

The Development of an Australian Teacher Performance Assessment: Lessons from the International Literature

Meghan Stacey, Debra Talbot, John Buchanan and Diane Mayer

The Version of Record of this manuscript has been published and is available in
Asia-Pacific Journal of Teacher Education, 23-9-19 online first

<http://www.tandfonline.com/10.1080/1359866X.2019.1669137>

The Development of an Australian Teacher Performance

Assessment: Lessons from the International Literature

Abstract

The requirement for all Initial Teacher Education programmes in Australia to include a capstone teacher performance assessment (TPA) is relatively new. However, TPAs are common in other countries, particularly the United States. In this article, we report on a review of the literature instigated by the authors' involvement in the development of an Australian TPA. Through a systematic review of the international research literature, supplemented by a separate set of preliminary sources, we identify and explore a range of key considerations for the development of TPAs within the Australian context. We focus on common aspects of TPAs worldwide: planning and preparation, observations on and evidence of teaching practice, and student work samples. We then interrogate further issues related to TPAs including the role of schools, principals and teachers; relation to Initial Teacher Education coursework; fairness; validity and reliability; and rubric development. To conclude we present a series of guiding principles to support the development and implementation of such complex, high stakes, and increasingly mandated kinds of assessments.

Keywords: teacher performance assessment, initial teacher education, teacher education accountability, teacher education policy

Word count: 5490

Introduction

Between 2016 and 2018 Initial Teacher Education (ITE) programmes in Australia, which were previously accredited under state-based guidelines, were required to transition to a set of new national Program Standards. The new national system meant that ITE providers would now be required to respond to national Program Standard 1.2 which states that pre-service teachers (PSTs) must “successfully [complete] a final-year teaching performance assessment prior to graduation” (Australian Institute of Teaching and School Leadership, 2015, p. 10). This new requirement followed the endorsement of TPAs as examples of “good practice” in assessing the “overall competency” of PSTs internationally, in the report of the Teacher Education Ministerial Advisory Group into Australian ITE (Teacher Education Ministerial Advisory Group, 2014, p. 31). Thus, while final, summative, capstone teacher performance assessments (TPAs) are common in some international contexts, particularly the United States (US), they are relatively new in Australia.

Program Standard 1.2 states that Australian TPAs must accomplish the following: include “the elements of planning, teaching, assessing and reflecting”; “be a valid assessment that clearly assesses the content of the Graduate Teacher Standards”; have “clear, measurable and justifiable achievement criteria”; be reliable; and include “moderation processes that support consistent decision-making” (Australian Institute of Teaching and School Leadership, 2015, p. 10). To support ITE institutions’ responses to the new mandate, AITSL offered a grant programme, open to consortia of ITE institutions to develop a national TPA. After a competitive review process in 2017, AITSL announced the success of two consortia (Australian Institute of Teaching and School Leadership, 2017). Work on the literature review which is the focus of this article was begun by the Literature Review Working Group at the University of Sydney in 2017, in support of one of these successful consortia, consisting of the University of Melbourne, Charles Darwin University, Curtin University, Federation University Australia, the University of Canberra, the University of Newcastle, the University of Sydney, the University of Technology Sydney, the University of Western Australia and

Victoria University. Given the newness of TPAs in the Asia Pacific, and the directive from AITSL that such assessments must now be implemented by all Australian ITE providers, our aim in this article is to support the research-informed development and implementation of such complex, high stakes assessment tasks.

We begin with a contextualising discussion of existing TPAs internationally. We then explain the methodology employed for developing our systematic review of the literature, aligning it with, and adding to, the preliminary literature gathered for the consortium project. The remainder of the article presents the findings of the review, first in relation to the core dimensions of planning and preparation, observations on and evidence of teaching practice, and student work samples. Each of these dimensions was required by AITSL, as outlined above; they are also common among existing TPAs. Reflection on practice, a further dimension required by the national Program Standard, is considered relevant to each of these sections and so is discussed within each of these sub-headings. The second major findings section explores remaining issues and concerns. Finally, we conclude with the presentation of some key guiding principles drawn from this literature and the broader consortium development process in order to guide the development of TPAs within the Asia-Pacific context.

