

# InSights

## **Evidence of impact of teacher education programs: A focus on classroom observation**

A paper prepared for the Australian Institute  
of Teaching and School Leadership

Professor Jenny Gore  
The University of Newcastle, Australia  
June 2015



**Citation**

Gore, J 2015, *Evidence of impact of teacher education programs: A focus on classroom observations*, Australian Institute for Teaching and School Leadership, Melbourne.

The Australian Institute for Teaching and School Leadership (AITSL) was formed to provide national leadership for the Commonwealth, state and territory governments in promoting excellence in the profession of teaching and school leadership with funding provided by the Australian Government.

ISBN 978-1-925192-15-5

© 2015 Australian Institute for Teaching and School Leadership (AITSL).

AITSL owns the copyright in this publication. This publication or any part of it may be used freely only for non-profit education purposes provided the source is clearly acknowledged. The publication may not be sold or used for any other commercial purpose.

Other than permitted above or by the Copyright ACT 1968 (Commonwealth), no part of this publication may be reproduced, stored, published, performed, communicated or adapted, regardless of the form or means (electronic or otherwise), without prior written permission of the copyright owner.

Address inquiries regarding copyright to: AITSL, PO Box 299, Collins Street West, VIC 8007, Australia.

This project was funded by the Australian Institute for Teaching and School Leadership Limited (AITSL) with funding provided by the Australian Government.

# InSights

## **Evidence of impact of teacher education programs: A focus on classroom observation**

A paper prepared for the Australian Institute for Teaching and School Leadership

Professor Jenny Gore  
The University of Newcastle, Australia



# Contents

<b>Principles to underpin the re-accreditation system</b> .....	<b>1</b>
<b>Standard and 'classroom readiness'</b> .....	<b>2</b>
<b>Available evidence types</b> .....	<b>3</b>
<b>The case for an emphasis on classroom observations</b> .....	<b>5</b>
<b>On the use of classroom observations as evidence of program effectiveness</b> .....	<b>9</b>
<b>Summary</b> .....	<b>11</b>
<b>References</b> .....	<b>12</b>



The task set for this stimulus paper was to provide a perspective on evidence of impact within an Initial Teacher Education context, specifically addressing:

- The principles/framework that might underpin the selection of evidence of impact against the Graduate level of the *Australian Professional Standards for Teachers* (APST) for the purposes of accreditation of ITE programs
- The types and/or range of evidence that could be considered to demonstrate evidence of impact against the Graduate level of the APST.

As framing for this paper, I have outlined a brief set of principles to be considered in the development of a national approach to the provision of evidence. The remainder of the paper focuses on three issues. First, I consider the relationships among the *Accreditation of Initial Teacher Education Programs in Australia Standards and Procedures*, the *Australian Professional Standards for Teachers*, and the idea of 'classroom readiness.' Second, I outline an argument for using classroom observations as a key form of evidence of the impact of teacher education programs. Finally, I offer preliminary thoughts on how such observations might be used for making judgments about the quality of graduates and effectiveness of teacher education programs.

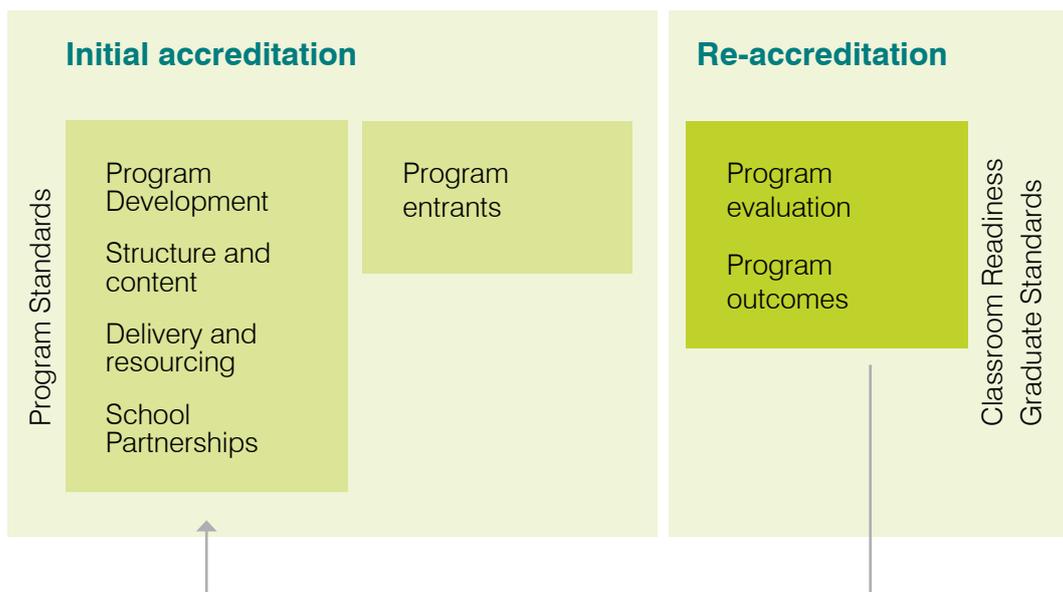
## Principles to underpin the re-accreditation system

