

An Analysis of the Structure and Assessment of Standards for Teacher Candidates and Programs ⁱ

DRAFT (please do not cite or quote without permission)

Chapter to appear in *Preparing Teachers to Implement College and Career Readiness Standards: Integrating Research, Policy, and Practice*, will be published by Sense Publishers (Rotterdam, The Netherlands) in 2013.

Edited by: Hersh C. Waxman, Beverly L. Alford, Danielle Brown, and Kayla Rollins, Texas A&M University

Kip Tellez

University of California-Santa Cruz

Abstract.

Educational standards are largely ignored until policy makers decide to measure students (or teachers) against them and then use the results to make important decisions. Teacher education was swept into the standards movement over two decades ago and today each and every state has a comprehensive set of standards for teachers; many also have specific standards for initial licensure. In this chapter, I have three goals. First, to my knowledge, no one has completed a textual analysis of the discourse and argumentation we find in state teacher education standards. I will conduct a qualitative analysis of all 50 state standards seeking to find and analyze common elements. Second, perhaps most important, I explore the development of teacher performance assessments developed from recent state standards. In particular, I focus on the development of the Performance Assessment for California Teachers, which is now the model for a nationwide performance assessment. Finally, I explore the efforts of the National Center for Teacher Quality to grade each of the nation's teacher education programs using its own set of standards, developed mostly in isolation from teachers and teacher educators.

Exploring the Development of Standards in Teacher Education

The teacher cannot produce in the learner a given experience without having first produced in him[her]self that experience...The teacher passes into some act or state of experience, and the pupil rises, at the touch of the teacher, into the same experience. (p. 3-4).

Arnold Tompkins (1903), *The Philosophy of Teaching*. Boston, MA: Ginn and Company.

Like many readers of this volume, I have served on several panels or committees during my career whose goal was to determine the knowledge and skills needed by teachers. I have been especially involved with the development of standards for prospective teachers, which must, by extension, be reflected in programs that prepare teachers. From my experience, committees charged with establishing teaching standards can typically determine its list in a day or two, depending on the size of the group and the time allotted for the task. But regardless of the committee membership or external conditions, all arrive at more or less the same 15-20 “must haves” on the list.

The striking sameness of teaching standards, which I will demonstrate later in this paper, has done little to prevent their proliferation. But why have standards become so ubiquitous? I would suggest the turn from twentieth to twenty-first century is likely to be remembered for the intense attention paid to accountability in education. In fact, unlike other reforms in education that were based on some curricular (e.g., Sputnik) or instructional (e.g., the “reading wars”) foundation, the current reform effort rests on the measurement of learning itself, resulting in an irritating but common begging of the question.

But in order to measure or evaluate learning, we must decide what is to be taught. Although standards alone will not produce a vast accountability schemes such as those we find in K12 education (e.g., No Child Left Behind) or in teacher preparation, it is impossible to create a reliable and valid assessment of one’s knowledge and skills without first developing standards by which we can judge whether the standards have been met.

As contemporary policymakers demanded assessments to gauge the overall

achievement of public school students in the US, a concomitant effort sought a narrowing of the achievement gaps (e.g., between rich and poor). This linkage and the reforms it has inspired have not been well received by most educators and educational researchers. One needs only to raise the topic of No Child Left Behind (NCLB) in any group of educators to find evidence of resistance, even outright anger over the level of surveillance now found in the nation's public schools (Pease-Alvarez, Samway, & Cifka-Herrera, 2010).

Although each educator will object to the accountability movement in his or her own way, many will suggest that standards and their high-stakes assessment push important purposes of schools aside at the expense of increasing academic achievement as measured by select-response, standardized tests. In this criticism they are joined by Jameson, whose classic work *Postmodernism: The Cultural Logic of Late Capitalism* (Jameson, 1991) notes that what was once (i.e., before modernity) conceived as marvelously, complex intellectual growth has come to be reconceived in utterly concrete terms, as the equivalent of a test score; much as the value of a corporation—an assemblage of human beings' skills and talents, successes and failures—is reduced to its exchange value at the moment.

Educators will generally disagree with the policies and accountability schemes pressed upon them by policy makers and others, but that does not stop the former from making mandates. Mitchell (1984) points out exactly how this happens:

“Policy makers cannot teach students, and they cannot manage school programs unless they change jobs and join the school staff. Hence, they must find policy mechanisms that may be used to indirectly restructure the school system through influencing the actions of educators and students by changing the cultural and material environment...” p. 154 (Mitchell, 1984).

In the current climate, the most common policy “mechanism” to influence the actions of educators in our era is the development of standards and their assessment. In fact, I have argued that standards have become a near addiction for politicians and other policymakers (Tellez, 2003), and it is easy to see why. Politicians can commission a panel and charge it with creating high learning or performance standards. The immediate benefit of this strategy is that it creates a newsworthy event. The secondary benefit is that convening a standard setting panel costs very little when compared with other reforms such as new materials or technology, lower class sizes, or professional growth among educators. With the panel's recommendations in place, wise politicians

can claim that they sponsored legislation that demands higher standards for learners or teachers (resulting in a fool-proof stump speech), and then sit back while educators do their best to meet those standards, most often without the aid of additional resources. If the standards are not met, the blame lies with the educators and the systems in which they work. The standards, now immutable, sit in judgment of the actors. Again, the wise politician can win once more avoid complicated questions about resources and simply restate the value of the high standards education needed to keep the nation competitive, drive our economy, and so on.

To place this movement within a theoretical frame, I turn to the work of Bourdieu, whose overarching theory of social structures suggests that a by-product of cultural invention in a bureaucratic ageⁱⁱ is a numbing sort of resignation to the way things are, a phenomenon he called “habitus.” For Bourdieu, habitus produces a common a set of expectations that structure our predispositions. These expectations become transparent, but are capable of quick reproduction and, in his words, “extort what is essential while seeming to demand the insignificant” (1990, p. 69). To my mind, this describes precisely the habitus produced by accountability schemes such as NCLB. We become beholden to the standards and grow accustomed to their exacting but largely meaningless rules, and yet those opposed to the scheme fail to generate enough resistance to alter the course.

