

Three Themes on Standards in Teacher Education: Legislative Expediency, the Role of External Review, and Test Bias in the Assessment of Pedagogical Knowledge

By Kip Tellez

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)

Kip Tellez is an associate professor in the Education Department of Crown College at the University of California, Santa Cruz, Santa Cruz, California.

Robert Roth's chapter on standards in teacher education, which appeared in the 1996 *Handbook of Research on Teacher Education*, was prophetic. I am thankful that Bob used his considerable research and

Three Themes on Standards in Teacher Education

analytical skills to develop these ideas, and I encourage all interested readers to read or reread Bob's work on standards and accountability in teacher education. You will be impressed by his analysis and shocked by the accuracy of his predications.

As many readers of this journal know, Bob passed away several years ago, not long after completing the chapter cited here. A long battle with melanoma finally bested him. And while I did not know Bob well, I read his work regularly and now take inspiration from his writings. Given his deep and enduring interest in the work of standards, I invite all sympathetic readers to consider our efforts at "reclaiming" the profession's standards and assessments as a testimony to Bob's attempts to awaken us to what was coming; to what has indeed now come.

Those looking for the simple "standards-are-evil" rhetoric will be disappointed by Bob's work. In fact, he was involved in many of the standard setting projects of the 1980s and early 1990s. But he did not view standards in teacher education as a way to exact external control on the profession. He saw standards as a way to legitimize our work in the eyes of those who use terms like "Mickey Mouse" when referring to teacher education. Bob and those with whom he worked initiated movements to place standards in teacher education because they believed that pedagogical knowledge was critically important. He believed that a Conant-like¹ view of teacher education would weaken the process of learning to teach, and ultimately fail to serve the children in our public schools.

Like many of us, he saw both the good and bad in common goals or standards set by well-meaning professionals. He also recognized, as other papers in this special issue of *Teacher Education Quarterly* will point out, that standards, in and of themselves, are not the problem. Just as standardized tests, such as the Stanford-9, are not the problem *per se*. It is what we *do* with the *results* of such tests that matter. If the results of the Stanford-9 test were simply pasted in students' cumulative folders, who would care? It is only when we begin to rank and reward schools with "high" scores and punish those with lagging scores that the results are noticed. Similarly, until standards are measured and used in a way that affects our lives, they are given little notice.

My goal in this paper is to expand on three themes in the area of standards and accountability as they apply to the current condition of teacher education in the state of California and elsewhere. I realize that these themes are not necessarily related, but I believe that they are the most pressing issues as state legislatures and school boards grow increasingly involved in teacher education.

First, I hope to share my insights on why the teacher education standards movement has gained inertia so quickly. I address this topic because I want our profession to understand the seduction of standards among policy makers. Second, I would like to address the consequences of standards on existing accreditation processes such as the National Council for Accreditation of Teacher Education. Could state standards and assessment in teacher education undo our efforts at self-governance? Finally, and perhaps most importantly, I address the issues of bias in

paper-and-pencil tests of teaching knowledge. All evaluation is necessarily biased. What bias can we expect to find in the paper-and-pencil tests of pedagogical knowledge (e.g., the Teaching Performance Assessment now in development by the Educational Testing Service).

Standards: What's the Counterposition?

While I agree with writers such as Linn (2000) who suggest that accountability and tests have been a key part of educational reform in the past 100 years, I would go further than Linn by arguing that the reform in the present era *is* accountability. A growing segment of contemporary stakeholders in education are not interested in the content of learning goals, but rather in the idea that goals, whatever they are, are being measured and tracked. The lack of any specific destination for our nation's schools seems to set our era apart from other reform eras. But how did the era of accountability consume an interest in the direction of educational endeavors?