The following principles, derived from my various experiences as teacher, teacher educator, researcher, and dean of education, attempt to articulate a viable and defensible approach to assessing the impact of teacher education programs while also mediating the diverse concerns of the various stakeholders who will necessarily be involved in its enactment.

1. While informing accountability, the major purpose of any national approach to gathering evidence of the impact of teacher education programs should be to provide feedback for those programs in the ongoing quest for delivery of high quality teacher education programs and preparation of high quality ('classroom ready') graduates and, ultimately, improve the quality of the teaching in Australian schools;
2. The approach must have reasonable levels of validity, reliability, and fairness, recognising both the desire for scientific integrity and the messy reality of the social worlds of schooling and teacher preparation;
3. The approach needs to have credibility with the profession and its various stakeholders. Without such credibility and buy-in, distrust among sectors of the education enterprise could be exacerbated with negative consequences for the status of teaching, which would potentially worsen current recruitment and workforce concerns;
4. The approach should not be cumbersome or overly burdensome for universities, systems or schools. Nor should it be overly rigid, locking institutions too tightly into agreed forms of evidence in ways that prevent refinement during an implementation phase;
5. The focus should be on the quality of teaching, not the quality of teachers. The focus on teachers, rather than teaching, in much government, policy, and academic argument, is misguided and damaging to the growth of the profession. It is the teaching (understood broadly) that matters most;
6. The focus here should be on the assessment of programs rather than on the assessment of individuals, which must remain the province of the teacher education provider.

## Standard and 'classroom readiness'

Three distinct but related constructs are at the centre of discussions about evidence of the impact of teacher education, as borne out in the TEMAG report: program standards, graduate standards, and classroom readiness. How these constructs, and the relationships among them, are understood will underpin the development of an approach to systematically gathering evidence. A preliminary mapping of their relationship is outlined below in Figure 1:



**Figure 1. Graphic representation of the relationship between Program Standards, Graduate Standards, and classroom readiness**

The Program Standards (outlined in the shaded rectangles) guide the accreditation of teacher education programs at both initial and re-accreditation phases. At re-accreditation, program evaluation and program outcomes become particularly pertinent. Program evaluation methods will contribute to accounting for how program outcomes are determined by the provider. Program outcomes are constituted primarily in terms of whether the Graduate level of the *Australian Professional Standards for Teachers* has been met and, in recent rhetoric, whether graduates are 'classroom ready.'

A key implication of this recent addition to the teacher education review discourse is that a graduate might meet the Graduate Standards (which the teacher education program has been accredited to deliver) but might not be 'classroom ready', might not be successful in the classroom, might not have "the practical skills needed" for the classroom (Teacher Education Ministerial Advisory Group, 2014). While in an ideal world they might be the same, it becomes clear that the Standards and being 'classroom ready' are not seen as synonymous. If this is the case, and if teacher educators, teachers, governments, and other stakeholders concur, and agree that both the Standards and classroom readiness are important, then forms of evidence are needed to demonstrate, at the program level,

that graduates can both meet the Graduate Standards and show their practical skill in the classroom. However, there can be no simple distinction between the Standards and being ready for the classroom. To a considerable extent, the Graduate Standards require demonstration of classroom skill. And demonstrating practical skill in the classroom requires understanding of multiple aspects of teaching and learning, as articulated (substantially) in the Graduate Standards.

A foundational consideration in developing an approach to the compilation and assessment of evidence of the effectiveness of teacher education programs is whether the Graduate Standards are taken as indicative of classroom readiness, or not. The simple fact that “classroom readiness” is now a key phrase in the latest push to improve teaching and teacher education signals their difference.

