March, 2011

Assessing Teachers for Professional Certification: Achieving National Consistency

Lawrence C Ingvarson, ACER

Available at: http://works.bepress.com/lawrence_ingvarson1/184/
On February 9 this year the Hon. Peter Garrett, the Minister for School Education, Early Childhood and Youth, welcomed the release of new National Professional Standards for Teachers. Garrett said "the Standards will drive future reform in the areas of national accreditation of teacher education programs, nationally consistent teacher registration, and the certification of highly accomplished teachers.”

The Australian Institute for Teaching and School Leadership (AITSL) has been charged with responsibility for these reforms. At the launch of the new National Standards, Tony Mackay, the Chair of AITSL, announced that:

“AITSL will be working in partnership with the various regulatory authorities, teacher unions, and all school sectors, to implement the standards in all jurisdictions. This work will include explaining and elaborating on the standards, providing support materials, providing the processes, and deciding on how judgements are made.”

This article focuses on the challenge of developing and implementing a nationally consistent and credible voluntary system for the certification of highly accomplished teachers. What might be involved and how might this be achieved?

Lawrence Ingvarson looks at the challenges involved in implementing the national standards.

An important first step is to recognise that teaching standards are not fully developed or defined until it is clear how they are to be used to judge teachers’ knowledge and practice. When standards are used for professional certification, there are three essential steps in their development.

These are:
1. Defining what is to be assessed – as an example, what do highly accomplished teachers know and do. (This is what the National Professional Standards for Teachers aim to do. These are often called content standards);
2. Developing valid and consistent methods for gathering evidence about what a teacher knows and is able to do in relation to the standards; and
3. Developing reliable procedures for assessing that evidence and deciding whether a teacher has met the standard. (This will depend on developing performance in addition to content standards).

Clearly we are only at the first stage in developing standards that can be used to assess teacher performance consistently.

The decathlon as an analogy
The Olympic Decathlon provides a good analogy of the steps involved in developing a respected standards-based system for judging performance.
The origins of the decathlon go back to early 20th Century when King Gustav V of Sweden told the American Jim Thorpe, “You, sir, are the World’s Greatest Athlete”.

People used to argue, apparently, about what makes a great all-round athlete, just as they still argue about what makes a good teacher. They realised the concept needed definition. What should all-round athletes be able to do? After a lot of debate they decided that the main elements of what constituted a great all-round athlete were strength, speed, stamina, endurance and perseverance. In other words, they defined what should be assessed – the content standards as it were – if one was to judge whether someone was a good all round athlete.

The next step was to reach agreement on how to judge the all-round athlete; how to assess strength, speed, etc. What should an athlete be asked to do to provide evidence that they are good all-round athlete? Ten events over two days were decided upon somehow, and the concept was thereby operationalised. On day one the events are 100 metres, long jump, shot put, high jump, and 400 metres. On day two, the events are 110 metres hurdles, discus, pole vault, javelin, 1500 metres.

A set of 10 events was seen as a sufficient sample of evidence on which to make judgments about an athlete’s overall ability as an all-round athlete. The events are carefully structured to cover each of the elements (i.e. the standards) more than once. Consequently, the evidence about each element (standard) is gathered in more than one way. This increases the reliability of the assessment.
Finally, they needed to set the performance standards, which meant reaching agreement on the level of performance in each event that met the standard. A performance standard is the level of performance on the criterion being assessed that is considered satisfactory in terms of the purpose of the evaluation. Another performance standard would be the level of performance on the criterion being assessed that is considered superior. So far as the decathlon is concerned, several “benchmark” levels have been set for each event. (For example, if you run 100 metres in 10.395 seconds you earn 1000 points, 10.827 seconds earns 900 points, 11.278 earns 800 points and 11.756 earns 700 points, and so on for each event).

The overall level of performance is determined by weighting and combining the performance across all events. Performance standards not only need to specify how well an athlete must do in each event to qualify; they need to specify how well they must do across all events on the average to be rated a good all round athlete. Athletes must participate in the same set of events; there is no choice. However, a good performance in one event can compensate for a poor performance in another.

