'a tool for rendering appropriately precise the making of judgements and decisions in a context of shared meanings and values'. This is a

useful reminder that a set of standards needs all three components listed above. A full set of standards points not only to what will be measured, but also to how evidence about capability and performance will be gathered, and how judgments will be made about whether the standards have been met. There are only a few examples of teaching standards in Australia currently that are complete in this sense and useful, therefore, for measuring teacher quality. Examples include the standards developed by AAMT, ASTA and the WA Education Department's Level 3 Classroom teacher standards. Among international developments, the most highly regarded standards for measuring highly accomplished teaching are those developed by the National Board for Professional Teaching Standards (NBPTS; Available at: www.nbpts.org).

Characteristics of well-written standards

Following is an extract from one standard from the set of standards for accomplished teachers developed by the Australian Science Teachers Association (2002):

Accomplished teachers of science engage students in scientific inquiry. . . Their teaching reflects both the excitement and challenge of scientific endeavour and its distinctive rigour. They both teach and model practices that allow their students to approach knowledge and experiences critically, recognise problems, ask questions and pose solutions. They actively involve students in a wide range of scientific investigations . . . (p. 18).

Several features of a standard such as this are noteworthy. The first is that it points to a large, meaningful and significant "chunk" of a science teacher's work – it is an example of the challenging educational aims they are trying to achieve. It is not a micro-level competency, or a personality trait. Science teachers readily identify this type of standard as referring to an authentic (i.e. valid) example of the kind of work they do (or aspire to do). The second is that the standard is context-free, in the sense that it describes a practice that most agree accomplished science teachers should follow no matter where the school is. By definition, a professional standard applies to all contexts in which teachers work (which is not to say context does not affect practice). No matter where a school is, engaging students in scientific inquiry is likely to be regarded as a core responsibility of science teachers.

The third feature is that the standard is non-prescriptive about how to engage students in "doing science" and "thinking scientifically"; it does not standardise practice or force teachers into some kind of straightjacket. There are many ways to engage students in scientific enquiry. While the standard identifies an essential element of good science teaching, it does not prescribe how the standard is to be met. In this way, the standard also allows for diversity and innovation. Teachers are invited to show how they meet this standard; how they engage students in scientific enquiry. The fourth feature is that, as a standard, it points to something that is measurable, or observable. It is possible to imagine the kinds of evidence that a science teacher will assemble over time to show that they meet the standard, such as samples of students' work or videotape segments over time provided by the teacher.

These features apply to standards in all teaching fields, whether primary or secondary. In summary, using science teaching still only as an example, good standards for teachers should:

- be grounded in clear guiding conceptions of what it means to do (e.g. science);
- be valid; that is, represent what (science teachers) need to know and do to promote quality learning opportunities for students to learn (science);
- identify the unique features of what (science teachers) *know* and *do*;
- delineate the main dimensions of development the profession expects of a teacher of (science) – what (science teachers) should get better at over time, with adequate opportunities for professional development; and
- be assessable; that is, point to potentially observable features and actions.

Recent research on the validity of teaching standards developed by teachers indicates that the profession is building a stronger capacity to develop content standards that meet these criteria. The NBPTS standards, for example, provide examples of standards in 26 separate levels and

fields of teaching that meet these criteria. They also provide elaborations of what the standards mean, that reflect the complexity of what good teachers' know and do. (The NBPTS website list the extensive research conducted on the measurement characteristics of its standards certification procedures).

Methods for measuring teacher quality against the standards

The National Board for Professional Teaching Standards in the USA provides an example of a fully functioning system for providing certification that teachers have attained high standards of performance. Internationally, this is the only system for measuring teacher quality that has been subjected to extensive research on the validity, reliability generalisability of its methods for assessing *teacher quality*.

The National Board for Professional Teaching Standards (NBPTS)

The NBPTS was formed in 1987 to advance the quality of teaching and learning in the USA by developing professional standards for accomplished teaching, creating a voluntary system to certify teachers who meet those standards and integrating certified teachers into educational reform efforts. It is an independent, non-profit, non-partisan and non-governmental national organization with a broad membership base that includes practising teachers, state governors, school administrators, teacher unions, school board leaders, college and university officials, business executives, foundations and concerned citizens.

Most states and a growing number of districts in the USA now offer extra rewards, including annual bonuses and higher salaries to encourage teachers to apply for National Board Certification. There is a growing market for National Board Certified Teachers. Carefully trained peer teachers, who have already demonstrated accomplishment in their field of teaching, carry out the assessment of teachers' performance under NBPTS supervision. History teachers evaluate history teachers, early childhood teachers evaluate other early childhood teachers, and so on. NBPTS certification processes ensure that teachers are evaluated by those with an in-depth knowledge of what is being evaluated. This encourages teachers' confidence in the validity and fairness of the processes.

The NBPTS approach to assessing teacher quality

Below is an outline of a typical set of NBPTS teaching standards, in this case, standards for highly accomplished science teachers (the NBPTS website provides the full version). It is only one of 26 sets developed in various teaching fields. Noteworthy, as in this typical example, is that each set of standards seeks to define, not only what is in common with other fields, but also what is unique about what teachers know and do in that field of teaching.