Like nearly every other reform of the twentieth century, the accountability reforms of today did not emerge from the ranks of local educators' wishes or outcries of student need. Rather, such reforms, in retrospect, have their origins in groups or organizations with enough power, money, or combination of the two to dictate the reform dimensions. For example, reforms in science and math education during the late 1950s and 1960s were fashioned not out of a demonstrated student need, but rather as a response to a perceived threat to the nation's preeminence in the sciences. Of course, the U.S. ultimately won the space race, and may have even without the math and science reforms. Science teachers, while they may have welcomed the additional money and interest shown toward their work, were not at the forefront of the "Sputnik" era reforms.

As another more insidious example, the back-to-basics reforms of the 1980s did not emerge from the ranks of teachers or parents, but were instead initiated by the well-crafted rhetoric of influential political leaders such as former U.S. Secretary of Education William Bennett. This reform was a particularly good illustration of how ideology alone, when shaped into policy, can affect the ways schools work.

As we now know, the back-to-basics reform was a stunning victory for Bennett and those who share his views on education. Not only were the "frills" eliminated from education, but the budgets that funded them could be slashed. Vocational programs, the arts, and so-called self-esteem curricula were no longer needed; now everyone was on the academic track to the university. Of course, college participation may have increased as a result of the back-to-basics movement (we have no way of knowing for certain), but the real benefits for both fiscal and social conservatives alike was the reduction in funding for public schools, especially urban schools. By hiding behind a policy that publicly claimed to be "detracking" the schools, the real goals of curriculum and fiscal reform could be met.

Three Themes on Standards in Teacher Education

Then emboldened conservative pundits in the 1990s moved beyond finding fault with specialized segments of school curriculum to narrating public education's total failure to promote student achievement. Through carefully designed and calibrated "objective" measures, they drew a line between test scores and a lack of rigor in the schools.

With many programs dismantled, a window was opened for a new era of reform, one that could hide behind the rhetoric of equal opportunity, just as the back-to-basics movement did. This reform, too, would require no new resources and could, its backers believed, allow for further financial cuts to schools. Thus, the accountability movement was born.

The accountability movement allowed policymakers to proclaim, "We are challenging the teachers, those who know best, to help our students meet the high standards we have set." Perhaps the most beneficial political consequence of such a high-minded statement is that it required no additional funding. Beyond administering a cheap (when compared to new books, buildings, or teachers) testing scheme, this stump speech is essentially foolproof. Who could possibly be against high standards? To the lament of the teachers and others who must work toward high standards with no additional resources, there is no counterposition to the high standards refrain.

Many front-line educators, especially those who worked with poor students of color, found that questioning whether the new high standards should be equally applied to all students led to a swift response from the accountability cabal, who retorted, "If you don't have high expectations for your low-income students, then perhaps you shouldn't be teaching."

The so-called success of the standards movement in K-12 has, I believe, led to the creation of standards in teacher education. The political expediency of the accountability movement has encouraged policy makers, many of whom are otherwise friendly to the issues teachers and teacher educators hold dear, to embrace standards wherever they are found. Legislators have found a hammer in the accountability movement and everything now looks like a nail. If standards and accountability have worked in the K-12 system, then they should be applied to all the endeavors funded by the state, including teacher education.

When political debate is reduced to sound bytes and television commercials, the standards movement has its greatest opportunity to take hold. Legislators become quickly enamored of the fist-pounding potential of a slogan such as, "I voted for (or, better still, authored) a bill that raised the bar for teacher quality." Modern political discourse, of course, does not encourage a protracted discussion of the height of the bar, who can and cannot leap it now, or even what the bar looks like.

In California, the K-12 curriculum standards were shoved through the legislative process by groups with plenty of money, power, or both. Educators may lament the route taken and the standards themselves, but they are now bound to them. With success in the K-12 realm ensured, the same groups who made sure that the

standards took hold had grown frustrated in their attempts to control and manage teacher education. SB 2042 offered an opportunity to further strengthen the teaching of the content standards. And legislators and other policy groups took out their hammer.

Whither NCATE?

