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

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 its own comprehensive set of standards for teachers; many also have specific standards for initial licensure. We might imagine that variations in the states' educational contexts would result in wide variations.

In fact, we find a striking sameness of teaching standards, which I will demonstrate in the full paper. The larger question, however, is 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 begging of the question. 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, "minority" and "dominant" cultural ethnic groups). 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).

But it is true that 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.

In this paper, I pursue three related investigations. First, I conduct a 50 state policy analysis of teaching standards. This analysis reveals that the standards across states are more or less uniform, reflecting a traditional and somewhat conservative view of teaching. This result is predictable, but also disappointing in that no state standard was even slightly titled 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. These concerns are addressed in the full paper.

Second, I explored the development of teacher performance assessments developed from recent state standards. 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 measureable; 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. In particular, I focus on the development of the Performance Assessment for California Teachers (PACT), which is now the model for a nationwide performance assessment. My analysis demonstrates the challenges of turning standards into an actual assessment while also offering a brief and weak test of the validity of the PACT. I complied the PACT scores, ratings of university supervisors, and the ratings of cooperating teachers for the 85 students who completed our 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. As an early and very preliminary test of concurrent validity, these findings are surprising and perhaps worrisome.

I do not claim that my small 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 policymaker will turn to student achievement data linked to preparation program. Such a linkage is impossible to create, a topic I discuss at length in the paper.

In the final section of my paper, I explore the efforts of the National Center for Teacher Quality (NCTQ) 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. 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), 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 reading 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 do not think such a report will have much influence. First, prospective teachers rarely conduct a statewide search for the best teacher education program. Instead, they seek out the least expensive and closest route to the license, regardless of the quality. They seem to know in advance that school administrators tend not to be 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 seems to tell them that there is far more variability among candidates from a single program than any differences that might exist across programs.

I admit that this lack of attention to the quality of credentialing institutions may point out the very problem that NCTQ is trying to solve, but I am 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.

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, but then critiques those states lacking a mandate for a minimum number of weeks candidates must student teach.

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. We must reassess our role in preparing teachers within a policy environment that will only become more beholden to standards and their assessment.

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, 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 or actively preparing those who do.

References (from the full paper)

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.