Domain 1: Preparing the way for productive student learning

- Understanding students
- Knowledge of science
- Instructional Resources

Domain 2: Establishing a favourable context for learning

- Engagement
- Learning environment
- Equitable participation

Domain 3: Advancing student learning

- Science inquiry
- Expanding fundamental understandings
- Contexts of science

Domain 4: Supporting teaching and learning

- Assessment

- Family and community outreach
- Contributing to the profession
- Reflective practice.

As with each set of NBPTS standards, these standards were developed by a national committee of expert teachers and researchers in the relevant field of teaching. Once established, the task of developing the methods of assessment for each set of NBPTS standards is handed to independent Assessment Development Teams consisting of other expert (science) teachers and specialists in educational measurement.

The NBPTS approach to measuring teacher quality relies on teachers providing two types of evidence. The first is a portfolio containing four “entries”. Three are classroom exercises, one based on samples of student work, two based on videotapes of classroom practices and one based on documented contributions to the profession and school community outside the classroom. Following are three examples of portfolio entries:

Entry 1: Designing Science Instruction

Teachers are asked to choose three activities from an instructional sequence and work samples from two students that demonstrate how they link instructional activities together to promote students’ understanding of one important scientific concept along with the development of one or more related process skills.

Entry 2: Probing Student Understanding

Teachers are asked to submit a 20-minute Videotape of a lesson in which they introduce an important idea in science, and demonstrate how they use classroom discourse and questioning to elicit students’ initial conceptions of an important idea in science, and how they use their understanding to influence their instruction. Optional Instructional Artefacts may also be submitted.

Entry 3: Inquiry Through Investigation

Teachers are asked to submit a 20-minute Videotape of a lesson in which they conduct an investigation of an important scientific concept and demonstrate how they support students in a scientific inquiry discussion as they interpret data that have been collected during the course of the investigation. Any Instructional Artefacts used by the students may also be submitted.

For the second ‘entry’ method of assessment, teachers attend an ‘Assessment Centre’ for three hours where they respond to six exercises on-line designed to gather evidence about their subject matter knowledge and pedagogical content knowledge. This mode of assessment gathers evidence that can not be covered well through the portfolio entries. Below is an example of one of the six assessment centre exercises for teachers applying for NBPTS certification that assesses a teacher’s knowledge about helping students to learn science.

Assessment Centre Exercise 4 – Misconceptions

- *Focus:* This exercise focuses on candidates’ ability to recognise student misconceptions and to appropriately address them through subsequent instruction.
- *Prompts:* Candidates are asked to identify the misconception(s) in a piece of student work, to develop the next lesson to address the misconception, and to develop an assessment to judge whether the student’s understanding has changed following instruction.

Points to note about these two methods for gathering evidence about teacher quality developed by the NBPTS, the portfolio entry and the assessment centre exercise:

- the tasks are authentic and, therefore, complex;
- the tasks are open-ended, allowing teachers to show their own practice;
- the tasks provide ample opportunity and encouragement for analysis and reflection;

- subject-matter knowledge underlies all performances;
- the tasks encourage teachers to exemplify good practice;
- each task assesses a cluster of standards; and
- each standard is assessed by more than one task.

In endeavoring to provide a valid assessment of accomplished practice, the NBPTS has aimed to develop methods of assessment that:

- allow for the variety of forms sound practice takes;
- sample the range of ways teachers know their content; and
- provide appropriate contexts for assessments of teaching knowledge and skill.

The NBPTS assessment processes engage candidates in the activities of teaching – activities that require the display and use of teaching knowledge and skill and that provide teachers with the opportunity to explain and justify their actions.

Setting performance standards

As described above, candidates for NBPTS certification complete ten assessment tasks: four portfolio entries and six assessment centre exercises. This number helps to provide a guarantee that NBPTS certification is a reliable assessment of teacher quality. Each NBPTS task assesses a cluster of the Standards, and each standard is assessed by more than one task. This also helps to ensure the reliability of the assessment.

Assessors undertake a week's training and are only invited to continue with 'live' scoring in subsequent weeks if they reach a high level of consistency in scoring benchmark entries. Two scorers, using standards-based rubrics, independently assess each exercise until they consistently agree. This means that between 10 to 20 assessors may be involved in assessing a teacher's total application. A weighted total score is calculated across all ten exercises. Pairs of scorers assess only one exercise, they do not examine all of a candidate's work. A wide-ranging and thorough research program ensures the technical quality and integrity of the measurement processes.

Setting performance standards involves establishing processes for distinguishing between levels of performance. The NBPTS is the only example of a certification system for accomplished teachers to have made a serious attempt to ensure the psychometric quality of its standards setting processes. The Board initially used the Judgmental Policy Capturing procedure (Jaeger, 1982, 1995). More recently, it has used the less complex 'direct judgment' method. Both methods involved weighting and benchmarking exercises based on the judgment of panels of expert teachers. The NBPTS takes care to ensure the validity of its standards, the processes for developing the standards, and the validity of the assessment tasks and scoring rubrics, especially the congruence between the assessment tasks and the standards that are being assessed.

All National Board assessments have been subject to validation studies in which panels of expert teachers in the relevant certification areas are asked to respond to a series of questions about the relevance, representativeness, necessity and importance of the standards and assessment processes. The panels found that the exercises and scoring rubrics were appropriate for the content being assessed (Crocker, 1997).