Contextualisation

The development of performance assessments for PSTs has been an emergent policy priority in teacher education over the past twenty years. A focus on the standardised assessment of student achievement has arguably led to scrutiny of ITE providers in some countries, requiring them to be accountable for producing teachers who are 'classroom-ready' (Buchanan & Schuck, 2016). In the US, Linda Darling-Hammond and colleagues have devised a career-long, standards-based assessment trajectory for teachers (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Darling-Hammond et al., 2011),

beginning with “high-quality preparation, evaluated authentically through performance assessments that both develop and measure beginning teacher effectiveness” (Darling-Hammond, 2012, p. 9). These assessments have been identified as arising from “a context wherein the construction of the problem of teacher education was lack of accountability and standardization of expectations” (Cochran-Smith, Piazza, & Power, 2013, p. 16) and where “anecdotally informed ‘teacher education is failing us’ headlines” abound (Mayer, 2013, p. 9). However, such tasks have also been seen as a way of enabling authentic, valid assessment of the actual work of PSTs in the classroom (see e.g. Darling-Hammond, 2012). In this way, TPAs have the potential to position teacher education accountability largely within the hands of the teaching profession.

The largest and most controversial teaching performance assessment currently in use is the edTPA in the US. The edTPA, developed by the Stanford Centre for Assessment Learning and Equity and administered by Pearson, is currently used in a number of states across the US, including – and increasingly – for teacher and programme ‘licensure’ (Behney, 2016). The edTPA straddles three domains of teaching – planning, instruction and assessment (Ratner & Kolman, 2016) – and was itself developed from the Performance Assessment for California Teachers (PACT), which has also influenced the development of a number of other TPAs including the Authentic Teacher Assessment (ATA) at Deakin University in Australia (Mayer, Allard, Moss, & Dixon, 2015).

These existing teaching assessments (edTPA, PACT, ATA) have many similarities, such as the use of a structured portfolio approach and alignment with a particular set of professional standards. But they also have important differences. The outsourcing of examining the edTPA is one obvious difference; this is discussed further below. Another distinction is the number and nature of domains. For example, the edTPA subsumes the ‘Context for Learning’ and ‘Planning and Instruction’ aspects of PACT into its own ‘Planning’ task, and the ‘Assessing Student Learning’ and ‘Reflecting on Teaching’ aspects into its ‘Assessment’ task, while ‘Instructing Students and Supporting Learning’ in PACT becomes the ‘Instruction’

task in the edTPA (Lim, Stallings, & Kim, 2015). Another difference is the conceptualisation of the TPA and what it is aiming to measure – teaching in its entirety, simply a ‘snapshot’, or something in between? For instance, both the edTPA and the PACT consist of two elements, a ‘Teaching Event’ as well as ‘Embedded Signature Assessments’. Embedded Signature Assessments take place throughout coursework in support of the capstone Teaching Event, a programmatic feature that acknowledges the inability of a single task to capture *all* of teachers’ work (Mayer, 2013).

There is a final, important point to be made here in contextualising our discussion of TPAs within the international space. This concerns the temptation, evident internationally although not yet in Australia, to assess teachers, and sometimes ITE providers, based on the standardized test results of school students through value-added modelling (VAM). The problems with the validity and reliability of such measures are many and are well-documented (see e.g. Berliner, 2013; Betebenner et al., 2012; Darling-Hammond et al., 2012; Darling-Hammond et al., 2011; Forman & Markson, 2015; Palardy & Peng, 2015). The consensus within the literature is that ‘value-added’ assessment can only ever be used for validation purposes within a suite of other measures, for information rather than direct assessment (Betebenner et al., 2012), and that any measure of growth of student learning needs to be based on teachers’ actual work with students in classrooms (Darling-Hammond, 2015).

Methodology

In 2017 the authors of this article were members of a working group within a consortium developing an Australian TPA. The working group was tasked with developing a set of design principles to inform the development of the TPA. This was first done through gathering a preliminary body of literature, initially through input from the wider consortium members. Given the range of information available, however, a systematic review was

warranted, and subsequently conducted. It is the results of this systematic review that forms the bulk of the data presented in the second half of this article, although these are supplemented by some of the preliminary sources (a process for which we provide more detail below).