Of course, we cannot separate the standards and accountability movement in K12 from the parallel activity in teacher education, and my task in this chapter is to consider standards in teacher education. As an introduction, I’d like to share a paragraph I reprinted in an earlier paper on this topic (Tellez, 2003). Robert Roth wrote these prophetic words nearly 20 years ago:

The domain encompassed by standards, accreditation, licensure, and certification is being reconstructed in fundamental ways. The impact is the creation of an entire historical era in the profession, equal in significance to other major periods in education history such as the development of normal schools. The standards movement is so pervasive and powerful that it may appropriately be termed the *Era of Standards*. The movement in general may be characterized by several salient features. Among these are a deep-seated and growing distrust of teacher education; a change in the locus of control, with national policy emerging as a dominant influence; restructuring of licensing and governance;

and reconceptualizing the nature of standards, with performance and outcomes assuming a preeminent role. (Roth, 1996, p. 242).

One year after this chapter was published, Roth passed away, and the teacher education community lost an important voice in the debate on standards and accountability. To my mind, his presages are nothing short of astonishing: it's almost as if he *willed* these events to happen, and I often wonder what Bob would predict for us now if he were alive.

One of the key points Roth made was his final sentence in the passage: that outcomes would become preeminent as standards in teacher education became operationalized. This link is crucial because standards have little meaning or force until they are measured. After determining the skills we expect of teachers, then we can go about the process of deciding whether a particular teacher education program is capable of developing this knowledge. In an edited book that represented Roth's last scholarly contribution to the field (Roth, 1998), Frazier politely extolls the view that standards are fine by themselves and even required for public confidence, but much less justified is the specificity of those standards and their fixation in programs. To quote, "Teacher standards, objected to by some reformists, are a legitimate expression of a public expectation. Less justified is the legislative specification of the program elements (Frazier, 1998, p. 141) (Roth, 1998).

Of course, this is precisely what has happened. Through state accreditation or other accountability strategies, program elements have become mandated. State agencies are typically not at fault; more often, it is the politician's zeal for standards and accountability that have resulted in such mandates. But it is important to note that the standards movement in teacher education has not been fomented by outside forces alone.

In his excellent book, *Teaching by Numbers*, Taubman documents the shift towards standards in teacher education, suggesting that teacher educators have been "seduced" into a movement that is now strangling them (Taubman, 2009). He points to David Imig's comments, made in 2005, in which Imig admits that we have achieved national attention and a perceived seriousness of purpose by developing teaching standards. And while it's true that we have gained a measure of legitimacy by "guaranteeing our products" through standardization, some of the consequences have been negative and unintended. Standardization, Imig argued, will mostly benefit

commercial interests in teacher education, who can now market products to be sold nationwide, allowing for economies of scale, that result, of course, in additional profit. In a more recent paper, Imig and his colleagues (Imig, Wiseman, & Imig, 2011), remark that the organization he led for 35 years, the American Association of Colleges of Teacher Education (AACTE), has now joined forces with several partners, the most important of which is the for-profit, publicly traded Pearson Assessments, whose subsidiaries include the Pearson Teacher Education Group, which is currently developing a nationwide performance assessment of teacher candidates. What has become evident during our “era of standards” is that developing standards is mostly symbolic; developing instruments for evaluating learners (or teachers) based on those standards is where the resources can be generated. I’ll return to Taubman’s analysis of standards in teacher education later in the final section of the chapter.

In the remaining sections, I have three tasks. First, I will explore the content of teaching standards across all 50 states, looking for common elements and interesting outliers. I next examine the development and implementation of the Performance Assessment of California Teachers, a project that I have participated in for over 12 years. The final section offers an in depth analysis of the National Center Teacher Quality, a new policy group with an innovative set of standards and compelling criteria for judging the effectiveness of teacher education programs.

I admit that my chapter has a distinctly narrative quality and an idiosyncratic flow that mirrors my own understanding of the topics. However, I think that readers of this volume will find it useful to match their experiences to mine, consider whether they agree or disagree with my conclusions (I’m always interested in hearing alternative viewpoints), and decide if my review and analyses might help to move the field of professional teacher education forward by analyzing what we’ve done (both good and bad) and what’s being done to us.

An exploration of teaching standards

I’ve been a teacher educator for 27 years, and in that time, I’ve seen several waves of reforms—perhaps better described as shifting themes—and wondered why such ideas and mandates come and go in teacher education. A particular favorite is to recall is the attention to reflection in teaching, a view very popular among teacher educators in the 1980s and early 1990s. I won’t share much about the varying

conceptions of reflection in teaching because I have written on this topic elsewhere (see Tellez [2007] for a review), but I would like to note the extraordinary influence of this view on teacher educator, and also note the equally extraordinary contrast between a reflective stance on teaching and one that defines teaching by standards.

We can credit Donald Schön's wide ranging work for virtually remaking our understanding of the professions, claiming that teaching, like many other professional fields, is not teachable, only coachable. Professions, he argued, demand a holistic understanding, guided by feedback from the conditions of the work. Professions cannot be learned in an atomistic way--by first performing individual skills, only later putting them all together. Therefore, given the non-routine nature of teaching, there is no proper way to describe or prescribe it with any degree of precision, which Schön shares in this passage:

. . . what teachers can *tell* you about their work--to themselves as well as to novice teachers--must fall short of what they know tacitly and show in action. Between a rule like "First get the kids' attention" and its concrete application in skillful teaching, there is always a gap of meaning. (1983, p. 202)

Contrast this view, so influential just 20 years ago, with a conception of teaching based on standards, which, by definition, suggests that a bright line can be drawn between standards—rules, in fact—and actual teaching. A reflective view of teaching, at least as Schön conceived it, argues the obverse.

In retrospect, the claims made about reflective teaching appear to me to have been our last chance at avoiding standards and an overarching accountability in teacher education. In Schön we had a very thoughtful ally, a professor the Massachusetts Institute of Technology who was not an education professor, telling the wider academic community that teaching was far too complex to be codified or standardized, too subtle and nuanced to be measured with any accuracy. Teaching, we all came to agree, existed in a professional half-light, somewhere between a complicated craft and genuine artfulness, both constructs difficult to define and even more difficult to measure. Schön even compared the work of a teacher to a jazz soloist, a metaphorical invention that seemed to please everyone.

So how did we move from Schön's description of teaching to the state of the profession as described by Roth? Like Imig, Darling-Hammond (1999) has argued that

the proliferation in standards is our own doing. And in recent decades, several major reports calling for the professionalization of teaching have argued that teachers must take hold of professional standard-setting if teaching is to make good on the promise of competence to the general public (Carnegie Forum, 1986; Holmes Group, 1986). The argument was that professionals must define high standards for those entering the profession, set rigorous expectations, and then hold peers to these standards and expectations.

With the widespread development of standards, it seemed to me that it would be useful to have a catalog of the states' efforts in this area. If standards have grown in importance, what do they look like nationwide? In fact, I wondered if every state had developed teaching standards (the answer is yes), and if some common themes could be found.