Other validation exercises involved panellists of experienced teachers working in pairs, independently of the assessment panels ranking a sample of portfolio exercises and Assessment Centre exercises. When compared with the scores awarded by the original assessors, the panellists' assessments, with rare exceptions, demonstrated the accuracy and the consistency of the scoring system. (Jaeger, 1998). In another psychometric validation study Jaeger (1998), it was found that among the 258 candidates in the study, there was a 13% chance of misclassification, which is relatively low.

Validation studies of the NBPTS system for assessing teacher quality for professional certification

The NBPTS has long agonised over the question of whether the students of National Board Certified Teachers (NBCTs) perform better on external measures of achievement than applicants who do not gain certification. It has only been relatively recently that the Board has been able to claim that its certification is a valid indicator of teacher who are more effective. The following examples come from some of the most recent research that has been carried out in this contentious field.

One of the best known studies is from a project by Bond, Smith, Baker & Hattie (2000), where the researchers compared samples of student work from a group of students taught by teachers who gained certification with work samples from another group taught by teachers who did not. The results of this study found that NBCTs significantly outperformed their non NBCT colleagues on 11 out of 13 key dimensions of teaching expertise, and out-performed them on all 13 measures (Bond *et al.*, 2000).

More recently, Goldhaber and Anthony (2004) used outcomes data from standardized tests for students in the third, fourth and fifth grades in North Carolina, the state with the largest number of NBCTs in the USA. They examined data for the years 1996-1997 through 1998-1999 using multivariate analysis to compare the effects of NBCTs on student achievement in mathematics and reading with those of non-NBCTs. The students taught by the NBCTs performed better and showed more growth in performance than those taught by the non NBCTs. The researchers concluded that the NBPTS certification process is an effective means of identifying teachers of high quality (Goldhaber & Anthony, 2004).

I Vandervoort and his colleagues (Vandervoort *et al.*, 2004) compared the achievement data of the students of 35 NBCTs with those of non certified teachers in Arizona. In three quarters of the comparisons, the elementary school students of the NBCTs performed better in reading, language arts and mathematics than students of non NBCTs. The authors of this study concluded that:

The preponderance of the evidence suggests that students of NBPCTs achieve more (Vandervoort *et al.*, 2004; p. 36).

Evidence that NBCTs make a major contribution to successful students' learning continues to mount. The most recent study, conducted by Cavalluzo (2004), used data from a large urban school district – Miami-Dade Public Schools – to assess the contribution made by teachers' professional characteristics to student achievement in mathematics in the ninth and tenth grades. One of the strengths of the data set used was the detail regarding each student. In addition to standard demographic indicators, Cavalluzo and colleagues were able to control for a number of indicators of student motivation and performance that might influence student achievement. The study found that, when compared with students whose teachers had never been involved with National Board Certification, the achievements of students of NBCTs were higher:

After taking into account differences in the characteristics of their students, such comparisons show that students who had a typical NBC teacher, made the greatest gains, exceeding gains of those with similar teachers who had failed NBC or had never been involved in the process. Students with new teachers who lacked a regular state certification, and those who had teachers whose primary job assignment was not mathematics instruction made the smallest gains (Cavalluzo, 2004; p. 3).

From this work, it was concluded that:

In this study, (National Board Certification) proved to be an effective signal of teacher quality. Indeed, seven of nine indicators of teacher quality that were included in the analyses resulted in appropriately signed and statistically significant evidence of their influence on student outcomes. Among these indicators, having an in-subject teacher, NBC and regular state certification in high school mathematics had the greatest effects (Cavalluzo, 2004, p. 3).

A full list of independent research projects about the validity of the NBPT standards and certification procedures are available at: http://www.nbpts.org/research/research_archive.cfm.

Board certified teachers are in high demand and are often mentors and leaders in their schools. This is largely because members of the education and wider communities are confident that the Board's stringent efforts to ensure the rigour, fairness, validity and reliability of its assessments can be depended upon to provide credible guarantees of *teacher quality*. Board certified teachers are thus rewarded in terms of enhanced status and expanded employment opportunities as well as financial remuneration.

A study commissioned by the Board in 2001 sampled the views of 10,000 National Board Certified Teachers. This study found that teachers believed the certification process had:

- made them better teachers (92 per cent);
- was an effective professional development experience (96 per cent);
- enabled them to create better curricula (89 per cent);
- improved their ability to evaluate student learning (89 per cent); and
- enhanced their interaction with students (82 per cent), parents (82 per cent) and colleagues (80 per cent).

Typical comments included:

The National Board Certification process was by far the best professional development I have been involved in. I did not realise how much I still needed to learn about impacting student learning. I learned so much through hours of analysing and reflecting.

I gained valuable insight of myself as a teacher. The process helped me to assess my teaching abilities as no administrator could have. Most importantly, my students benefit from my self-improvement.

Working with other teachers in my school who were also working on certification was rewarding.

It was the hardest thing I have ever done and it is something I am so glad that I tried. I am immensely proud of the work I turned in – even if I did not make the needed grade. It has made me a better teacher and colleague.