How does this apply to teaching?
Similar steps apply to any system that aims to provide valid, reliable and fair assessments. While AITSL has completed the first step of defining, in broad terms, what is to be assessed, it recognises that there will be a need for the standards to be elaborated before they can be used to assess teacher knowledge and practice. For example, they will need to elaborate on what teachers in various fields are expected to know about what they teach and what their students are expected to learn.

Teacher associations have an important role to play here. National curriculum documents should go some way to meeting this need, but expert teachers and researchers with up-to-date knowledge of recent research in the various teaching areas will need to be called upon if the standards are to be valid and useful for assessment purposes – and credible to teachers. (What, for example, should an early year’s primary teacher know about diagnosing and dealing with problems in learning to read? What should a high school science teacher know about research on common misconceptions in science and how to deal with them? And so on.).

Nearly 20 teacher associations have developed their own standards and they want them to be used for certification purposes. This represents a tremendous resource that AITSL can draw upon.

The decathlon analogy illustrates the cardinal rules that apply to valid and fair evaluations of teaching; multiple, independent sources of evidence (assessment tasks) are needed for reliable judgments; and assessments should be conducted by multiple, independent assessors trained to interpret the evidence consistently.

The same fundamental issues arise in planning a certification system for highly accomplished teachers as for great all-round athletes.

Do the standards distinguish between highly accomplished and proficient teaching?
What types methods of assessment (c.f. events) provide valid evidence of meeting the standards?
How will we gain a representative sample of what a teacher knows and does in relation to the standards?
How will we ensure that the evidence covers all the standards, and in more than one way where possible?

How do we do this without placing an unreasonable burden on teachers?
How do we ensure that assessors are trained to high levels of consistency in interpreting the evidence and making judgments?
Should teachers meet a certain level on every standard, or should they be able to compensate for a poor performance on some standards with a good performance in others?

Ensuring reliable assessments of highly accomplished teaching
Table 1 extends the decathlon analogy by listing the elements of the National Professional Standards and suggesting a range of assessment methods that might be used to provide evidence against each standard. These methods fall into two main groups: those where the evidence is gathered independently of the teacher (e.g. classroom observation, test scores and ratings from student questionnaires).

Table 1 serves to illustrate how the assessment methods need to be chosen so that, in total, they provide several independent sources of evidence about each standard. Every standard needs to be assessable and to be assessed. It will also be important that assessment requirements for candidates are comparable across certification fields, such as highly primary generalist teaching and secondary art specialists. (This requirement can be greatly facilitated by the development of common assessment “shells”; i.e. portfolio tasks with a similar structure, but adapted to each certification field.)

Professional knowledge
Knowledge of content that a teacher is expected to teach and how to help students learn that content can be assessed directly and reliably by written assessments. Structured portfolio tasks, classroom observation and student questionnaires can also provide such evidence to a limited extent (as indicated by the smaller ticks). Structured portfolio tasks based
on planning and teaching a unit of work and analyses of student work over time can also provide evidence about teachers’ knowledge of their students.

**Professional practice**
Evidence about Professional practice can be gained by means of carefully structured portfolio tasks based, for example, on videotapes, examples of student work and assessments of student learning over time, and direct observation. Student questionnaires have proved to be a reliable source of evidence about the learning environment.

**Professional Engagement**
The Professional Engagement standards call for yet another type of evidence. Structured portfolio tasks, endorsed by principals or supervisors, in which, for example, teachers provide several brief case studies of their work with students’ families and community and link these to improved student learning have proved reliable. The typical CV rarely makes these links explicit.

There is evidence to indicate that assessors can be trained to make consistent judgments based on the evidence provided by each of the assessment methods in Table 1 (Ingvarson & Hattie, 2008). And, of course, the reliability of overall judgments about a teacher’s practice increases with the number of independent sources of evidence, as well as the fairness. Testing the validity of these judgments is a much more complex question.

Some have advocated value-added approaches to providing evidence about a teacher’s performance, based on gain scores on state or national standardised tests such as NAPLAN, but these are not included here. In the Australian context, it is not possible to use NAPLAN to provide reliable evidence about the performance of individual teachers, since many do not teach in the subjects or the year levels assessed by NAPLAN. Scaling up NAPLAN so that it was administered every year from grade 1 to grade 9, and so that it included more areas of the curriculum, would be very expensive and unlikely to gain support.