In order to conduct my analysis of state teaching standards, I downloaded each of the state's preservice teacher credential standards. In some instances, I could not find standards for preservice teacher education and instead had to rely on the standards for practicing teachers. In the former cases, it was implied that preservice teacher education programs must abide by the standards for practicing teachers.

To assemble the table below, I created a text file containing all the state standards. I then used a text processing application (TextWrangler) to find instances of repeated words and phrases. I then selected a state standard that I believed was a good representation of the content. The following web site was useful in locating the states' standards

(http://www.ccsso.org/Resources/Digital_Resources/Web_Links_to_State_Teacher_Licensing_Standards.html) although I caution readers that several links led to empty pages.

Table 1. Common (or meta-) state standards, illustrative example, quoted text, and my analysis.

Meta-Standard	Sample State	Quoted Text from State Sample	Analysis
Knowing content	Alabama	To improve the learning of all students, teachers master the disciplines related to their teaching fields including the central concepts, important facts and skills, and tools of	This standard is almost always the very first mentioned. Although I could not find a standards list that explicitly prioritized one standard over another, the position of content

		inquiry; they anchor content in learning experiences that make the subject matter meaningful for all students.	knowledge in the lists appears to be an informal tool for emphasizing that this standard is most important (i.e., teachers <i>must</i> know their content).
Knowing classroom "management"	Kansas	The educator understands the principles of effective classroom management and can use a range of strategies to promote positive relationships, cooperation, and purposeful learning in the classroom.	Preservice teachers often report that they don't receive enough help in learning the effective tools of classroom management, a view corroborated by many school administrators of beginning teachers.
Knowing human development	Florida	Identify patterns of physical, social, and academic development of students; Identify motivational strategies and factors that encourage students to be achievement and goal oriented; Identify activities to accommodate different learning needs, developmental levels, and experiential backgrounds.	This standard seems obvious to everyone: If teachers do not understand the developmental landmarks of the children and youth they are to teach, then they will be unlikely to guide meaningful instruction. This knowledge is often imparted to preservice teachers in the form of Piagetian tasks, which, I would argue, are of dubious value.
Knowing teaching strategies for specific content	California	Candidates...demonstrate the ability to teach the state-adopted academic content standards for students in English-Language Arts (Grades K-8). They understand how to deliver a comprehensive program of systematic instruction in word analysis, fluency, and systematic vocabulary development; reading comprehension; literary response and analysis; writing strategies and applications; written and oral English Language conventions; and listening and speaking strategies and applications.	At the elementary teaching level, nearly all standards are detailed in the requirements for teaching literacy, including specifics that we don't find in other standards. In California's case, candidates must also pass a test of reading pedagogy, the Reading Instruction Competence Assessment, paper-and-pencil test titled largely towards a phonemic approach to literacy instruction. (see O'Sullivan, S., & Jiang, Y. H. [2002] or a review.
Knowing	Colorado	The teacher is skilled in	I don't exactly know when

technology		technology and is knowledgeable about using technology to support instruction and enhance student learning.	this standard became commonplace, but I could not find a single set that did not mention technology, albeit with varying degrees of emphasis.
Knowing how to teach diverse learners	Wisconsin	The teacher has a well-grounded framework for understanding cultural and community diversity and knows how to learn about and incorporate students' experiences, cultures, and community resources into instruction.	Interestingly, the California Standards are missing any mention of diversity. By contrast, Wisconsin, one of the least diverse large states, requires teachers to understand how to build on student culture to connect to student learning.
Knowing how to assess learning	Ohio	Teachers are knowledgeable about assessment types, their purposes and the data they generate. Teachers analyze data to monitor student progress and learning, and to plan, differentiate and modify instruction.	As a result of NCLB, test score data have grown in importance, and now many states have emphasized or re-emphasized the importance of teachers' capacity to develop and utilize the results of tests, both standardized and teacher-made, both formative and summative.
Knowing legal mandates	Michigan	Understand and uphold the legal and ethical responsibilities of teaching (e.g., federal and state laws and SBE [State Board of Education] policies pertaining to positive and effective learning environments, appropriate behavioral interventions, student retention, truancy, child abuse, safety, first aid, health, and communicable disease).	Ubiquitous standard, often listed with very little explanation.
"Knowing" ⁱⁱⁱ the moral obligations of teaching	Illinois	Candidates are aware of and reflect on their practice in light of research on teaching, professional ethics, and resources available for professional learning; they continually evaluate the effects of their professional decisions and actions on students,	This standard, also ubiquitous, is similarly underdetermined. It is also impossible to assess without observing actual candidate behavior, making it unique among the common standards.

		families and other professionals in the learning community and actively seek out opportunities to grow professionally	
--	--	---	--

As I predicted, the standards across states are more or less uniform, reflecting a traditional and somewhat conservative view of teaching. This result is predictable, but I was a bit disappointed to find that no state standard was even slightly tilted towards a critical stance of teaching, which we might expect given the influence of writers such as Paulo Freire and like-minded writers in the US. In every instance, teaching standards are written to ensure that teachers, both pre- and inservice, are beholden to state-sanctioned K12 standards. Teachers are never encouraged to develop curricula that emerge from student interests or concerns, even if such a curriculum matches the adopted curriculum.

As I analyzed the standards further, I found a few interesting examples that were represented by just a few states but bear mentioning nonetheless. For instance, the Massachusetts standards require that a teacher must convey “knowledge of and enthusiasm for his/her academic discipline to students.” I think that we would find wide agreement among educators on this standard, but I don’t see any obvious way to assess it. I was also interested in Standard Five (Virginia) that requires teachers to “collaborate with colleagues to improve instruction, assessment, and student achievement.” Here we have in policy a reflection of the professional learning community movement, but, again, the valid assessment of such a standard would be a challenge

What we find in the most common standards, however, is a set of teaching targets that can be assessed with varying degrees of difficulty. It is an easy task for a sophisticated assessment company such as the Educational Testing Service (ETS) to develop valid basic skills and content knowledge tests (e.g., Praxis) in a “paper-and-pencil” or computer-based select-response format (including some supply response or “essay” questions). On the other hand, assessing classroom “management,” for example, is best measured by examining a teacher’s performance; in this instance, a teacher may be able to report the correct procedures but fail utterly in the context of the classroom. Recalling Schön’s example, we want to know whether a teacher can “get the kids attention” in advance of instruction. Assessing this standard in a performance setting is not impossible, but it’s much more difficult than measuring whether a preservice teacher

knows the causes of the Civil War. The assessment of teaching performance is the topic of our next section.