The systematic review aimed to ascertain what lessons could be inferred from the implementation and evaluation of existing teacher performance assessments, which might guide the design of TPAs within the Asia-Pacific context. Search terms for this systematic search included: teach* AND performance AND assess*, edTPA, and PACT. These terms were searched within JSTOR and ProQuest Education databases, controlling for peer reviewed articles published within the last ten years. This process followed a range of recommended systematic literature review strategies, for instance by enabling various truncations of the keywords, targeting the primary databases relevant to the field, and ensuring that the results would be recent (O'Brien & McGuckin, 2016). This process returned a total of 2160 search results, of which a final 31 were found to be relevant and useful after applying the following additional inclusion and exclusion criteria (O'Brien & McGuckin, 2016). Items were discarded if they were not substantively focused on a TPA or similar teacher evaluation system, either in whole or in part, and if the TPA in question simply provided the context or the data for an unrelated analysis.

In the results sections below, we first present some of the common themes which emerged, grouped under headings reflective of the common components of existing TPAs: planning and preparation, observations on and evidence of teaching practice, and student work samples. This is followed by a section exploring remaining issues and concerns in two parts. First, we consider issues and concerns raised by the literature gathered through the systematic review. Second, we examine further issues and concerns, not raised in the systematic review, but which were evident in literature gathered through the preliminary literature gathering process. The inclusion of this additional material serves to round out our discussion of relevant concerns, whilst also highlighting how leaders in the field can

contribute to the literature review process so as to provide more comprehensive results (Onwuegbuzie & Frels, 2016). Indeed, while systematic reviews of the literature are efficient and have reduced scope for bias, considering a field more broadly can lead to “insights that can be neglected or passed over in the steps towards exclusion and quality control that are required in the systematic review model” (Jesson, Matheson, & Lacey, 2011, p. 15).

Common Aspects of a TPA

Planning and Preparation

Planning documentation is one of the most common aspects of teacher portfolio assessments and is generally seen to be a core part of more authentic approaches to teaching assessments (Darling-Hammond, 2012). The inclusion of planning documentation is noted in the literature which discusses models like the edTPA and PACT (Darling-Hammond, 2012; Noel, 2014; Stewart, Scalzo, Merino, & Nilsen, 2015). In current teaching assessment models, the focus of preparation is commonly in pursuit of a particular ‘central focus’ or learning objective (Hébert, 2017). Lessons planned are generally expected to be consecutive or at least sequential, and usually stipulate a number range, such as 3-5 lessons for the edTPA (Behney, 2016). The types of documentation include lesson plans and other ‘materials’ relating to pedagogy and assessment (Stillman et al., 2013), as well as, often, commentaries and justifications of choices and selections in relation to context, such as in the edTPA (Behney, 2016), and, more explicitly, in the PACT (Bunch, Aguirre, & Téllez, 2009). Thus, the provision of planning documentation by PSTs provides one opportunity to focus on issues of context and student diversity. This is an essential and ongoing consideration (Stillman et al., 2013), especially given that TPAs are frequently critiqued for failing to foreground issues of race and racism, for instance (Tuck & Gorlewski, 2016).

Issues that have arisen regarding the inclusion of planning documentation in teaching assessments are primarily related to length, manageability and the selection of materials.

One concern within the literature has been the “labor intensive” (Ginsberg & Kingston, 2014, p. 40) nature of such work, with the provision of planning documentation and commentary potentially having the perverse effect of “detract[ing] from time candidates need to prepare lessons and evaluate students” (Hébert, 2017, pp. 74-75) in their day to day work. Similar concerns have been noted by Roosevelt & Garrison (2018) as reducing PSTs’ sense of ‘pedagogical responsibility’ by encouraging a performative focus on oneself rather than one’s students. There is therefore a potential tension regarding what PSTs choose to include or not include, with a danger that they may be tempted “to make compromises” so as to “be compliant with the assessment expectations” (Sato, 2014, p. 430); indeed, commentary on TPAs has suggested that PSTs can feel pressured to teach ‘palatable’ lessons (Snyder, 2009). In the light of these critiques, one can conclude that planning documentation needs to be tied closely to PSTs’ actual classroom work, with any ‘messiness’ or lack of cohesion seen as indicative of the complexities of real classrooms. This is also important given related concerns that contexts “stressed by poverty and low student performance” can be perceived as “unfair”, “when other candidates are teaching in contexts that are perceived to be ‘easier’ or ‘less challenging’” (Sato, 2014, p. 430). On the contrary, ‘messiness’ should ideally be viewed as providing important opportunities for meaningful, nuanced reflection.