## Available evidence types

Major types of evidence currently in use for gauging the effectiveness of teacher education programs, and some of their associated challenges, are listed in Table 1. I have not systematically provided advantages or strengths of the first three types of evidence, although this would provide a more balanced view. Such balance is likely to come through the full set of stimulus papers. My goal in this short paper is to provoke, to stimulate. While each of these types of evidence could play a role in the assessment of teacher education program effectiveness, I argue that classroom observations offer the greatest potential and align best with the principles articulated above, as well as with the assessment of classroom readiness.

Of these forms of evidence, value-added measures (VAM) hold great appeal with many politicians, policy-makers, and psychometricians. However, there is mounting concern about their adequacy and their unintended consequences, particularly when used for teacher assessment in high-stakes contexts (e.g., Teachers College, Columbia University, 2015). Classroom observations, on the other hand, are increasingly favoured as a source of evidence of the quality of teaching and/or teachers. Observations are seen to have the following advantages:

- Consistency of the data (Goldring et al., 2015);
- Specificity of evidence in observation data (Goldring et al., 2015);
- They can provide both individualised (teacher) and institutional (provider) feedback data to facilitate improvement;
- Their timeliness, relevance, and transparency (Goldring et al., 2015);
- They provide a ‘bigger picture’ than can be provided by VAM (Goldring et al., 2015); and,
- They are considered to be fairer by principals (Goldring et al., 2015) and by teachers (Jiang et al., 2015).

Note that nearly every state in the US requires classroom observation as part of teacher evaluation (Doherty & Jacobs, 2013). Also, most places in the US that are using value-added measures are weighting it 30–50%, with classroom observations given greater or equal weight (Goldhaber, 2015).

Against this background, my ‘stimulus’ to this discussion is to offer expanded consideration of the potential value of classroom observations and to provide an account of how they might be used to assess the impact of teacher education programs, particularly in terms of classroom readiness, but also in relation to the Graduate Standards.

**Table 1. Challenges associated with the dominant forms of evidence used in assessment of teaching and teacher education**

Type of evidence	Major challenges
<p><b>Surveys</b></p> <p>(of teacher educators, student teachers, co-operating teachers, principals, students)</p>	<ul style="list-style-type: none"> <li>• Adequacy of survey items and scales and their interpretation</li> <li>• Adequacy of response rates</li> <li>• Potential impact of biases, personal conflicts, power relations</li> </ul>
<p><b>Student portfolios</b></p>	<ul style="list-style-type: none"> <li>• Capacity to adequately represent the graduates' classroom readiness</li> <li>• Tendency for piecemeal, ad-hoc collection of cherry-picked materials</li> <li>• Can favour those who are strong at writing</li> <li>• Students can meet the requirements of portfolio presentation but not be 'good' teachers</li> </ul>
<p><b>Value-added measures</b></p> <p>(where the major source of data in Australian schools is NAPLAN results)</p>	<ul style="list-style-type: none"> <li>• Accounting for diverse contexts</li> <li>• Time lag between data collection and any link with graduates, when programs have already moved on</li> <li>• Adequacy/precision of measures (Ballou &amp; Springer, 2015)</li> <li>• Spurious connection to what is happening in the classroom (Darling-Hammond (2015)/connection to teachers, let alone to teacher education</li> <li>• Availability of student test data, at all, and only in limited subject areas, including in Australia</li> <li>• Understanding and complexity of interpretation (Goldring et al., 2015) which makes them less acceptable/less useful to principals and others</li> <li>• Instability of estimates (Goldhaber, 2015)</li> <li>• Volatility between statistical models (Sass, 2008)</li> <li>• Lacking in specificity of evidence (Goldring et al., 2015)</li> <li>• Palatability/acceptability among teachers, teacher educators</li> </ul>
<p><b>Classroom observations</b></p>	<ul style="list-style-type: none"> <li>• Adequacy of instruments, scales and their interpretation</li> <li>• Adequacy of observation methods (including length of observations, whole lesson versus walkthrough, protocols, training)</li> <li>• Accuracy in assessing teaching quality</li> <li>• Potential impact of biases, personal conflicts, power relations</li> <li>• Willingness to embrace instruments across states and jurisdictions in Australia</li> <li>• Current use of idiosyncratic, ad-hoc, and narrative accounts of teaching quality</li> </ul>

## The case for an emphasis on classroom observations

While a combination of data sources is likely to be appropriate in providing evidence of the effectiveness of teacher education programs, I contend that the inclusion of robust observation data might be required, mandated, or even privileged, on the basis of its significant potential for providing evidence of the impact of teacher education on the quality of teaching carried out by candidates or graduates. Classroom observations, done well, can both help to enhance the practice of teaching candidates throughout their program and as graduates and, through feedback at the program and/or institutional level, inform refinement of the quality of teacher education. These are the dual concerns of a national approach to judging the effectiveness of teacher education; namely, attending to the capacities and/or performance of graduates, and attending to the capacity and/or quality of the teacher education program to achieve the desired graduate outcomes.