However, before moving on to that discussion, I would like to mention an ambitious effort undertaken by the Association of Teacher Educators (ATE), a professional association of teacher educators somewhat smaller than the AACTE (800 institutions represented vs. 650, respectively). ATE has developed standards for teacher educators. (<http://www.ate1.org/pubs/uploads/tchredstds0308.pdf>). This effort is compelling because these standards are not NCATE-like markers designed to assess the programs that they work in, but designed to assess the teacher educators *themselves*. These standards are likely a response to the continuous criticism of teacher educators, who are often cited as the reason for teacher education's weak reputation; it's an ambitious goal and yet one with a clear logic: If we are going to make our teacher candidates prove themselves worthy of a license, then it stands to reason that teacher educators who are teaching and supervising them should undergo a similar assessment. A recent volume includes the testimony of several teacher educators who have considered how their own qualities measure up to the ATE standards (Klecka, 2009). It remains to be seen whether an assessment of the quality of teacher educators can keep the critics at bay.

Standards Made Concrete: The Performance Assessment for California Teachers

The establishment of standards as a tool for advancing a human endeavor cannot be realized until we devise a strategy for assessing those standards. But the development of assessment tools is not a trivial problem. To wit:

“There is still considerable confusion about the definition of clinical competence and most of the methods in use to define or measure competence have not been developed systematically and issues of reliability and validity have barely been addressed. The assessment of clinical competence remains almost universally accepted in the nurse education literature as a laudable pursuit yet there are aspects of it that remain at odds with the higher education of nurses.” (Watson, Stimpson, Topping, & Porock, 2002)

Although this paragraph was written a decade ago and in reference to an entirely

different professional field, the conditions for assessing clinical competence remain dubious across disciplines and fields.

Evaluating teacher performance must balance tensions between reliability and validity, the meaningful and the banal, and perhaps even the sacred and the profane. The metaphors we use to describe teaching itself signal such a predicament: If teaching is art, then it is probably impossible to measure; if it is a craft, then it might be measurable; if it is skill, then its measurement should be straightforward. But educational researchers, educators *a fortiori*, do not agree on the essence of teaching, and we are therefore bound to disagree on how to evaluate it.

I am not suggesting that issues of teacher assessment and evaluation must be necessarily binary or even that evaluating teaching is alone in creating such tensions. Nevertheless, teacher evaluation has complicated measurement in education by inviting important questions about whether its assessments are genuine and authentic (Télez 1996).^{iv} Teachers, in particular, are dubious that any evaluation of their work can capture the scope, detail, and subtleties of the profession (Heneman & Milanowski, 2003). Assessment specialists, on the other hand, recognize that all evaluation will always be underdetermined and include error. This is of course why psychometricians refer to any score or ranking as an *estimate* of an individual's true achievement or aptitude.

But the caveats by testing specialists on the very specific and limited uses of assessment data often fail to stop policymakers who enact laws or create policies that use assessments and the resulting data in unintended or invalid ways. And in our era of accountability, there seems to be no educational evaluation, no accountability scheme, or no augmenting of standards that politicians and policymakers do not embrace. And they are particular fond of creating laws that evaluate teachers. But evaluating the knowledge or skills of practicing teachers is very difficult, especially in those states where teachers are highly unionized. And even in "open shop" states, where unions are less powerful, history has shown that attempts to evaluate practicing teachers' teaching performance--and then use the data for merit or even termination decisions--rarely last (see Popham [1988] for an example).

What follows is an account of the development of teacher candidate standards (which I did not help to write) and a performance assessment (which I did). The context is the state of California, where I now work, but I also recall the development of the "performance assessment" in Texas, known as the ExCET tests. I won't address the latter assessments because they are paper-and-pencil tests of pedagogical knowledge,

which are largely based on the reading of teaching cases, and then choosing the correct answer; the EcXET exams were a performance assessment only in the loosest interpretation of the term.

In the late 1990's, when legislators in the state of California determined to rewrite the standards for initial teacher licensure, they took a very different strategy than Texas had. Indeed, the legislation, known as Senate Bill (SB) 2042, included a mandated teacher performance assessment designed to evaluate preservice teachers based on the new standards. If a new teacher could not pass the performance evaluation, then no credential could be given. Like policymakers everywhere who mandate standards and the accountability schemes, the psychometric realities of creating a valid and reliable performance assessment that could stand the bright light of such high stakes was not a concern, and legislators marched on. With the new Teacher Performance Expectations (<http://www.ctc.ca.gov/educator-prep/standards/adopted-TPEs-2013.pdf>) and the mandate to assess them, conceptualizations of teaching, in California at least, had moved, in a decade or so, from Schön's reflective practitioner, to standardization, to a mandate to assess those standards in a "pass or else" performance assessment.

I don't mean to be glib or necessarily pessimistic about standards and assessments, but the transformation has surprised me. And when we first read the SB 2042 legislation, which was supported by both sides in two very divided state houses, it struck me that the Democratic side had succumbed to the "higher standards" refrain I have mentioned, while the Republican side believed that they had chipped away at the "monopoly" of certificate providers, arguing that if individuals who want to become teachers can pass a performance assessment without formal teacher preparation, then why can't they be given a credential by simply passing the assessment, without the aid of a professional preparation program? The latter has not, or not yet, happened.

Missing from the legislation was an admission that creating a performance assessment that could withstand the validity challenges in a high-stakes environment would not be an easy task. And yet each of the roughly 250 teacher preparation programs throughout California, regardless of their location, size, or student population, would now be required to assess whether their credential candidates have met the defined California state standards of teaching expectations. Teacher preparation programs in California have the choice to use either the state-sponsored Teaching Performance Assessment (TPA) system developed by the California Commission on Teacher Credentialing (CCTC), in partnership with the ETS, or an alternative

assessment that meets the CCTC's assessment quality standards. In response, the Performance Assessment for California Teachers (PACT) consortium designed and constructed subject-specific performance assessments modeled after the portfolio assessments of the Connecticut State Department of Education, the Interstate New Teacher Assessment and Support Consortium, and the National Board of Professional Teaching Standards. Initially, in 2001, the Consortium consisted of 12 professional teacher preparation programs in 8 University of California campuses, 2 California State Universities, Stanford University, and Mills College. Currently, it includes 30 universities (both public and private, large and small), 1 district internship program, and 1 charter school network.