Observations on and Evidence of Teaching Practice

Demonstrations of PSTs’ enacted classroom practice are also commonly assessed during teacher preparation programmes and as part of formalised TPAs. Historically, demonstration and assessment occurred simultaneously, in real-time, and involved an observer who utilised an observation schedule to record and judge the performance of the PST. In contrast, the majority of the TPAs cited above involve the PST recording 10-15 minutes of video of their classroom practice.

The use of video to record a segment of enacted classroom practice and learning has advantages in terms of providing both evidence and a prompt for PST reflection on their practice (Jackson, Kelsey, & Rice, 2018; Noel, 2014; Okhremtchouk, Seiki, Gilliland, Ateh, &

Wallace, 2009). Video as a form of evidence does present concerns, however. For instance, if it is unfamiliar to candidates, it can be viewed as intrusive (Sato, 2014). Okhremtchouk et al. (2009) point to logistical difficulties related to videoing within individual schools, and potential ethical issues around PSTs focusing more on filmed classes than the rest of their teaching load. Noel (2014), on the other hand, discusses the “surprising...lack of comments about the challenges of videoing” and the relative ease with which most candidates used their phones or laptops to record their videos (p. 367). Noel’s (2014) work also discusses the considerations of PSTs regarding what should or should not be included in the video segment, which raises questions over how the video segment may be constructed: what is included, and what is left out.

Student Work Samples

If teaching is framed as knowledge-informed intellectual work involving professional decisions in relation to student learning, teacher evaluation should include some evidence of student learning (or not) and teacher reflection upon, and resultant decision making in relation to that learning. In the edTPA, teacher candidates:

submit assessment data; analyse whole class performance, in particular students’ interpersonal and presentational communication; document examples of differentiated instructional practices using data and work samples for three focus students; state the implications of the data; and explain their plans for further instruction. (Behney, 2016, pp.272-273)

For the TE aspect of PACT, candidates submit “an in-depth analysis of the work of several individual students” (Bunch et al., 2009, p. 106). Collecting student work samples and annotating them in terms of future planning and teaching is a further opportunity for candidates to demonstrate reflective processes, and has been shown to enhance PSTs’ ability to gather and analyse information to inform instruction (Campbell et al., 2016). However, in a discussion concerning the establishment of face validity of TPAs, Sato (2014)

draws attention to the importance of privileging student work samples based on their actual, regular work in the classroom, over any use of standardised test results (see also Brady, Heiser, McCormick, & Forgan, 2016; Ginsberg & Kingston, 2014). It is important to note that the assessment component of TPAs (designing assessment and interpreting the results) are generally found to be more difficult than other aspects of the task (Kim, 2019). This is, however, also one component argued to have particular predictive validity for the ongoing 'effectiveness' of PSTs in their future work as teachers (Darling-Hammond, Newton, & Wei, 2013).

Additional Issues and Concerns

Arising from the Systematic Review

The literature gathered in this systematic review suggests that a further issue is the various roles of schools, principals and teachers, and the kind of support they a) could and should be given, and b) could and should give in the assessment process (Chizhik, Chizhik, Close, & Gallego, 2017; Kissau, Hart, & Algozzine, 2019; Seymour, Burns, & Henry, 2018). A related question is how the TPA can be integrated into, and supported by, ITE provider coursework. 'Frontloading' TPA-like experiences has been the approach of some teacher education programmes internationally (Wahl, 2017), while evidence from TPAs has also been used in programme evaluation (Bunch et al., 2009). The use of TPAs for programme design has been recommended by some of their early architects (Peccheone & Whittaker, 2016), however the potentially negative impact of TPAs on teacher preparation programmes for 'teaching to the test' and 'co-opting' teacher education programmes has also been highlighted (Hébert, 2017; Krise, 2016; Ledwell & Oyler, 2016). A further issue indicated by the literature is the question of fairness in relation to age, gender, race and ethnicity (Denner, Norman, & Lin, 2009; Goldhaber, Cowan, & Theobald, 2017), as well as practical considerations around cost (Snyder, 2009), time commitment and turnaround for feedback (Campbell et al., 2016). In relation to fairness, further concerns include whether TPAs better