Observation of classrooms and feedback to teachers is a sensible platform within an Australian approach to assessing the impact of teacher education given that it is built in to the *Australian Teacher Performance and Development Framework* (AITSL, 2012) which requires that “every teacher, every year, in every school receives regular, effective and constructive feedback on their performance, as well as opportunities to identify areas for further development.” It is also built in to the *Australian Professional Standard for Principals* (AITSL, 2011) and is central to ACER’s *National School Improvement Tool* (Australian Council for Educational Research [ACER], 2012). Observation of lessons is also a ‘signature pedagogy’ (Shulman, 2005) of teacher education, with university-based and school-based supervisors already engaging in observations of student teachers as a regular and accepted practice of determining achievement against the Graduate Standards and success in the classroom. This means that the policy environment and the culture of both schools and teacher education are already at least comfortable with, if not advocating enhancement of, the observation of teaching as a key mechanism for judging and enhancing quality.

In the context of teacher assessment, Goldring et al. (2015) note: “For years, the field has lamented that principals do not spend time in classrooms and do not sufficiently attend to teaching and learning. Structured teacher observations, as integral components of teacher evaluations, are poised to be a very powerful lever for changing principal leadership and the influence of principals on schools, teachers, and learning” (pp. 102–103). I argue, similarly, that observations as an integral part of teacher development – especially implemented in ways that involve teams of teachers in conducting structured, reciprocal, comprehensive analysis of lessons – are poised as a powerful lever for changing teaching cultures and the quality of teaching (see Gore & Bowe, in press). It should be noted also that such ‘profession-shaping’ impacts have the potential to change who opts to pursue a career in teaching, the effectiveness of those currently in teaching, and who chooses to stay (Goldhaber, 2015; Gore, Barron, Holmes and Smith, forthcoming). These latter issues must be taken into account in efforts to reform teacher education, lest unintended negative consequences outweigh the benefits of a more fully-articulated system of accreditation.

If a major goal of this initiative is to ensure the ‘classroom readiness’ of teacher education graduates, then conceptualising what constitutes ‘readiness’ is fundamental to the development of a national framework. Determining this locally, with every institution generating its own model or protocol, is unlikely to advance the field. As City et al. (2009) assert, “Gaining an explicit and widely held view of what constitutes good teaching and learning in your setting is a first step toward any systematic efforts to scaling up quality” (p. 173). Among teacher educators also, even within the one institution, there is typically no clear, shared conception of good teaching and learning (Darling-Hammond & Bratz-Snowden, 2005). Simply strengthening regulatory processes is unlikely to generate greater agreement. Indeed, widespread agreement is often thwarted by the historical development

of the field of teacher education in which programs have aligned with particular paradigms (academic, efficiency, apprenticeship, social reconstructionist) (Feiman-Nemser, 1990; Zeichner, 1993), rather than recognising the commensurability and indeed desirability of attending to all four paradigms in the design and conduct of teacher education (Gore, 2001). Despite this tension, a 'combined' view of what matters in teaching – that attends to the intellectual basis for teaching, the 'science' of teaching, the 'craft' of teaching, and the moral and political effects of teaching – is entirely consistent with the breadth of the Australian Professional Standards for Teaching, organised as they are around professional knowledge, professional practice, and professional engagement.

It should be noted that under current initial accreditation processes, teacher education providers are already systematically addressing the Standards (any weaknesses in the adequacy of this approach might lie in the processes of accreditation, preparation of panel members, etc.). Classroom readiness is also already being addressed through internal assessment mechanisms, although often conflated with assessment against the Graduate Standards and often carried out by classroom teachers and/or sessional university staff during professional experience/practica/internships. The way in which teacher education providers address classroom readiness is therefore typically left to the each institution. A key question to be addressed however is 'what forms of evidence are used to judge classroom readiness and what tools are available for doing so?'