By 2006 nearly 120,000 teachers had applied for National Board Certification and around 45 per cent had been successful. Many who miss out the first time apply again. The application fee for National Board Certification is about \$US2,500. This may seem expensive, but it is much less than the costs of a Masters degree. A recent independent study of relative costs of different approaches to professional development by Cohen and Rice (2005) found that:

the candidacy process and candidate support programs . . . incorporate elements of high-quality professional development identified in the research literature and are no more costly than other forms of professional development. . . Our findings on design and cost suggest policy makers should consider the NBC model as an alternative way to target professional development and salary rewards.

Completing an NBPTS portfolio takes at least twelve months. The portfolio tasks engage applicants in challenging, site based learning that centres on gathering, analysing and reflecting on evidence of their students' and their impact on that learning. Tasks were designed to be vehicles for professional learning. There is considerable evidence that teachers who have been through the National Board system regard the experience as one of the most powerful professional experiences they have ever had (Tracz & Associates, 1995).

Concluding Comments

A recent publication of The Education Trust in the USA by Haycock (2004) was titled, "The Real Value of Teachers: If good teachers matter, why don't we act like it." The evidence described and outlined in this paper (and growing evidence from Australian professional associations such as ASTA and the AAMT), indicates it is not because of a lack of capacity to

measure *teacher quality*. This paper indicates that the profession can define good teaching in all the specialist fields of teaching, including early childhood, primary teaching, as well as secondary teaching. It can gather valid evidence of good teaching, and it can assess that evidence validly and reliably.

The capacity to develop standards and credible methods for assessing teacher performance is growing, but more investment needed to translate this capacity into viable systems for registration and advanced certification. Australia needs a major research program focused on developing better methods for assessing *teacher quality*. This paper began by listing a number of reasons why we need better methods for assessing teacher quality. The need is clear. Policies aimed at improving salaries, lifting the attractiveness of teaching as a career, the quality of teacher education, and the effectiveness of professional learning, will amount to little without guarantees that they are linked to valid and reliable measures of better quality teaching. Without better methods for evaluating teaching, it will be difficult to ask the public to place greater value on it.

Above all, given the social and economic importance of *teacher quality* and *quality teaching* at both national and individual levels, our teachers and their students require no less. Further, since teachers are the most valuable resource available to schools and higher education institutions, there is a crucial need for a substantive and methodological refocus of the prevailing economic *teacher-quality/student-performance/merit-pay* research and policy agenda to one that focuses on the need for capacity building in teacher professionalism in terms of what teachers *know* and can *do* via the specification and evaluation of quality teaching standards.

References

- Access Economics (2005). *Review of higher education outcome performance indicators*. Canberra, ACT: Australian Government Department of Education, Science and Training.
- Alton-Lee, A. (2002). *Quality teaching: Impact of teachers and schools on outcomes*. Wellington, NZ: Ministry of Education.
- Alton-Lee, A. (2005). *Overview of key messages from quality teaching for diverse students in schooling: Best evidence synthesis*. Wellington, New Zealand: Ministry of Education. Available for download at: www.minedu.govt.nz/goto/bestevidencesynthesis/S.
- Australian Science Teachers Association (2002). *National professional standards for highly accomplished teachers of science*. Canberra, ACT: Australian Science Teachers Association.
- Berliner, D. (1992). The nature of expertise in teaching. In F.K.Oser, A. Dick, & J. Patry (Eds), *Effective and responsible teaching: The new synthesis* (pp 227-249). San Francisco, CA: Jossey-Bass.
- Bishop, J. (2007, February). 'Education is a key driver of economic prosperity' *The Australian*, 2 February 2007, p. 14.
- Bond, L., Smith, T., Baker, W., & Hattie, J.A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro, NC: Center for Educational Research and Evaluation.
- Braun, H.I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Policy Information Centre, Educational Testing Service. Princeton, NJ: Educational Testing Service.
- Brinkworth, P. (2004). *AAMT teaching standards assessment evaluation project 2004*. Canberra, ACT: Quality Schooling Branch, Department of Education, Science and Training.
- Cavalluzzo, L. (2004). *Is national board certification an effective signal of teacher quality?* Washington DC: The CNA Corporation.
- Center on Education Policy (2003). *State and federal efforts to implement the No Child Left Behind Act*. Washington, DC: Author.
- Cohen, C.E. & Rice, J.K. (2005). *National Board Certification as Professional Development: Design and Cost*. Washington D.C.: The Finance Project www.financeproject.org
- Coltheart, M., & Prior, M. (2007). *Learning to read in Australia*. Occasional Paper 1/2007 (Policy Paper #6). Canberra, ACT: The Academy of the Social Sciences in Australia. Available for download at: <http://www.assa.edu.au/>.