reflect a candidate's ability to negotiate the requirements of the task, including the heavy emphasis on written responses, rather than their overall ability to be an effective teacher (Clayton, 2018; Sandholtz, 2012). It is therefore considered that having multiple forms of evidence within the overall assessment is important (Sandholtz, 2012), as it ensures a more holistic view deriving from various aspects of the assessment, which ultimately should provide a range of ways of seeing and assessing the interrelated sequence of lessons. Finally, ensuring reliability through the training of assessors and other moderation procedures is another issue of central importance and this was critiqued within some of the literature gathered through the systematic review (e.g. Swanson & Hindebrandt, 2017). Validity is also addressed to some extent, such as in guidelines for establishing face validity through emphasising authentic, contextualised flexibility, and analyses of content and construct validity as related to teaching standards (Sato, 2014; Wilkerson, 2015). However, issues of validity, reliability and scoring were not addressed in a more technical sense in the literature gathered for the systematic review. Next we review literature that was not gathered through the systematic review but which was compiled during preliminary work to inform the guiding principles for the development of the consortium's TPA.

Beyond the Systematic Review

Issues of validity and reliability, and associated questions concerning rubric development, scoring and moderation so as to ensure that our considerations were as all-encompassing in nature as possible, were a key consideration in our preliminary work. Insights can be garnered from the literature on assessment. For instance, Humphry and Helsinger (2014) have demonstrated flaws within the common matrix design of rubrics and their use of the same number of potentially rather arbitrary qualitative gradations across criteria. They further assert that "it is desirable for criteria to contain descriptions of performances free of obvious overlap or redundancy to allow raters to focus on distinctive and complementary aspects of students' performances" (Humphry & Helsinger, 2014, p. 256). Similarly, Pufpaff et al.'s findings suggest that simpler rubrics, which measure "only one performance dimension per row" (p. 135) are more effective. In relation to TPAs more specifically, the technical report

produced for PACT (Pecheone & Chung, 2007) is of particular utility. In PACT, a 'pass' is determined through the following process: a panel of teacher educators first formulate initial recommendations; a review of these recommendations produces a decision regarding passing standards and a cutscore model, reviewed by all participating programmes, and; these decisions are then reviewed by policymakers across the consortium (Pecheone & Chung, 2007). To establish benchmarks, Teaching Events from the first pilot year were used, teams assessing these independently before selecting samples for scorer training (Pecheone & Chung, 2007). Fail grades are double marked, as are a selection of random samples each cycle (Pecheone & Chung, 2007). A further, related consideration is how a holistic assessment can be achieved across the separate (but essentially interrelated) aspects of a TPA. Pecheone and Chung (2007) describe a sequential scoring process for the PACT, in which scorers read or view each task in order, scoring according to rubrics and taking notes. As part of the PACT evaluation process, assessors' overall judgement of a PST was also used as an assessment of criterion-related concurrent validity against the usual scoring process to counter anomalies (Pecheone & Chung, 2007). While developing a holistic sense of a PST's teaching within a particular school context is a core goal of TPAs, we note that concerns persist in trying to ascertain PSTs' 'competency' beyond individual school settings. In part to address this, the preliminary scoping of literature for the consortium also included some material on the use of video-based 'teaching scenarios' (see e.g. Borko, 2016).

In short, this additional literature provided deeper insight into mechanisms for enabling valid and reliable assessment of individual TPAs and moderation within and across cohorts. These insights, in turn, further inform design considerations for a TPA that aims for an authentic assessment of preservice teachers' practice.

Conclusion

To conclude, we briefly summarise some of the principles and ongoing concerns arising from the literature to guide the development and implementation of TPAs within the Asia-Pacific.