There are dozens of classroom observation protocols and pedagogical frameworks available to assist with the task of judging ‘classroom readiness’ or the quality of teaching (not that these are necessarily understood as the same thing). These include:

- Classroom Assessment Scoring System (US) (Pianta, La Paro, & Hamre, 2008)
- Framework for Teaching (US) (Danielson, 2007)
- Mathematical Quality of Instruction (US) (Hill et al., 2008)
- Protocol for Language Arts Teaching Observation (US) (Grossman et al., 2010)
- Marzano Teacher Evaluation Model (US)
- 5 Dimension Teacher Evaluation Rubric (US)
- Expert Science Teaching Educational Evaluation Model (US)
- Rapid Assessment of Teacher Effectiveness (US)
- Project for Enhancing Effective Learning (AUS)
- Biological Sciences Curriculum Study 5E (US)
- Authentic Pedagogy (US) (Newmann et al.)
- Dimensions of Learning (US)
- Five Standards of Effective Pedagogy (US)
- 3 Dimensional Pedagogy (AUS)
- Productive Pedagogies (AUS)
- Effective Teaching Profile Maori (NZ)
- 5 Dimensions of Effective Teachers (NZ)
- Learning by Design (AUS/US)
- Quality Teaching (AUS) (NSW DET, 2003)
- Quality Teaching for Diverse Students in Schooling: Best Evidence Synthesis (NZ)
- Learners Framework
- Pedagogy of Engagement
- Reflect, Respect, Relate (AUS)
- South Australian Teaching for Effective Learning (AUS)
- 8 Aboriginal Ways (AUS)
- E5 Instructional Model (AUS)
- 6 Principles for Teaching Mathematics (AUS)
- 5 Dimensions of Teaching and Learning (US)
- Innovative Teaching and Learning Model (US/AUS)
- Teach like a Champion (US)

However, not all observations tools/protocols are alike. A common problem with many existing observation approaches is that everyone is judged to be effective (in part because observation protocols are typically criterion-referenced [Cohen & Goldhaber, in press]), with little specificity as to what effective means. The same might be said of the *Australian Professional Standards for Teachers* which, as currently written, do not provide specificity about, or a clear position on, what constitutes high quality classroom (or assessment) practice; for example, the Standards do not articulate what counts as “effective” (Ladwig & Gore, 2009). While there is greater specification in many observation frameworks than many sets of standards for teachers, the construct validity of a particular protocol “rests on the assumption [that] the practices featured are consequential in terms of students’ experiences” (Cohen & Goldhaber, in press). Hence, what is addressed in the protocol must matter in terms of impacting on the teaching-learning experience.

An observation tool/protocol also needs to be acceptable to teachers and teacher educators, to resonate with their (diverse) views of good teaching. It needs to have construct validity and provide specificity without rigidity, so as to allow for diverse teaching

contexts. In the United States, only CLASS and FFT have been recognised as having sufficient technical power for use in teacher performance evaluations (Worrell et al., 2014), although it is not clear how many frameworks have been tested.

In Australia, the major pedagogical frameworks currently in use, and all conceptually related through their shared origins in 'Authentic Pedagogy' (Newmann & Associates, 1996), are: 'Productive Pedagogy/ies' (Qld); 'Quality Teaching' (NSW, ACT); and 'Teaching for Effective Learning' (SA) (derived from Quality Teaching/Productive Pedagogy).

The Quality Teaching framework (NSW DET, 2003) has been used in a decade-long program of research showing that:

- The scales that make up the framework, Intellectual Quality, Quality Learning Environment, and Significance, have technical power in terms of validity and reliability;
- The conceptual basis of the framework resonates with teachers;
- Using the framework strengthens the basis for productive professional conversations among student teachers, teachers, and teacher educators by providing a shared set of concepts and language for talking about teaching;
- Student performance increases when Quality Teaching is high; and
- Equity gaps for low SES and Aboriginal students narrow when Quality Teaching is high. (see Gore, 2014a; Ladwig, 2007, 2010; Ladwig et al., 2007)

Furthermore, when used in teacher professional development, through the approach we call 'Quality Teaching Rounds' (Gore, 2014b), we have found measurable improvements in the quality of teaching. Our current cluster randomised controlled trial of teachers participating in Quality Teaching Rounds has produced effect sizes in the improvement of teaching quality of 0.3–0.4 for the intervention groups, a substantial measurable effect given the complexity of the teaching enterprise and difficulties of measurement.