-
- Curtis, D.D., & Keeves, J.P. (2000). The Course Experience Questionnaire as an institutional performance indicator. *International Education Journal*, 1(2), 73-82.
- Darling-Hammond, L. (1992). Creating standards of practice and delivery for learner-centred schools. *Stanford Law and Policy Review*, 4, 37-52.
- Darling-Hammond, L., & Baratz-Snowden, J. (Eds.) (2005). *A good teacher in every classroom: Preparing the highly qualified teachers our children deserve*. San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L., & Bransford, J. (Eds.) (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L. & Youngs, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us? *Educational Researcher*, 9(3), 13-25.
- DEST (2003). *Australia's teachers: Australia's future*. Canberra, ACT: Australian Government Department of Education, Science and Training.
- DEST (2006) Attitudes to teaching as a career. Canberra, ACT: Australian Government Department of Education, Science and Training.
- Dolton, P., Chevalier, A. & McIntosh, S. (2001). *Recruiting and retaining teachers in the UK: An analysis of graduate occupation choice from the 1960s to the 1990s*. London: Department of Education and Science.
- Fenstermacher, G.D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186-213.
- Fullan, M., Hill, P.W., & Crévola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8(4), 369-395.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd edn.). London: Hodder-Arnold.
- Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. With discussion. *Journal of the Royal Statistical Society, A*, 159(3), 385-443.
- Hanushek (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, 61(2), 280-288.
- Hanushek, E.A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.
- Hanushek, E.A. (2004). *Some simple analytics of school quality*. Background paper to keynote address presented at the *Making Schools Better Summit Conference*, Melbourne Business School, the University of Melbourne, 26-27 August 2004.
- Hart, P.D. & Teeter, M. (2002). *A national priority: Americans speak on teacher quality*. Princeton, NJ: Educational Testing Service.
- Hattie, J.A. (1987). Identifying the salient facets of a model of student learning: A synthesis of meta-analyses. *International Journal of Educational Research*, 11(2), 187-212.
- Hattie, J.A. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36, 5-13.
- Hattie, J.A. (2003, October). *Teachers make a difference: What is the research evidence?* Background paper to invited address presented at the 2003 ACER Research Conference, Carlton Crest Hotel, Melbourne, Australia, October 19-21, 2003. Available at: <http://www.acer.edu.au/documents/TeachersMakeaDifferenceHattie.doc>.
- Hattie, J.A. (2005a). What is the nature of evidence that makes a difference to learning? *Research Conference 2005 Proceedings* (pp. 11-21). Camberwell, VIC: Australian Council for Educational Research. Available at: <http://www.acer.edu.au>.
- Hattie, J.A. (2005b). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43(6), 387-425.
- Hill, P.W., & Rowe, K.J. (1996). Multilevel modeling in school effectiveness research. *School Effectiveness and School Improvement* (Leading article) 7(1), 1-34.
- Hill, P.W., & Rowe, K.J. (1998). Modeling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement*, 9(3), 310-333.
- Hoad, K-A., Munro, J., Pearn, C., Rowe, K.S., & Rowe, K.J. (2005). *Working Out What Works (WOWW) Training and Resource Manual: A teacher professional development program designed to support teachers to improve literacy and numeracy outcomes for students with learning difficulties*. Canberra,

-
- ACT: Australian Government Department of Education, Science and Training; and Australian Council for Educational Research.
- Ingvarson, L.C. (Ed.) (2001). *Assessing teachers for professional certification: The National Board for Professional Teaching Standards*. Amsterdam, the Netherlands: Elsevier Press.
- Ingvarson, L. (2003). A professional development system fit for a profession. In V. Zbar and T. Mackay (Eds.), *Leading the education debate: Selected papers from a decade of the IARTV Seminar Series* (pp. 391-408). Melbourne, VIC: Incorporated Association of Registered Teachers of Victoria (IARTV).
- Ingvarson, L.C., Beavis, A., Danielson, C., Ellis, L. & Elliott, A. (2005). *An evaluation of the Bachelor of Learning Management at Central Queensland University*. Canberra, ACT: Australian Government Department of Education, Science and Training. Available for download in PDF format at: http://www.acer.edu.au/research/documents/BLM_280905.pdf.
- Ingvarson, L.C., & Chadbourne, R. (Eds.), *Valuing Teachers' Work*. Camberwell, VIC.: Australian Council for Educational Research (forthcoming).
- Ingvarson, L.C., Elliot, A., Kleinhenz, E., & McKenzie, P. (2006a). *Teacher education accreditation: A review of national and international trends and practices*. Canberra, ACT: Teaching Australia (Australian Institute for Teaching and School Leadership Ltd).
- Ingvarson, L. Elliott, A. Kleinhenz, E. & McKenzie, P. (2006b). *Accreditation of teacher Education: A Review of national and international trends and practices in other professions*. Report prepared for Teaching Australia (Australian Institute for Teaching and School Leadership Ltd). Available at: <http://www.teachingaustralia.edu.au/ta/go/home/projects/teacheraccreditation>
- Ingvarson, L.C. & Hattie, J. (Eds.). (in press). *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards*. Amsterdam, the Netherlands: Elsevier Press.
- Ingvarson, L.C., & Kleinhenz, E. (2006a). *Advanced teaching standards and certification: A review of national and international developments*. Report to Teaching Australia (Australian Institute for Teaching and School Leadership): <http://www.teachingaustralia.edu.au/ta/go/home/projects/standards>.
- Ingvarson, L.C., & Kleinhenz, E. (2006b). *Standards for advance teaching: A review of national and international developments*. Canberra, ACT: Teaching Australia (Australian Institute for Teaching and School Leadership Ltd.).
- Ingvarson, L.C. & Kleinhenz, E. (2006c). *A Standards-Guided Professional Learning System*. Melbourne, VIC: Centre for Strategic Education. Available at: www.cse.edu.au.
- Ingvarson, L.C., Kleinhenz, E., Khoo, S.T. & Wilkinson, J. (2007). *The VIT Program for Supporting Provisionally Registered Teachers: Evaluation of implementation in 2005*. Melbourne, VIC: Victorian Institute for Teaching.
- Jaeger, R.M. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis*, 4(4), 461-475.
- Jaeger, R.M. (1995). Setting performance standards through two-stage judgmental policy capturing. *Applied measurement in education*, 8(1), 15-40.
- Jaeger, R.M. (1998). Evaluating the psychometric qualities of the National Board of Professional Teaching Standards' assessments: A methodological accounting. *Journal of Personnel Evaluation in Education*, 12(2), 189-210.
- Kleinhenz, E., & Ingvarson, L.C. (2004). Teacher accountability in Australia: Current policies and practices and their relation to the improvement of teaching and learning. *Research Papers in Education*, 19(1), 31-49.
- Kupermintz, H. (2002). Teacher effects as a measure of teacher effectiveness: Construct validity considerations in TVAAS (Tennessee Value-Added Assessment System). *Centre of the Study of Evaluation Technical Report 563*. Los Angeles, CA: National Centre for Research on Evaluation, University of California.
- LaTrice-Hill, T. (2002). *No Child Left Behind Policy Brief: Teaching quality*. Denver, CO: Education Commission of the States. Available at: <http://www.ecs.org/clearinghouse/>.
- Macklin, J. (2006, October). *Teaching standards: Recognising and rewarding quality teaching in public schools*. Australian Labor Party. Available at: www.alp.org.au.
- Marsh, H.W., Rowe, K.J., & Martin, A. (2002). PhD students' evaluations of research supervision: Issues, complexities and challenges in a nationwide Australian experiment in benchmarking universities (Leading article). *Journal of Higher Education*, 73(2), 313-348.