First and foremost, the collective literature emphasises that a TPA must be designed to capture the actual work of teaching in school classrooms. However, there is ongoing and increasing concern about whether this is possible. For example, Hebert (in press, 2018) suggests the edTPA mischaracterizes student teaching as 'real' teaching by positioning candidates as autonomous agents. She suggests that the edTPA might be better utilized as a means of assessing novice employed teachers rather than preservice teachers who operate within a range of external restrictions. A second principle is that any TPA should aim for maximum authenticity in which any measurement of growth in student learning remains strictly context-specific. When a task is removed from context, or becomes viewed as primarily performative or centred, for instance, on writing skills rather than teaching, validity is compromised. TPAs therefore need to include multiple forms of evidence, and integrate knowledge of content, students and context. In relation to this emphasis on context, a third principle is that an Australian TPA should also provide, if not insist on, the opportunity for PSTs to demonstrate their ability to teach in diverse contexts with diverse students with diverse learning needs. This links to a fourth principle, that the complexity of real classrooms must be recognised. As Paugh et al. (2017) found, reflection is often stymied for many candidates who decide to prioritize easier cases or ignore complexity, essentially choosing between compliance with the TPA format and genuine inquiry-based decision-making. Fifth and finally, we note that practical issues, including the TPA's relationship to broader ITE coursework and moderation processes, also need to be taken into consideration.

The development and implementation of a valid and reliable TPA that meets Australian accreditation requirements is no straightforward matter. However, the lessons from the literature outlined in this article offer some directions as to how this might be done. The significance of our contribution here is to draw together such sources of evidence at a critical

moment in Australian education policy, with all institutions now required to implement a TPA, and to present this collection as a resource for doing so. Given the breadth of the literature reviewed, it is our belief that the resulting principles presented might also have applicability beyond the Asia-Pacific context.

Acknowledgements

The authors acknowledge the seed funding from the Australian Institute for Teaching & School Leadership (AITSL) in the first year of the project (2017) and the significant and ongoing financial support provided by member institutions of the Assessment for Graduate Teachers (AfGT) Consortium. The authors also thank members of the AfGT Consortium for the enthusiastic and collaborative work in developing the AfGT.

The authors would also like to acknowledge the contribution of the wider Literature Review Working Group at the University of Sydney in preparing the original literature review upon which this article is based: Janette Bobis, Patrick Brownlee, Wayne Cotton, and Alyson Simpson.

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the interpretation of the data by all members of the AfGT Consortium.

Funding

This work was supported by the Australian Institute for Teaching and School Leadership (AITSL).

References

- Australian Institute of Teaching and School Leadership. (2015). *Accreditation of ITE programs in Australia: Standards and Procedures*. Melbourne: AITSL.
- Australian Institute of Teaching and School Leadership. (2017). Teaching Performance Assessment - AITSL Grant Program Announcement 30 June. Retrieved from https://www.aitsl.edu.au/docs/default-source/default-document-library/tpa-announcement---30-june-2017.pdf?sfvrsn=c862e83c_0
- Behney, J. (2016). Perceived influence of cooperating teachers on edTPA performance. *Foreign Language Annals*, 49(2), 271-286.
- Berliner, D. C. (2013). Problems with value-added evaluations of teachers? Let me count the ways. *The Teacher Educator*, 48(4), 235-243.
- Betebenner, D., Braun, H., Corcoran, S., Darling-Hammond, L., Friedman, J., Goldhaber, D., . . . Rothstein, J. (2012). *Learning from recent advances in measuring teacher effectiveness*. Washington: Institute of Education Sciences.
- Borko, H. (2016). Methodological contributions to video-based studies of classroom teaching and learning: a commentary. *ZDM Mathematics Education*, 48(1-2), 213-218.
- Brady, M. P., Heiser, L. A., McCormick, J. K., & Forgan, J. (2016). Value-added models for teacher preparation programs: Validity and reliability threats, and a manageable alternative. *The Educational Forum*, 80(3), 339-352.
- Buchanan, J., & Schuck, S. (2016). Preparing a 'classroom ready' teacher: The challenge for teacher educators. In Gibbs, I. S. (Ed.), *Teacher Education: Assessment, Impact and Social Perspectives* (pp. 1-12). Hauppauge: Nova Science Publishers.
- Bunch, G. C., Aguirre, J. M., & Téllez, K. (2009). Beyond the scores: Using candidate responses on high stakes performance assessment to inform teacher preparation for English learners. *Issues in Teacher Education*, 18(1), 103-128.
- Campbell, C., Ayala, C. C., Railsback, G., Freking, F. W., McKenna, C., & Lausch, D. (2016). Beginning teachers' perceptions of the California Teaching Performance Assessment. *Teacher Education Quarterly*, 43(2), 51.
- Chizhik, E. W., Chizhik, A. W., Close, C., & Gallego, M. (2017). SMILE (Shared Mentoring in Instructional Learning Environments): Effectiveness of a lesson-study approach to student-teaching supervision on a teacher-education performance assessment. *Teacher Education Quarterly*, 44(2), 27-47.
- Clayton, C. D. (2018). Policy meets practice in New York State: Understanding early edTPA implementation through preservice candidates' eyes. *Teacher Education Quarterly*, 45(3), 97-125.
- Cochran-Smith, M., Piazza, P., & Power, C. (2013). The politics of accountability: Assessing teacher education in the United States. *The Educational Forum*, 77(1), 6-27.
- Darling-Hammond, L. (2012). The right start: Creating a strong foundation for the teaching career. *The Phi Delta Kappan*, 94(3), 8-13. doi:10.1177/003172171209400303
- 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., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan*, 93(6), 8-15.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. H., Rothstein, J., National Academy of, E., & American Educational Research, A. (2011). *Getting Teacher Evaluation Right: A Background Paper for Policy Makers*. Retrieved from <http://files.eric.ed.gov/fulltext/ED533702.pdf>
- Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: the potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25(3), 179-204. doi:10.1007/s11092-013-9163-0