Even before ETS had completed their assessment system, the response from the state's teacher education faculty was swift and resolute: The performance assessment was an infringement on faculty autonomy and program integrity. Further, the California State University system concluded that the implementation of the CalTPA was tantamount to an unfunded mandate, and the system's 22 campuses initially refused to require the performance assessment until the state legislature provided funding for its implementation. As the date for mandatory passage grew closer, it became clear that the performance assessment would indeed be required for all credential earners and that no funds would be appropriated. The law was clear, and in 2007, each of the state's teacher education programs had to implement one of the performance assessments approved by the CCTC, which now stands at four.

Yet in spite of the objections, some research evidence suggests that teacher education programs are in need of a powerful summative evaluation of candidates. For example, Raths and Lyman (2003) found that far too many student teachers receive weak and positively biased formative evaluations throughout their program and yet earn a license because these formative evaluations fail to coalesce into a negative summative appraisal. The PACT and other summative, performance assessments may prevent such instances.

And objections from teacher educators on the grounds that the mandated performance assessment would be ruinous for teacher education seem not to have materialized. For instance, Hafner and Maxie, (2006) found that most teacher educators in California believed they had strong programs before SB 2042, and that the assessment had sacrificed attention to equity and, as predicted, required additional resources to administer, but had not altered their programs significantly. I should note

however, that this study was conducted when the performance assessment was in place before the performance assessment became mandatory.

As PACT has become routine, attention has turned away from broad critiques of summative assessments of teaching to studies of PACT itself. The most widely cited study thus far (Pecheone & Chung, 2007) investigated several issues related to PACT, including content validity, bias and fairness, construct validity, criterion-related concurrent validity, and score consistency and reliability. They found evidence for PACT's content validity, demonstrating consistency with the California Teaching Performance Expectations. No systematic biases were found based on the type of classroom in which the candidate completed PACT percent of English Language Learner students in their classrooms, grade level taught (i.e., elementary versus secondary), students' academic achievement level, or months of previous paid teaching experience. They did, however, find that females scored significantly higher than males and that candidates teaching in high socio-economic/suburban schools scored higher than candidates teaching in low socio-economic or urban/inner-city schools. In terms of construct validity, through exploratory factor analysis of the pilot data from the Elementary Literacy Teaching Event, they found that two distinct factors emerged—one for Planning, Instruction and Academic Language and another for Assessment and Reflection—indicating that at least some of the domains are tapping into distinct constructs of teaching.

Other studies have focused on aspects of rater reliability evidence. Porter (2010) examined the inter-rater reliability of PACT scores and found poor to moderate evidence for consistency in rater scores. Several quasi-validation studies have used candidate responses to the PACT Teaching Event in specific or holistic evaluations of teacher candidates' performance. Bunch, Aguirre, and Tellez (2009) conducted a qualitative case study using submitted materials for eight teacher candidates' Elementary Mathematics Teaching Events to further assess their Academic Language ability—ability to teach and meet the needs of linguistically diverse students in both their academic and English language vocabulary. They found the PACT Teaching Event tasks provided useful information in evaluating teacher candidates on this important, but often overlooked, teaching skill.

Sandholtz and Shea (2012) explored the relationship between supervisors' predictions and candidates' performance on PACT. The findings indicate that university supervisors' perspectives did not always correspond with outcomes on the performance

assessment, particularly for high and low performers. Though this result might be expected given the greater variability and lower reliability of scores at the ends of the distribution, the study represents an effort to triangulate among different data sources. Similarly, Darling-Hammond, Newton and Wei (2010) have argued that the PACT should be used in concert with several other measures of candidate learning. My colleagues and I (Duckor, Castellanos, Téllez & Wilson, 2013) explored the dimensionality of the assessment, using the Partial Credit Model (PCM), which models item responses to polytomous items. The PCM is the polytomous version of the simple Rasch model for dichotomous items. Specifically, the PCM models the probability of going from score level j to $j + 1$, such as 2 to 3, given the examinee has completed the previous step for each item. Whereas our research confirms that PACT is reliably discriminating candidates' scores across domains, we do raise some concern regarding the internal structure, noting that PACT might be measuring candidates' analytic discourse capacities rather than their actual performance in teaching.

Finally, as part of a PACT study group composed of University of California teacher educators, I conducted a very simple test of concurrent validity, computing simple zero-order correlations between PACT scores, university supervisor ratings of preservice teachers, and the ratings of cooperating teachers for the 85 students who completed the UC Santa Cruz program in 2011. For the university supervisors and cooperating teachers, the correlation between their ratings was statistically significant ($p < .01$) at .41. The PACT, on the other hand, correlated weakly with supervisors and cooperating teachers at .26 ($p < .05$) and .03, respectively.^v As an early and very preliminary test of concurrent validity, these findings are surprising and perhaps worrisome.

I do not claim that my tiny data set should cause us to reconsider the validity of the PACT. Most importantly, we do not know if the PACT is a better predictor of genuine skill in teaching than human ratings, although I suspect not. But even if we can imagine that the PACT and other performance measures lack validity, what might be the consequences? Specifically, how would policymakers respond. My guess is that the inventive (and perhaps self-serving) policymaker will turn to student achievement data linked to preparation program. In fact, I suspect this may happen regardless of the fate of performance assessments. Much as value-added scores have come to dominate the conversation around in-service teacher quality (Lockwood et al., 2007), policymakers will likely take this strategy for determining which of their states' programs are effective.

It's a logical step, but one that I will argue is impossible for several reasons. First, program preparation, no matter how effective, has but a limited influence. In fact, I would compare it to the principle of physics known as half-life, an exponential function often used to describe the decay of radioactive material. By the strict formula, in the first period (the first year of teaching, in our case), the effect of a teacher education program can assumed to be at its maximum. But for each year thereafter, its effects decrease by half, so that in five to six years, only 1-3% of its effects can be found, although its influence never disappears entirely. And I would suspect that many of the lasting effects have little to do with what a candidate learned in the program.

Second, school districts themselves exert a powerful influence on a beginning teacher's practice. Nearly all have their own professional development programs designed to push curriculum and instruction towards their own goals. Beginning teachers, who have yet to earn permanent status in the district and thus eager to please administrators, typically take their district's professional development and curricular goals very seriously, even if it contradicts the foundational beliefs promoted in their preparation program (Brouwer & Korthagen, 2005).

Third, I find two general objections to VAM models. Using student gain scores as the only measure of teacher effectiveness ignores the importance of *multiwave* data (Willett, 1989). I would also question the value of the rankings produced by VAM models. We have evidence that the teacher effects found in VAM models are normally distributed in any given year. Is there any valid reason to sort two teachers, whose percentile ranks are 50th and 51st? How could we possibly distinguish between these two teachers? One did rank higher than the other, but the difference is so small as to be insignificant. Measuring the quality of beginning teachers using achievement data presents far too many challenges, but these challenges may not prevent policymakers from demanding it.