-
- McCaffrey, D., Lockwood, J.R., Koretz, D.M., & Hamilton, L.S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: Rand Corporation.
- Millmann, J. (Ed.) (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc.
- Monk, D.H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Education Evaluation and Policy Analysis*, 14, 307-332.
- Mortimore, P. (1991). School effectiveness research: Which way at the crossroads? *School Effectiveness and School Improvement*, 2(3), 213-229.
- National Board for Professional Teaching Standards (2001). *Early Childhood /Generalist Standards (for teachers of students ages 3-8)*. 2nd ed. www.nbpts.org
- National Board for Professional Teaching Standards (2005). *Framework of National Board Standards and Certificates*. Available at: www.nbpt.org/standards/stds.
- Nelson, B. (2002). *Quality teaching a national priority*: Media Release, 4 April 2002, MIN 42/02. Available from: http://www.dest.gov.au/ministers/nelson/apr02/n42_040402.htm.
- Nelson B (2004). *New Carrick Institute for Learning and Teaching in Higher Education*: Media Release, 11 August 2004: MIN 851/04. Available at: <http://www.dest.gov.au/Ministers/Media/Nelson/2004/08/n851110804.asp>.
- Newcombe, G. (2006, November). Letter to *Teacher: The National Education Magazine*. Camberwell, VIC: Australian Council for Educational Research.
- OECD (2001). *Teachers for tomorrow's schools: Analysis of the World Education Indicators, 2001 edition*. Paris: Organisation for Economic Cooperation and Development and UNESCO Institute for Statistics.
- OECD (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2006). *Education at a glance: OECD indicators 2006*. Paris: Organisation for Economic Cooperation and Development.
- Podgursky, M., Monroe, R., & Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review*, 23(5), 507-518.
- Parliament of Victoria, Education and Training Committee (2005). *Step Up, Step In, Step Out: Report on the inquiry into the suitability of pre-service teacher training in Victoria*. Melbourne: Parliament of Victoria, Education and Training
- Popham, W.J. (1997). The moth and the flame: Student learning as a criterion of instructional competence. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* (pp 264-274), Thousand Oaks, CA: Corwin Press, Inc.
- Purdie, N., & Ellis, L. (2005). *A review of the empirical evidence identifying effective interventions and teaching practices for students with learning difficulties in Years 4, 5 and 6*. A report prepared for the Australian Government Department of Education, Science and Training. Camberwell, VIC: Australian Council for Educational Research. Available for download in PDF format at: <http://www.acer.edu.au/research/programs/documents/literaturereview.pdf>.
- Ramsey, G. (2000). *Quality matters – revitalising teaching: Critical times, critical choices*. Report of the Review of Teacher Education. Sydney, NSW: New South Wales Department of Education and Training.
- Raudenbush, S.W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioural Statistics*, 29(1), 121-129.
- Raudenbush, S.W., & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.), *Review of Research in Education 1988-1989, Vol. 15* (pp. 423-475). Washington, DC: American Educational Research Association.
- Raudenbush, S.W., & Willms, J.D. (Eds.). (1991). *Schools, Classrooms and Pupils: International Studies of Schooling from a Multilevel Perspective*. New York: Academic Press.
- Raudenbush, S.W., & Willms, J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (Eds.) (2002). *World class schools: International perspectives on school effectiveness*. London: Routledge-Falmer.
- Richardson, V. (Ed.). (2001). *Handbook of research on teaching* (4th edn.). Washington: American Educational Research Association.