- Denner, P., Norman, A., & Lin, S. (2009). Fairness and consequential validity of teacher work samples. *Educational Assessment, Evaluation and Accountability*, 21(3), 235-254.
- Forman, K., & Markson, C. (2015). Is "effective" the new "ineffective"? A crisis with the New York State teacher evaluation system. *Journal for Leadership and Instruction*, 14(2), 5-11.
- Ginsberg, R., & Kingston, N. (2014). Caught in a Vise: The Challenges Facing Teacher Preparation in an Era of Accountability. *Teachers College Record*, 116(1).
- Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, 68(4), 377-393.
- Hébert, C. (2018). Assessment in the clinical experience: Student teaching and the edTPA. *Teaching Education*. doi: 10.1080/10476210.2018.1505842
- Hébert, C. (2017). What do we really know about the edTPA? Research, PACT, and packaging a local teacher performance assessment for national use. *The Educational Forum*, 81(1), 68-82.
- Humphry, S., & Heldsinger, S. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher*, 43(5), 253-263. doi:10.3102/0013189X14542154
- Jackson, E. D., Kelsey, K. D., & Rice, A. H. (2018). A case study of technology mediated observation in pre-service teaching experiences for edTPA implementation. *NACTA Journal*, 62(1), 1-10.
- Jesson, J. K., Matheson, L., & Lacey, F. M. (2011). *Doing your literature review: Traditional and systematic techniques*. London: SAGE.
- Kim, N. (2019). Challenges of teaching and preparing edTPA. *Art Education*, 72(1), 21-27.
- Kissau, S., Hart, L., & Algozzine, B. (2019). Investigating the impact of edTPA professional development on classroom practice and student teaching experience. *Journal of Teacher Education*, 70(2), 102-114.
- Krise, K. (2016). Preparing the standardized teacher: The effects of accountability on teacher education. *JCT*, 31(2), 24-32.
- Ledwell, K., & Oyler, C. (2016). Unstandardized responses to a "standardized" test: The edTPA as gatekeeper and curriculum change agent. *Journal of Teacher Education*, 67(2), 120-134. doi:10.1177/0022487115624739
- Lim, W., Stallings, L., & Kim, D. J. (2015). A proposed pedagogical approach for preparing teacher candidates to incorporate academic language in mathematics classrooms. *International Education Studies*, 8(7), 1.
- Mayer, D. (2013). Policy driven reforms and the role of teacher educators in reframing teacher education in the 21st century. *Waikato Journal of Education*, 18(1), 1-18.
- Mayer, D., Allard, A. C., Moss, J., & Dixon, M. (2015). Initial teacher education and assessment of graduates in Australia. In J. C. K. Lee & C. Day (Eds.), *Quality and change in teacher education* (pp. 149–162). ProQuest Ebook Central: Springer International Publishing.
- Noel, A. M. (2014). Teacher Performance Assessment (edTPA): An instructor's development and evaluation of an Embedded Signature Assessment in an Early Childhood literacy course. *Journal of Early Childhood Teacher Education*, 35(4), 357. doi:10.1080/10901027.2014.968299
- O'Brien, A., & McGuckin, C. (2016). The systematic literature review method: Trials and tribulations of electronic database searching at doctoral level. In *SAGE research methods cases*. London: SAGE.
- Okhremtchouk, I., Seiki, S., Gilliland, B., Ateh, C., & Wallace, M. (2009). Voices of pre-service teachers: Perspectives on the Performance Assessment for California Teachers (PACT). *Issues in Teacher Education*, 18(1), 39-62.
- Onwuegbuzie, A., & Frels, R. (2016). *Seven steps to a comprehensive literature review: A multimodal and cultural approach*. London: SAGE.
- Palardy, G. J., & Peng, L. (2015). The effect of summer on value-added assessments of teacher and school performance. *education policy analysis archives*, 23, 92. doi:10.14507/epaa.v23.1997