I conclude this section not by forecasting the wide adoption of VAM tools for evaluating teacher education programs, but point to the expanded use of performance assessments for teacher candidates, in spite of much evidence of their validity. AACTE, through its partnership with Pearson Assessments^{vi}, is moving quickly to establish the "edTPA" as a national performance assessment for teacher candidates. I have not participated in the edTPA, but its essential design is similar to the PACT (Pecheone, personal communication, 2010), and I am confident that with time and the vast resources behind the edTPA effort that its measurement properties will be deemed

sufficient. What strikes me most about the edTPA effort is how its national reach and standards-based assessment parallels the current Common Core effort. Is teacher education forever bound to the K12 policy weights and pulleys?

“We are all participating, whether we want to or not”: The efforts of the National Center for Teacher Quality.

Professional behavior may be defined in terms of four essential attributes: a high degree of generalized and systematic knowledge; primary orientation to the community interest rather than to individual self-interest; a high degree of self-control of behavior through codes of ethics internalized in the process of work socialization and through voluntary associations organized and operated by the work specialists themselves; and a system of rewards (monetary and honorary) that is primarily a set of symbols of work achievement and thus ends in themselves, not means to some end of individual self-interest. (Barber, 1963)p. 672)

As if some teacher educators were not already frustrated with the mix of standards and accountability procedures, we now behold a wholly new and different entity developing standards for teacher education, one that is in direct competition with state accrediting bodies, NCATE, and all other collaborations established “through voluntary associations organized and operated by the work specialists themselves” (see Barber above). Known as the National Center for Teacher Quality (NCTQ), this organization exists outside of most traditional authoritative sources in teacher education, engaging in a strategy of evaluating teacher education that some have described as “evaluation by extortion.”

Who is NCTQ and why are they evaluating teacher education? A review of the membership on NCTQ’s Board of Directors and Advisory Board reveals several obvious choices if one wished to assemble a group critical of traditional professional school teacher education (e.g., Chester Finn, Eric Hanushek, Frederick Hess, E.D. Hirsch, Wendy Kopp, Michael Podgursky, Michelle Rhee) but also some surprises, including Suzanne Wilson, (Chair of the Department of Teacher Education at Michigan State University) and three K12 teachers, one of whom attended a teacher education program that NCTQ rates as “weak.” Among the funders I find several obvious choices (e.g.,

Abell Foundation) but others who have supported university-based educational researchers (e.g., the Heinz Endowments). NCTQ's president is Kate Walsh, an established critic of traditional teacher preparation (see Walsh & Podgursky, 2001). She leads a professional staff of about 20.

The goal of NCTQ includes advocacy for "reforms in a broad range of teacher policies at the federal, state and local levels in order to increase the number of effective teachers...with a research agenda that has direct and practical implications for policy." (<http://www.nctq.org/p/about/index.jsp>). Unlike some critics of professional school teacher education, NCTQ argues "that formal teacher preparation can and should add value" (<http://www.nctq.org/p/>)

Among its chief initiatives is a thorough cataloging and ranking of the nation's teacher preparation programs, based on standards that NCTQ developed. Perhaps the most controversial aspect of the evaluation is the manner in which NCTQ has gone about gathering the data for their reports. First, NCTQ invites preparation programs to send them a range of data and data sources, including course syllabi, admissions criteria, and other data that will assist them in rating the program against their criteria. But if an institution refuses to share the requested data, NCTQ promises to send a Freedom of Information Act (FOIA) request to the institution and force it to provide the data. And if the institution requires a fee to assemble the data, and if NCTQ believes that the fee is unreasonable, NCTQ places them on a list of "most secretive" institutions. fn

I recall vividly the meeting of the University of California Deans and Directors of Teacher Education when NCTQ's request was discussed. Some were adamant that we should not participate and not send any data, awaiting NCTQ's request via the FOIA; others pointed out that we would be participating regardless.^{vii} In the end, most of UC institutions decided to send the requested data, if only to be sure that NCTQ had the correct information.^{viii} But never before had anyone heard of an evaluation conducted in this way. Because NCTQ was not affiliated with any university, no informed consent was required. And because no one at NCTQ is a member of a professional organization (at least I could find no evidence) such as American Educational Research Association, they are not bound by the ethical standards required of members of a research or evaluation professional organization, which, in all cases, allows evaluation participants to decline participation or discontinue their involvement at any time without consequences.

NCTQ has taken, some might say stolen, teacher education's attention (or is it now a compulsion?) with standards, coopted it, and fashioned it into a tool for evaluating

teacher education programs, whether or not they agree with NCTQ's standards.

The most complete state-wide report from NCTQ details the state of Illinois' teacher preparation programs ("examine in unprecedented detail"), and the harshest criticism is reserved for three programs operated by Olivet Nazarene University, because, in the words of NCTQ, "despite repeated requests, the institution would not cooperate." The rest of the report, I believe, might be confusing for the intended reader, but not because the criteria are unclear or that NCTQ was necessarily capricious or biased in their assessments (the standards themselves might be tilted towards NCTQ's biases^{ix}, but I believe they represent, more or less, fair criteria for judgment, and resemble the teacher education program standards we find in many states. But after a hundred pages of ratings, comments, and evidence, we learn primarily that Northwestern University is the winning program in Illinois: it got an A-.

As I read through the Illinois report, I grew confused and a bit tired. And I wondered if a prospective teaching candidate or school administrator would be similarly nonplussed.

NCTQ claims that it is providing important information for consumers, both prospective teachers seeking a quality program and school administrators looking to hire the best teachers. I understand the intent here, but I simply don't think such a report will have much influence. First, prospective teachers rarely conduct a statewide search for the best teacher education program. In my experience, they seek out the least expensive route to the license, regardless of the quality. They seem to know in advance that schools and school districts are not overly concerned with the quality of their credentialing institution; rather, they are most interested in how individual preservice teachers perform in the practicum (student teaching). A strong letter from a cooperating teacher, especially one who works within the hiring school district, will be given far more weight than the license earner's program affiliation. Second, I find that school hiring panels generally do not care where candidates received their credential. Their experience tells them that there is far more variability among candidates from a single program than any differences that might exist across programs. In an article published in *Educational Leadership*, which is read widely by school administrators, Stronge and Hindman (2003) point out dozens of characteristics and experiences to look for when hiring teachers, but credentialing institution does not even make the list. A recent study confirmed the findings of the Stronge and Hindman work (Harris, Rutledge, Ingle, & Thompson, 2010).