-
- Rivkin, S.G., Hanusheck, E.A., & Kain, J.F. (2001). *Teachers, schools, and academic achievement*. Working Paper No. 6691. Washington, DC: National Bureau of Economic Research.
- Rowe, K. J. (2000). Assessment, league tables and school effectiveness: Consider the issues and let's get real! *Journal of Educational Enquiry*, 1(1), 72-97.
- Rowe, K.J. (2001). Educational performance indicators. In M. Forster, G.N. Masters and K.J. Rowe, Measuring learning outcomes: Options and challenges in evaluation and performance monitoring (pp. 2-20). *Strategic Choices for Educational Reform; Module IV – Evaluation and Performance Monitoring*. Washington, DC: The World Bank Institute.
- Rowe, K.J. (2002). The importance of teacher quality. *Issue Analysis*, No. 22, February 27, 2002. Sydney, NSW: Centre for Independent Studies; available at: <http://www.cis.org.au>
- Rowe, K.J. (2004a). *The importance of teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision – implications of findings from the international and Australian evidence-based research*. Background paper to invited address presented at the *Making Schools Better Summit Conference*, Melbourne Business School, the University of Melbourne, 26-27 August 2004. Available for download in PDF format at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2004b). *Analysing and reporting performance indicator data: 'Caress' the data and user beware!* Background paper to invited address presented at the 2004 Public Sector Performance & Reporting Conference (under the auspices of the International Institute for Research – IIR), Sydney, 19-22 April 2004. Available at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (Chair) (2005a). *Teaching reading literature review: A review of the evidence-based research literature on approaches to the teaching of literacy, particularly those that are effective in assisting students with reading difficulties*. A report of the Committee for the National Inquiry into the Teaching of Literacy. Canberra, ACT: Australian Government Department of Education, Science and Training. Available at: <http://www.dest.gov.au/nitl/report.htm>.
- Rowe, K.J. (Chair) (2005b). *Teaching reading: Report and recommendations*. Report of the Committee for the National Inquiry into the Teaching of Literacy. Canberra, ACT: Australian Government Department of Education, Science and Training. Available at: <http://www.dest.gov.au/nitl/report.htm>.
- Rowe, K.J. (2006a). *Effective teaching practices for students with and without learning difficulties: Constructivism as a legitimate theory of learning AND of teaching? Background paper to keynote address presented at the NSW DET Office of Schools Portfolio Forum, Wilkins Gallery, Sydney, 14 July 2006*. Available at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2006b). *School performance: Australian State/Territory comparisons of students' achievements in national and international studies*. Camberwell, VIC: Australian Council for Educational Research. Available at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Rowe, K.J. (2007). *Practical multilevel analysis with MLwiN & LISREL: An integrated course* (6th edition, revised). 23rd ACSPRI Summer Program in Social Research Methods and Research Technology, Australian National University, 15-19 January 2007. Camberwell, VIC: Australian Council for Educational Research.
- Rowe, K.J. (in press, a). School and teacher effectiveness: Implications of findings from evidence-based research on teaching and teacher quality. In A. Townsend and B. Caldwell (Eds.), *Building on the past to chart the future: A critical review of research, policy and practice in school effectiveness and improvement* (Vol 2, Chapter 41). New York: Springer.
- Rowe, K.J. (in press, b). Educational effectiveness: The importance of evidence-based teaching practices for the provision of quality teaching and learning standards. In D.M. McInerney (Ed.), *Research on Sociocultural Influences on Motivation and Learning (Volume 7, Standards in Education)*. Greenwich, Conn: Information Age Publishing.
- Rowe, K.J., & Hill, P.W. (1998). Modeling educational effectiveness in classrooms: The use of multilevel structural equations to model students' progress. *Educational Research and Evaluation*, 4(4), 307-347.
- Rowe, K.J., & Stephanou, A. (2003). *Performance audit of literacy standards in Victorian Government schools, 1996-2002*. A consultancy report to the Victorian Auditor General's Office. Melbourne, VIC: Australian Council for Educational Research.
- Rowe, K.S., Pollard, J., & Rowe, K.J. (2005). *Literacy, behaviour and auditory processing: Does teacher professional development make a difference?* Background paper to Rue Wright Memorial Award presentation at the 2005 Royal Australasian College of Physicians Scientific Meeting, Wellington,