Even if this could be done, there is increasing concern about the reliability and validity of value-added models when used as the main basis for high stakes decisions, such as teacher dismissal or certification (Baker et al., 2010; Rothstein, 2011). Such methods certainly do not provide the “gold standard” by which other methods should be judged (although they could have useful role to play in research studies designed to test the validity of standards-based methods such as those in Table 1). There is evidence that new portfolio methods, in which teachers, using customised tests, are asked to provide direct evidence that their students have progressed as a result of their teaching, may be a better way to go.

While there is no one best way to assess teacher performance, a key consideration is whether the methods used have a significant effect on improving teachers’ practice. This is where standards-based methods for evaluating teaching have a distinct advantage; because they engage the teacher in the active role of describing, analysing and reflecting on their teaching and showing how they meet the standards in their particular context (Lustick, 2011; Brantlinger, et al. 2011).

---

**TABLE 1: Ensuring reliable assessment against the AITSL standards.**

<table>
<thead>
<tr>
<th>METHODS OF ASSESSMENT</th>
<th>Written assessments</th>
<th>Portfolio entries</th>
<th>Student evaluation</th>
<th>Classroom observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AITSL STANDARDS</td>
<td>Tests of pedagogical/content knowledge</td>
<td>Student work or learning based</td>
<td>Video based</td>
<td>Documented activities</td>
</tr>
<tr>
<td>Professional knowledge</td>
<td>Know students and how they learn.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Know the content and how to teach it.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Professional knowledge</td>
<td>Plan for and implement effective teaching and learning.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Create and maintain supportive and safe learning environments.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Assess, provide feedback and report on student learning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Professional engagement</td>
<td>Engage in professional learning.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Engage with colleagues, parents/carers and the community</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

---
CONFRONTING 21ST CENTURY INEQUALITIES: TOWARDS MORE JUST SCHOOLS AND SOCIAL SERVICES

Opening Address: The Hon Tanya Plibersek MP

Keynote: Raewyn Connell, The University of Sydney

SCHOOLS AND SOCIAL JUSTICE IN THE 21ST CENTURY

As long as we live in an unequal society, there will be questions about the connection between schools and social justice. Social science has gathered a wealth of knowledge about this issue, but the knowledge is not always used. At this moment, social justice seems to be an inconvenient question. Yet the major structures of social inequality – class, gender, colonization – are still present. These are not static, but living structures that constantly produce new effects in economy, culture, and social power – and all are intertwined with education. So are the processes that marginalize people on grounds of disability, sexuality, or ethnicity. In the 21st century we see inequalities produced in new social arenas. These include the trans-national arenas of the world economy, the effects of the neoliberal dominance of politics, new communication technologies and networks of social relations, and arenas of environmental injustice. Questions for teachers (and educational policy makers) arise across this spectrum. What would a socially just curriculum like it? How can we abolish institutional hierarchies in education? What is socially just practice in the contemporary classroom? Australian education will be profoundly shaped by the answers we give to these questions – or our failure to answer them.

Keynote: Ross Homel AO, Griffith University

THE PATHWAYS TO PREVENTION PROJECT

Children’s development is influenced by many contexts and institutional settings, including families, schools, childcare centres, churches, and ethnic community groups. Positive development is more likely when the developmental system that shapes these settings and their interconnections works in a manner that enriches children’s environments and their access to resources. The capacity to work across contexts and maintain integrated systems of support for development is particularly important for children who experience family adversity and various forms of social and economic disadvantage that are known risk factors for healthy development. The Pathways Project has evolved since 2001 as a comprehensive service offered through a partnership between national community service agency Mission Australia, seven local primary schools and Griffith University, in several ethnically diverse, socially disadvantaged, and high crime Brisbane suburbs. Within its universal focus, the Pathways model emphasises comprehensive and integrated practice that supports development in a holistic way.