Quality Teaching Rounds, which can be used in preservice or inservice teacher development, involve observation of whole lessons in groups of at least four using the Quality Teaching framework to guide analysis, discussion, and refinement of teaching practice. This approach aligns with much of the literature on effective teacher professional development, which highlights the merits of teachers working collaboratively in the context of their own schools, with a focus on their students' learning (Bolam et al., 2005; Louis, 2006). Working in groups to evaluate and discuss teaching quality, guided by the specificity of the framework, not only helps to enhance teaching practice and build teaching culture, but also contributes to the reliability of judgements. Polikoff (2013, 2014), for example, found that classroom observations are more reliable when using two different raters than when using two observations for a single rater, while Kane and Staiger (2012) found teacher ratings to be fairly consistent when using generic protocols (applicable across subjects and grades), like Quality Teaching, and less consistent when they were assessing content-specific teaching practices.

There is not scope here to go into more detail about Quality Teaching and Quality Teaching Rounds. But as an exemplar of how the Quality Teaching classroom observation 'protocol' could be used in the context of assessing the effectiveness of teacher education, a few additional points are warranted. First, participation in Quality Teaching Rounds, a process that could easily be incorporated in existing practicum experiences, engages teaching candidates in addressing professional knowledge, professional practice, and professional engagement, thus aligning directly with the *Australian Professional Standards for Teachers*. Second, with the specification of good teaching provided through the Quality Teaching framework as an observation protocol, it provides a valuable means by which to assess 'classroom readiness.'

Given the demonstrable impact of Quality Teaching on student outcomes and its capacity to improve teaching quality – and in the absence of student test scores that could be usefully and defensibly be attributed to the teacher education program – the Quality Teaching framework offers a strong basis for observations as a key form of the evidence of teacher education program effectiveness. Indeed, it has potential beyond other observation tools (including those favoured in the US), because its conceptual bases goes beyond the technical delivery of teaching to the purposes and effects of teaching, as valued by teachers and articulated in the *Melbourne Declaration on Educational Goals for Young Australians* (Barr et al., 2008).

If teacher education programs are to be assessed either by their impact on student learning or by their impact on teaching practice, I propose that in Australia we are much better positioned to focus on the impact of teacher education on teaching practice (specifically those teaching practices known to produce learning/correlated with improved student performances).

## On the use of classroom observations as evidence of program effectiveness

In this section, I offer a few comments on how classroom observations might be used as evidence of program effectiveness. First, it should be clarified whether an application for second phase accreditation of teacher education programs should provide evidence of the judgements made by the teacher education provider about the performance of graduates, or cohort averages, or only provide an account of how (types of) evidence has been gathered, interpreted, and used internally for the assessment of teaching candidates/ graduates and for the further refinement of the teacher education program. With the former, with an observation instrument such as the Quality Teaching framework, which has been used in a 12-year program of research, the average Quality Teaching scores of practising teachers could be used as a benchmark against which to report on the level of achievement of graduates. It might be reasonable, for example, to report on the proportion of graduates whose lessons scored within ranges at, above, and below these averages. Alternatively, or additionally, a provider might set its own standards for quality, for instance dictating that a student is not “classroom ready” unless able to consistently achieve a certain level for each of the three dimensions of the Quality Teaching framework, adjusted for context. Teacher education providers might also choose to use such observations as a common form of assessment across institutions and a basis for expanding research in teacher education. In making these suggestions, I want to emphasise that I am not advocating a neat numerical system (just as such an approach does not account adequately for diverse contexts and other confounding variables with value-added measures).

Observational scales could also over- or underestimate the true instructional quality of a classroom. Teachers could also be assigned to classrooms in which it is systematically more difficult to engage in the high-quality teaching practices measured on the observation protocols described above. In addition, certain curricula might circumscribe instruction, making it easier or harder to engage in specific practices. (Cohen & Goldhaber, in press, p. 6)

Nonetheless, utilising robust observation data throughout a teacher education program, from beginning to end, would inform discussions with teaching candidates to enhance teaching practice and enable targeted program refinement.