- New Zealand, 8-11 May 2005. Available for download in PDF format at: <http://www.acer.edu.au/research/programs/learningprocess.html>.
- Semple, A. & Ingvarson, L.C. (2006). How can professional standards improve the quality of teaching and learning science? Conference Proceedings, ACER Research Conference 2006 *Boosting science Learning – what will it take?* (pp. 42-48). Camberwell, VIC: Australian Council for Educational Research. Available at: <http://www.acer.edu.au/workshops/conferences.html#past>.
- Scriven, M. (1994). Using the Duties-Based Approach to Teacher Appraisal. In L.C. Ingvarson and R. Chadbourne (Eds.), *Valuing Teachers' Work*. Camberwell, VIC.: Australian Council for Educational Research (forthcoming).
- Shulman, L.S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Education Review*, 57, 1-22.
- Shulman, L.S. (1991). Final Report of the Teacher Assessment Project. Palo Alto: Stanford University.
- Slavin, R.E. (2005). *Evidence-based reform: Advancing the education of students at risk*. Report prepared for Renewing Our Schools, Securing Our Future: A National Task Force on Public Education (A joint initiative of the Center for American Progress and the Institute for America's Future). Available at: <http://www.americanprogress.org/site/>.
- Stronge, J.H. (2002). *Qualities of effective teachers*. Alexandria, VI: Association for Supervision and Curriculum Development.
- Sykes, G., & Plastrik, P. (1993). *Standard setting as educational reform*. Washington DC: American Association of Colleges for Teachers of Education.
- Tracz S., & Associates. (1995). Improvement in teaching skills: Perspective from national board for professional teaching standards field test network candidates, *Annual Meeting Educational Research Association*. San Francisco.
- US Department of Education (2002). *No Child Left Behind: A desktop reference*. Washington, DC: Author. Available at: www.ed.gov/offices/OESE/reference.
- Vandervoort, L.G., Amerin-Beardsley, A., & Berliner, D. (2004). National board certified teachers and their students' achievement. *Educational Policy Analysis Archives*, 12(26).
- Wheeler, P.H. (1994). *Foundations upon which to build a teacher evaluation system* (TEMP D Memo 18). Kalamazoo, MI: Western Michigan University, The Evaluation Centre, Centre for Research on Educational Accountability and Teacher Evaluation.

Appendix

FOUNDATIONS OF MEASURES FOR EVALUATING TEACHERS

On what foundations should teachers be evaluated? If measures of *teacher quality* are to be used in making decisions that are critical to teachers' lives and careers, it is clear they must be based on valid criteria or defensible foundations. Wheeler (1994, pp. 3-4) provides a helpful classification of foundations or sources that have been used in the US for developing criteria for evaluating teachers, together with comments on their relative validity. Each provides a way of answering the question, 'how will we determine what teachers should know and be able to do?' Each provides a source for criteria to be used in determining the domains of performance and attributes to be covered by the standards:

Government regulations and requirements. This category covers state and federal laws, codes, and program guidelines. Examples are complying with safety codes for the handling and storage of chemicals; implementing categorical program requirements such as involving of parents of Chapter 1 [Disadvantaged] students in their educational program; following the state curriculum frameworks; using district adopted textbooks; and administering tests in accordance with specified procedures.

Professional standards. Specific examples of this category are (1) the professional standards for teaching mathematics developed by the National Council of Teachers of Mathematics [See Case 1 in this report]; (2) the standards for teacher competence in the educational assessment of students developed by the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association; and (3) the standards of the National Board for Professional Teaching Standards. Such professional standards can be helpful in developing a local teacher evaluation system. However, they may be narrowly focussed, may reflect the interests of the association, and may or may not be relevant to the local context.

Outcomes of teaching. Examples of outcomes are student assessment results, number and types of disciplinary referrals, implementation of skills learned in a training program, and amount of resources used. Such evaluation systems assume that promoting the attainment of those outcomes covered by the evaluation system is the primary function of the teacher. These systems can drive teaching behaviour rather than promote diverse teaching practices and curricula content for different teachers and students. They can also be constraining for teachers confronted with challenging situations and students with extensive behaviour problems, and it can be impossible to obtain valid and reliable assessment data for some students (e.g., disabled, non-English speaking, and highly mobile).

Theories grounded in practice. Theories of teaching, of learning and cognition, of the cognitive psychology of teaching, and of the cognitive development of teachers are examples of foundations in this category. However, theories are attempts to provide explanations of phenomena and are not, by themselves, adequate as foundations for systems to evaluate teachers.

What teachers are doing. Potential foundations in this category look at what teachers are doing and use the results of such efforts to build a teacher evaluation system. One type of study looks at effective and, in some cases, ineffective teachers, and identifies the practices and behaviours associated with these teachers (also called effective teaching research, or process-product studies). Another type of study looks at what teachers are doing (job analysis). A third is based on the consensus of practitioners concerning what they actually do as part of their teaching job. A fourth is based on what teachers at a particular school have been doing in the past and are expected to continue doing, that is, the norms of the school. All of these assume that what some

teachers are doing is a good approach for others in the profession of teaching, a questionable assumption that can lead to an invalid system (Scriven, 1994).

What others would like teachers to be doing. Examples of these include the use of certain teaching styles (e.g. cooperative learning groups, whole language instruction), preferences of peers and supervisors, and desires of clients and stakeholders (e.g. students, parents, future employers of students, community members). A foundation based on the styles, preferences and desires of others is clearly invalid, whether the approaches work well for an individual teacher or not.

What teachers should be doing. The duties and responsibilities of a teacher, as designated by the local school board, the superintendent and principal, and the state education agency, form the seventh type of foundation. Criteria and performance indicators derived from a foundation of teacher duties and responsibilities often overlap with the first type of foundation (governmental regulations and requirements). Teachers must be fully informed as to what their duties and responsibilities are. This can be done through well-written and comprehensive job-descriptions or an employee handbook. In some cases, teachers in some subject areas or specific individuals will have additional duties and responsibilities not common to all teachers; they must be made fully aware of these if they are to be evaluated on the basis of how well they perform these duties and responsibilities.