Keynote: Peter Freebody, The University of Sydney

NOT BY SCHOOLING ALONE: SERVICE COLLABORATION IN INEQUITABLE TIMES

Current economic conditions and political forces have intensified economic inequalities and access to quality services and supports in Australia. At the same time fragile status of agenda relating to equality, equity and social justice in the service and education sectors has existed over the last three decades in Australia. This means that only highly proactive and coordinated efforts on the parts of researchers, policy makers and practitioners can re-establish a sense of the significance and value of equity as a set of public discourses, commitments and activities. Professor Freebody will use this backdrop to summarise and draw out the major themes arising from the day’s discussions. He will also draw on the extended, on-going OECD future-of-schooling scenarios to put the argument for a widespread, cross-agency, cross-community movement around equity, inclusion, and social justice. Such a movement is crucial at a time when social divisions seem to be intensifying, when young Australians, particularly those in poor, migrant, Indigenous and remote communities, face increasingly complex, uncertain and challenging circumstances.

Concurrent workshops and papers:

- Dorothy Bottrell, Kelly Freebody & Susan Goodwin, The University of Sydney
  SCHOOL–COMMUNITY ENGAGEMENT FOR EQUITY

- Jon Callow, University of Sydney & Geoff Munns, University of Western Sydney
  TEACHERS FOR A FAIR GO – ENGAGING STUDENTS IN POVERTY

- Margaret Clark, Australian College of Educators
  RECLAIMING THE LANGUAGE OF THE NEO-LIBERAL EDUCATION AGENDA

- Denise Lynch, The University of Sydney
  REFUGEE CHILDREN IN THE AUSTRALIAN CONTEXT
Where to now?
Developing assessment methods along the lines indicated in Table 1 is no simple matter. AITSL faces a formidable challenge in building a profession-wide and nationally consistent system for the assessment and certification of highly accomplished teachers (the same challenge applies to the registration of new teachers). Setting up such a system will be as complex as setting up a national system for assessing student achievement like NAPLAN, perhaps more so. High levels of expertise in educational measurement will be required if the system is to gain professional credibility and recognition by employers (Ingvarson & Hattie, 2008).

Table 1 helps to indicate the scale of work involved before we can be confident that the assessment methods cover all the National Standards. Some of these methods have been used in some jurisdictions to a limited extent, but little research has been conducted on their validity and reliability.

A well-funded research program is required that will include the development and trialling of a range of common assessment exercises. For example, it will need to provide convincing evidence that the assessment exercises are valid measures of the intentions of the standards writers.

It will need to show that the sample of evidence gathered is sufficient to make a reliable judgment about whether a teacher meets the standards; i.e. that one can generalise from the evidence to a teacher’s typical performance.

Guidelines for the assessment exercises will need careful preparation if they are to be interpreted consistently by teachers. The demands that they make on teachers across different teaching fields and levels will also need to be consistent. The feasibility for teachers of assembling the required evidence about their teaching will be another important consideration.

Once trials have been completed and decisions made about the assessment methods, several crucial steps remain in developing a nationally consistent system. These focus on the complex tasks of weighting the different forms of evidence and setting the standard. It calls for a common understanding of what will count as meeting the standard. It calls for the development of scoring rubrics based on the standards and the selection of benchmarks illustrating different levels of performance. It calls for carefully planned programs for training assessors to levels where they can make consistent judgements of this evidence across schools and jurisdictions.

The research program will need to complete these tasks before fundamental questions about the validity of the assessment system can be addressed: is it able to distinguish between good teachers and poor teachers? Is it able to distinguish between teachers who meet the standards and those who do not? Addressing these questions adequately will, in turn, require further high quality research, but they will need to be answered positively if the system is to gain the trust and respect from the profession that will be essential to its survival.

The challenge of national consistency calls for a national research program that brings together leading researchers, expert teachers, regulatory authorities, teacher associations and all school sectors with a common interest in developing a high quality national certification system and ensuring comparability across jurisdictions.

A vital question that remains is whether Australia should have one profession-wide certification system, as in other professions, or, whether operation of the system should be delegated to each jurisdiction, or even to each school sector. It is difficult to see how a nationally consistent system that provides a widely respected certification to highly accomplished teachers can emerge if each jurisdiction develops its own assessment methods, using AITSL’s standards only as a framework – just as it is difficult to imagine a decathlon standards that would gain respect if each country decided on its events and performance standards.

The public will only support significantly higher salaries for nationally certified highly accomplished teachers if it is convinced that the certification system is consistent as well as rigorous.

Lawrence Ingvarson is Principal Research Fellow at ACER and Fellow of ACE.

REFERENCES