- Paugh, P., Wendell, K. B., Power, C., & Gilbert, M. (2017). 'It's not that easy to solve': edTPA and preservice teacher learning. *Teaching Education*, 29(2), 147-164. doi:10.1080/10476210.2017.1369025
- Pecheone, R., & Chung, R. (2007). Technical report of the Performance Assessment for California Teachers (PACT): Summary of validity and reliability studies for the 2003-04 pilot year. Retrieved from http://www.pacttpa.org/_files/Publications_and_Presentations/PACT_Technical_Report_March07.pdf.
- Pecheone, R., & Whittaker, A. (2016). Well-prepared teachers inspire student learning. *The Phi Delta Kappan*, 97(7), 8-13.
- Pufpaff, L. A., Clarke, L., & Jones, R. E. (2015). The effects of rater training on inter-rater agreement. *Mid-western Educational Researcher*, 27(2), 117.
- Ratner, A. R., & Kolman, J. S. (2016). Breakers, benders, and obeyers: Inquiring into teacher educators' mediation of edTPA. *education policy analysis archives*, 24, 35. doi:10.14507/epaa.24.2112
- Roosevelt, D., & Garrison, J. (2018). "Yet the root of the matter is not in them": Reclaiming the lost soul of inspirational teaching. *Educational Theory*, 68(2), 177-195. doi:10.1111/edth.12305
- Sandholtz, J. H. (2012). Predictions and performance on the PACT teaching event: Case studies of high and low performers. *Teacher Education Quarterly*, 39(3), 103-126.
- Sato, M. (2014). What is the underlying conception of teaching of the edTPA? *Journal of Teacher Education*, 65(5), 421-434. doi:10.1177/0022487114542518
- Seymour, C. A., Burns, B. A., & Henry, J. J. (2018). Cooperating teachers stakeholders in the edTPA? . *Issues in Teacher Education*, 27(1), 41-56.
- Snyder, J. (2009). Taking stock of performance assessments in teaching. *Issues in Teacher Education*, 18(1), 7-11.
- Stewart, A. R., Scalzo, J. N., Merino, N., & Nilsen, K. (2015). Beyond the criteria: evidence of teacher learning in a performance assessment. *Teacher Education Quarterly*, 42(3), 33. Stillman, J., Anderson, L., Arellano, A., Lindquist Wong, P., Berta-Avila, M., Alfaro, C., & Struthers, K. (2013). Putting PACT in context and context in PACT: Teacher educators collaborating around program-specific and shared learning goals. *Teacher Education Quarterly*, 40(4), 135-157.
- Swanson, P., & Hindebrandt, S. (2017). Communicative learning outcomes and world language edTPA: Characteristics of high-scoring portfolios. *Hispania*, 100(3), 331-347.
- Teacher Education Ministerial Advisory Group. (2014). *Action now: Classroom ready teachers*. Canberra: Australian Government.
- Tuck, E., & Gorlewski, J. (2016). Racist ordering, settler colonialism, and edTPA: A participatory policy analysis. *Educational Policy*, 30(1), 197-217. doi:10.1177/0895904815616483
- Wahl, R. (2017). What can be known and how people grow: The philosophical stakes of the assessment debate. *Studies in Philosophy and Education*, 36(5), 499-515. doi:10.1007/s11217-016-9525-4
- Wilkerson, J. R. (2015). The need for instructional sensitivity and construct clarity in PACT: A commentary on "Examining the Internal Structure Evidence for the Performance Assessment for California Teachers". *Journal of Teacher Education*, 66(2), 184-192. doi:10.1177/0022487114561029