Second, I want to propose that a robust observation protocol, used appropriately as introduced above, might be the major form of evidence required for satisfying reaccreditation requirements. While portfolios, surveys, and measures of student learning

(such as artefacts of their work) will be incorporated in the internal evaluation of graduates and teacher education programs, and other documentation regarding staffing and other resources including school partnerships might be provided for re-accreditation purposes, I contend that robust classroom observation data of the kind I am outlining might be the centrepiece of providing evidence of program impact. Such status can only be accorded to observation data if: (1) the observation protocol meets acceptable standards of construct validity, reliability, and fairness; (2) the observation protocol aligns with the *Australian Professional Standards for Teachers*; and (3) the observation protocol serves as a credible proxy for 'classroom readiness.' The Quality Teaching framework meets these conditions. If 'robust' classroom observation is built into both the learning experiences and the assessment of teaching candidates, then making the case for program impact with classroom observation as the centrepiece would also ensure that re-accreditation was not overly burdensome for universities, systems, or schools.

## Summary

If teacher education providers have a coherent, tested framework for guiding both program and graduate quality, and they have assessed graduates/students against that framework, and the framework aligns with the Standards, then a valid, achievable, feasible approach to demonstrating the impact of teacher education on ensuring graduates not only meet the Graduate level *Australian Professional Standards for Teachers* but are certified as 'classroom ready' is to report on how such a framework is used to (a) guide program design and delivery and (b) carry out graduate assessment, including during practicum/professional experience.

Value-added data, on the other hand, a major feature of teacher assessment in the US, provides no direct link to what teachers are doing in classrooms (Goldring et al., 2015) and therefore does not provide for 'actionability' (CAEP Evidence Guide) – program evaluation and improvement. Observation, done well, would provide timely, relevant, and transparent data – and be actionable!

## References

- Australian Institute for Teaching and School Leadership. (AITSL). (2012). *Australian Teacher Performance Development Framework*. Retrieved from [http://www.aitsl.edu.au/docs/default-source/default-document-library/fact\\_sheet\\_-\\_australian\\_teacher\\_performance\\_and\\_development\\_framework](http://www.aitsl.edu.au/docs/default-source/default-document-library/fact_sheet_-_australian_teacher_performance_and_development_framework)
- Australian Institute for Teaching and School Leadership. (AITSL). (2011). *Australian Professional Standards for Teachers* Retrieved from [http://www.aitsl.edu.au/docs/default-source/apst-resources/australian\\_professional\\_standard\\_for\\_teachers\\_final.pdf](http://www.aitsl.edu.au/docs/default-source/apst-resources/australian_professional_standard_for_teachers_final.pdf)
- Australian Institute for Teaching and School Leadership. (AITSL). (2011). *Australian Professional Standard for Principals*. Retrieved from [http://www.aitsl.edu.au/docs/default-source/default-document-library/australian\\_professional\\_standards\\_for\\_principals.pdf](http://www.aitsl.edu.au/docs/default-source/default-document-library/australian_professional_standards_for_principals.pdf)
- Australian Council for Educational Research (ACER). (2012). *National School Improvement Tool*. Retrieved from <https://www.acer.edu.au/files/NSIT.pdf>
- Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77–86. doi:10.3102/0013189x15574904
- Barr, A., Gillard, J., Firth, V., Scrymgour, M., Welford, R., Lomax-Smith, J., ... & Constable, E. (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Melbourne, Australia: Ministerial Council on Education, Employment, Training and Youth Affairs.
- Bolam, R., McMahon, A., Stoll, L., Thomas, S., Wallace, M., Greenwood, A., . . . Smith, M. (2005). *Creating and sustaining effective professional learning communities*. London, UK: University of Bristol/Institute of Education.
- City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard Education Press.
- Cohen, J., & Goldhaber, D. (in press). Observations on evaluating teacher performance: Assessing the strengths and weaknesses of classroom observations and value-added measures. In J. A. Grissom, & P. Youngs (Eds.), *Making the most of multiple measures: The impacts and challenges of implementing rigorous teacher evaluation systems* (under contract). New York: Teachers College Press.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching (2nd ed.)*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132–137. doi:10.3102/0013189x15575346
- Darling-Hammond, L., & Bratz-Snowden, J. (2005). *A good teacher in every classroom: Preparing the highly qualified teachers our children deserve*. San Francisco, CA: John Wiley and Sons.
- Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 212–233). New York: Macmillan.
- Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, 44(2), 87–95. doi:10.3102/0013189x15574905
- Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., &

- Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104. doi:10.3102/0013189x15575031
- Gore, J. M. (2001). Beyond our differences: A reassembling of what matters in teacher education. *Journal of Teacher Education*, 52(2), 124–135.
- Gore, J. (2014). Towards quality and equity: The case for *Quality Teaching Rounds*. In *Quality and Equity: What does research tell us? Proceedings of the Australian Council for Educational Research (ACER) Research Conference* (pp. 86–91). Melbourne, Australia: ACER.
- Gore, J. (2014b). Effective implementation of pedagogical reform through Quality Teaching Rounds. In *Proceedings of the annual national conference of the Australian College of Educators (ACE): What counts as quality in education?* (pp. 16–21). Melbourne, Australia: ACE.
- Gore, J. M., & Bowe, J. M. (in press). Interrupting attrition? Re-shaping the transition from preservice to inservice teaching through Quality Teaching Rounds. *International Journal of Educational Research*.
- Gore, J., Barron, R. J., Smith, M., & Holmes, K. (forthcoming). *Who should teach? Discourses of teacher selection and the aspirations of school students interested in becoming teachers*.
- Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English Language Arts and teachers' value-added scores* (NBER Working Paper No. 16015). Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w16015.pdf>
- Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). *Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An exploratory study*. *Cognition and Instruction*, 26(4), 430–511. doi:10.1080/07370000802177235
- Jiang, J. Y., Spote, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students [Special issue]. *Educational Researcher*, 44, 105–116. doi:10.3102/0013189X15575517
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.
- Ladwig, J. G. (2007). Modelling pedagogy in Australian school reform. *Pedagogies: An International Journal*, 2(2), 57–76.
- Ladwig, J. G. (2010). Beyond academic outcomes. *Review of Research in Education*, 34, 113–141.
- Ladwig, J. G., & Gore, J. M. (2009). Re-reading the standards agenda: An Australian case study. In M. Simons, M. Olssen, & M. A. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 772–734). Rotterdam, Netherlands: Sense.
- Ladwig, J. G., Gore, J. M., Miller, W. A., Griffiths, T. G., & Smith, M. (2007). *Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy*. Paper presented at the annual conference of the Australian Association for Educational Research, Fremantle, Australia. Retrieved from <http://www.aare.edu.au/data/publications/2007/>

[lad07283.pdf](#)

Louis, K. S. (2006). Changing the culture of schools: Professional community, organisational learning and trust. *Journal of School Leadership*, 16(4), 477–489.

Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco, CA: Jossey-Bass.

NSW Department of Education and Training (NSW DET). (2003). *Quality teaching in NSW public schools: A classroom practice guide*. Sydney, Australia: NSW Department of Education and Training/Professional Support and Curriculum Directorate.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Manual: Pre–K*. Baltimore, MD: Brookes.

Polikoff, M. S. (2013, March). The stability of observational and student survey measures of teaching effectiveness. Paper presented at the annual conference of the Association for Education Finance and Policy, New Orleans, LA. Retrieved from <http://www.aefpweb.org/sites/default/files/webform/Stability%20of%20observational%20measures.pdf>

Polikoff, M. S. (2014). Does the test matter? Evaluating teachers when tests differ in their sensitivity to instruction. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 278–302). San Francisco, CA: Jossey-Bass.

Sass, T. R. (2008). *The stability of value-added measure of teacher quality and implications for teacher compensation policy* (Report Brief 4). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from <http://files.eric.ed.gov/fulltext/ED508273.pdf>

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, 134(3), 52–59.

Teachers College, Columbia University. (2015, April 2). Sachs Lecture Series. David Berliner: Evaluating teacher education and teachers using student assessments [Video file]. Retrieved from <https://youtu.be/IQkSnpFvbcs>

Teacher Education Ministerial Advisory Group. (2014). *Action now: Classroom ready teachers*. Canberra, Australia: Department of Education. Retrieved from <http://www.studentsfirst.gov.au/teacher-education-ministerial-advisory-group>

Worrell, F., Brabeck, M., Dwyer, C., Geisinger, K., Marx, R., Noell, G., & Pianta, R. (2014). *Assessing and evaluating teacher preparation programs*. Washington, DC: American Psychological Association.

Zeichner, K. M. (1993). Traditions of practice in U.S. preservice teacher education programs. *Teaching and Teacher Education*, 9(1), 1–13.



 [aitsl.edu.au](http://aitsl.edu.au)

 [facebook.com/aitsl](https://facebook.com/aitsl)

 [twitter.com/aitsl](https://twitter.com/aitsl)

 [youtube.com/aitsleduau](https://youtube.com/aitsleduau)

**aitsl**  
Australian Institute  
for Teaching and  
School Leadership  
Limited