I admit that this lack of attention to quality of credentialing institutions may point out the very problem that NCTQ is trying to solve, but I'm not sanguine that NCTQ will have any success in getting school officials who make hiring decisions to care about where or how their teachers were prepared. Perhaps it's plain arrogance that leads principals and other educators to believe that they can discern the best applicant from a 45 minute interview and a few letters of reference while ignoring other important variables (which might include the quality of candidate's institution) but they have been using their methods for many years, and they will not alter practices as a consequence of NCTQ's reports.

And in high-need content areas such as mathematics and sciences, as well as special education and bilingual education, NCTQ's effort will have little or no effect. Teacher shortages in these fields suggest that schools will likely offer a position to all qualified applicants, regardless of where they earned their credential. In addition, hard-to-staff schools, especially those in rural regions of the country, where general teacher shortages are common, the schools simply cannot be selective, so that the quality of the licensing institution will have no bearing on hiring decisions. I also have some general comments to share on NCTQ's task, which I will connect to standards and NCATE to conclude the chapter.

As I read the NCTQ website repeatedly, I found the rhetoric to be an odd mix of neo-liberal and neo-conservative ideals. For instance, NCTQ makes arguments to deregulate teacher education by citing Teach for America as a model worth considering and perhaps replicating (see <http://www.nctq.org/p/tqb/viewStory.jsp?id=29234>), but then critiques those states lacking a mandate for a minimum number of weeks candidates should student teach. For teacher educators like myself, who hold a divided opinion on Teach for America (see Téllez, 2011), this contradictory stance is frustrating. How can NCTQ at once praise TFA, which proudly does not require a genuine clinical experience, while giving low grades to states that do not force teacher education programs into longer student teaching placements? This is not the first time that teacher educators have pointed out the contradictions in the neo-liberal and neo-conservative agendas when attacking professional school teacher preparation (e.g., Cochran-Smith & Fries, 2001), but I don't think that the reach, as evidenced in NCTQ's goals and products, has ever been so obvious.

These contradictions have caused to me wonder about the goals of the anti-teacher education "league." But I also have questions for the unfailing promoters of

professional school teacher preparation.

How do NCTQ's standards compare to the shared standards that I outlined in the first section? First, I want to recognize that the NCTQ standards are written for programs, not candidates, but if a program is beholden to a set of standards, then the candidates will be necessarily held to the same goals. For instance, Standard 10 from NCTQ requires that programs "train teacher candidates to successfully manage classrooms." (See <http://www.nctq.org/standardsDisplay.do>). If a program wishes to achieve this goal, the candidates must also demonstrate these skills.

Given this caveat, I would argue that NCTQ's standards are very similar to the candidate standards we find across the states. In fact, I would even argue that the NCTQ standards are similar to NCATE's, which is somewhat surprising, given that NCTQ has been highly critical of NCATE (<http://www.nctq.org/p/tqb/viewStory.jsp?id=29218>). NCTQ specifically suggests that some of the strongest (by their estimation, anyway) teacher education programs are not NCATE approved. Both include standards that require programs to prepare candidates for diversity (NCATE Standard 4: "The unit designs, implements, and evaluates curriculum and provides experiences for candidates to acquire and demonstrate the knowledge, skills, and professional dispositions necessary to help all students learn") and equity (NCTQ Standard 13: "The program ensures that teacher candidates experience schools that are successful serving students who have been traditionally underserved"). Both also require that candidates know their content well, manage classrooms effectively, understand assessment, and so on.

For my part, the only real difference between NCATE and NCTQ is the method by which they find evidence, and in this regard the two could not be more unlike. NCATE works directly with the institution, requiring a campus visit that famously requires dozens of interviews with candidates, cooperating teachers, faculty, local school leaders, and other stakeholders. And pity the teacher educator who is responsible for compiling the "document room" in advance of team's visit. And we can't overlook the herculean efforts of already overworked, harried staff members who must orchestrate complex schedules, travel needs, special meal requests and countless other details in order to make the whole event possible.^x In stark contrast, NCTQ does not make visits, interviews no one, and draws its conclusions from documents alone.

Early in the chapter I referenced Taubman's (2009) book and now draw attention to the section in which he shares his account of an NCATE visit to his university. In

particular, he laments both the direct and opportunity costs the visit required, but, in the end, his reaction is more personal, describing NCATE accreditation as an “intensely depressing and intellectually numbing experience” (p. 91). I would use exactly the same words to describe NCTQ’s effort.

Ending thoughts

Creating standards for teachers and teacher education programs is easy. Put a dozen people--they don’t even need to be educators--in a room and they come up with the roughly the same list. Developing paper-and-pencil tests, performance assessments, accreditation manuals, syllabi rubrics, or interview protocols designed to evaluate teacher candidates or teacher education programs is more difficult, but tractable tasks nonetheless. Evaluating teaching standards might be the policymaker’s best chance at exacting reforms in our schools, but it’s a very blunt tool. Only when the evaluation of teaching standards is placed directly in the path of an individual who wants to be a teacher is any notice taken, and then only by the candidate herself. The efforts of NCTQ, NCATE, legislators, policymakers, and neo-conservative and neo-liberal pundits alike are largely superfluous. Genuine concern for the education of our nation’s children and youth is manifest in teaching them.

References:

- Barber, B. (1963). Some problems in the sociology of the professions. *Daedalus*, 92(4), 669-688.
- Bourdieu, P. (1990). *The logic of practice*. Stanford, CA: Stanford University Press.
- Brouwer, N., & Korthagen, F. (2005). Can teacher education make a difference? *American Educational Research Journal*, 42(1), 153-224.
- Bunch, G. C., Aguirre, J. M., & Tellez, 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), 26.
- Cochran-Smith, M., & Fries, M. K. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. *Educational researcher*, 30(8), 3-15.
- Darling-Hammond, L. (1999). *Reshaping Teaching Policy, Preparation, and Practice. Influences of the National Board for Professional Teaching Standards*: ERIC.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5-6), 523-545.
- Darling-Hammond, L., Newton, X., & Wei, R. C. (2010). Evaluating teacher education outcomes: A study of the Stanford Teacher Education Programme. *Journal of Education for Teaching*, 36(4), 369-388.

- Darling, L. F. (2001). Portfolio as practice: the narratives of emerging teachers. *Teaching and Teacher Education*, 17(1), 107-121.
- Duckor, B., Castellano, K., Téllez, K., and Wilson, M. Examining the internal structure of the Performance Assessment for California Teachers (PACT): A multi-dimensional Item Response Model validation study of the Elementary Language Arts Event. Paper presented at the American Educational Research Association conference, San Francisco, April 2013)
- Ellett, C. D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17(1), 101-128.
- Frazier, C. (1998). University responsibility for a public purpose. In R. Roth (Ed.), *The role of the university in the preparation of teachers*. New York: Routledge.
- Habermas, J. (1987). *The theory of communicative action, Vol. 1: Reason and the rationalization of society*. (T. McCarthy, Trans.). Boston: Beacon.
- Hafner, A. L., & Maxie, A. (2006). Looking at answers about reform: Findings from the SB 2042 implementation study. *Issues in Teacher Education*, 15(1), 85-102.
- Harris, D. N., Rutledge, S. A., Ingle, W. K., & Thompson, C. C. (2010). Mix and match: What principals really look for when hiring teachers. *Education Finance and Policy*, 5(2), 228-246.
- Heneman III, H.G., & Milanowski, A. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17 (2) (2003), pp. 173–195
- Heneman, H. G., & Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(2), 173-195.
- Imig, D., Wiseman, D., & Imig, S. (2011). Teacher education in the United States of America, 2011. *Journal of Education for Teaching*, 37(4), 399-408.
- Klecka, C. L. (2009). *Visions for teacher educators: perspectives on the Association of Teacher Educators' standards*. Rowman & Littlefield.
- Kristjánsson, K. (2010). Educating moral emotions or moral selves: A false dichotomy? *Educational Philosophy and Theory*, 42(4), 397-409.
- Le, V. N., Buddin, R. J., Education, R., Corporation, R., Education, U. S. D. o., & Sciences, I. o. E. (2005). *Examining the validity evidence for California teacher licensure exams*: RAND Santa Monica, CA.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement*, 44(1), 47-67.
- Mansvelder-Longayroux, D. D., Beijaard, D., & Verloop, N. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and teacher education*, 23(1), 47-62.
- Mitchell, D. E. (1984). Educational Policy Analysis: The State of the Art. *Educational Administration Quarterly*, 20, 129-160.
- O'Sullivan, S., & Jiang, Y. H. (2002). Determining the Efficacy of the California Reading Instruction Competence Assessment (RICA). *Teacher Education Quarterly*, 29(3), 61-72.
- Pease-Alvarez, L., Samway, K. D., & Cifka-Herrera, C. (2010). Working within the system: teachers of English learners negotiating a literacy instruction mandate. *Language Policy*, 9(4), 313-334.
- Pecheone, R. and K. Stansbury. 1996. Connecting teacher assessment and school reform. *Elementary School Journal*. 97(2): 163–77.

- Peck, C. A., Gallucci, C., & Sloan, T. (2010). Negotiating Implementation of High-Stakes Performance Assessment Policies in Teacher Education: From Compliance to Inquiry. *Journal of Teacher Education*, 61(5), 451-463.
- Peters, B. G. (1988). *Comparing public bureaucracies: Problems of theory and method*: University Alabama Press.
- Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(3), 269-273.
- Porter, J. M. (2010). *Performance Assessment for California Teachers (PACT): An Evaluation of Inter-Rater Reliability*. Doctoral Dissertation: University of California at Davis.
- Raths, J., & Lyman, F. (2003). Summative Evaluation of Student Teachers An Enduring Problem. *Journal of teacher education*, 54(3), 206-216.
- Sandholtz, J. H., & Shea, L. M. (2012). Predicting Performance: A Comparison of University Supervisors. *Journal of Teacher Education*, 63(1), 12.
- Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*: Basic Books.
- Stronge, J. H., & Hindman, J. L. (2003). Hiring the best teachers. *Educational Leadership*, 60(8), 48-52.
- Taubman, P. M. (2009). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*: Routledge New York, NY.
- Téllez, K. (2003). *Three Themes on Standards in Teacher Education: Legislative Expediency, the Role of External Review, and Test Bias in the Assessment of Pedagogical Knowledge*.
- Téllez, K. (2007). Have conceptual reforms (and one anti-reform) in preservice teacher education improved the education of multicultural, multilingual children and youth? *Teachers and Teaching*, 13(6), 543 - 564.
- Téllez, K. (2011). A case study of a career in education that began with "Teach for America". *Teaching Education*, 22(1), 15-38.
- Walsh, K., & Podgursky, M. (2001). Teacher certification reconsidered: Stumbling for quality. A rejoinder. *The Abell Foundation*.
- Watson, R., Stimpson, A., Topping, A., & Porock, D. (2002). Clinical competence assessment in nursing: a systematic review of the literature. *Journal of Advanced Nursing*, 39(5), 421-431.
- Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, 49(3), 587-602.
- Wolf, K., & Dietz, M. (1998). Teaching portfolios: Purposes and possibilities. *Teacher Education Quarterly*, 25(1), 9-22.

ⁱ I wish to thank my colleague Mark Dressman of the University of Illinois, Champagne-Urbana, who helped to conceptualize my understanding of standards represented in this document. Furthermore, some of the text is taken from a previously unpublished paper we wrote together. I also wish to thank editors for... Errors, however, are mine alone.

ⁱⁱ Although I used the term bureaucratic as a negative term here, I would remind readers that bureaucracies often develop with an interest of fairness in mind. See (Peters, 1988) for a highly readable treatment of the development of state bureaucracies.

ⁱⁱⁱ I place knowing in quotation marks because knowing a moral obligation is different than knowing how to create good assessment, for instance. I recommend (Kristjánsson, 2010) for an excellent review of this concept.

^{iv} See Ellett and Teddlie (2003) for an excellent historical review of teacher evaluation in the US).

^v I would glad to share the survey items with interested readers.

^{vi} See <http://www.edtpa.aacte.org/> for a description of this effort.

^{vii} In the interest of full disclosure, I was a signatory on a document objecting to the NCTQ methods and overall purpose. See http://www.nctq.org/docs/University_of_California_System_First_Letter.pdf. In that letter, we mostly argued against the forced participation and asked why NCTQ was so fixated on the inputs in teacher education and so uninterested in the actual performance of candidates?

^{viii} My own institution, University of California at Santa Cruz, is reported to be “fully cooperating” with NCTQ.

^{ix} The standards can be found at
(<http://www.nctq.org/standardsDisplay.do?output=P>)

^x I have worked in four different teacher education programs, only one of which was NCATE approved, but I did assist in preparation for an NCATE review.