Before arriving at the University of California, Santa Cruz, I taught at the University of Houston for nine years. During that time, the state of Texas developed and initiated a drastic accountability plan for teacher education. Briefly described, teacher education candidates are required to pass a paper-and-pencil test of pedagogical and content knowledge in order to receive the teaching license. Furthermore, each university in the state is accredited based on its candidates' passing rates. In a paper written with Mark Dressman, formerly at Houston, now at the University of Illinois-Urbana Champaign, we document our experiences as university teacher educators who were invited to review teacher education programs in jeopardy of losing their state accreditation (Dressman & Téllez, 2000).

It is important to point out that we volunteered to review the "accredited-warned" universities not out of desire to deliver our knowledge of teacher education to those who had less, but rather to investigate the workings of the teacher education accountability system. As it turned out, we learned as much from the "accredited-warned" teacher educators as they did from us. I could say much more about our experiences but lack the space to do so here. I encourage interested readers to contact me for a copy of the full paper.

However, I would like to build on a finding that we uncovered as a result of the research we conducted in Texas, one that could have implications for California and other states implementing state-sponsored accountability systems.

Shortly after the state teacher education agency in Texas announced the pass rates and accreditation rankings for each teacher education program in the state, one of the largest and most respected teacher education programs in Texas decided to forego its NCATE review in favor of promoting its pass rates on the state teacher education test. At first glance, this may not appear to be a dramatic shift, but a deeper analysis suggests a potentially much greater impact.

NCATE has had some success in encouraging states to mandate its review for all credentialing institutions in its state. For instance, Arkansas has mandated that each teacher education institution must undergo NCATE accreditation. North Carolina, on the other hand, does not mandate NCATE review, but nearly every teacher education program in the state is NCATE approved. More importantly, NCATE has been promoted by teacher educators and many legislators alike as a way for teacher education to control its destiny. NCATE's rigorous review visits are big, complicated, and expensive. But they offer an institution the chance to learn what its doing well and where it needs to improve. True, NCATE has standards, to

Three Themes on Standards in Teacher Education

which the review teams hold fervently, but an NCATE review does not render an easy pass or fail judgment. Teacher educators who have been reviewed or participated in the review process come to understand the complexity of teacher education.

It is the complexity and richness that may be NCATE's undoing. Potential teacher education students, who are likely to know little if anything about NCATE accreditation, will be very interested in whether the program they attend will prepare them for a test they must pass. Further, they are not likely to read a long and detailed NCATE report on the university and believe it will prepare them appropriately, no matter how glowing it may be. They will be interested in whether the program they attend will lead to a license.

If additional states follow Texas' lead by developing a teacher education accountability scheme based on the passing rates of state mandated tests of pedagogy, I am fearful that NCATE could become a relic — our best attempt at self-regulation. In the case of California, whose public universities license over 10 percent of all teachers nationwide, NCATE is holding on, but not by much. Only 13 of the 22 California State University campuses have current accreditation from NCATE; three private institutions are accredited; none of the University of California campuses have undergone NCATE review thus far. If the test results emerging from the implementation of the SB 2042 standards, the Teacher Performance Activities, are publicly shared and promoted by the state as an indicator of program quality, NCATE may lose what little foothold it has in California. And if legislators and other policy makers in California and Texas can turn the public's attention towards paper-and-pencil assessments of teacher "quality," it will not be long before other states are pursuing a similar strategy.

Bias in Teacher Licensure Tests

Psychometricians uniformly admit to a degree of systematic bias inherent in all tests and assessments. So the question for tests of pedagogical or content knowledge for preservice teachers is not *if* they are biased, but rather who is disadvantaged or privileged by the test and is the level of bias acceptable. In my view, teacher educators must address head on this issue, especially as new licensure tests are developed.

My research, teaching, and policy endeavors are shaped by a desire to create a more equitable system of schooling in the U.S. Towards this effort, I believe that increasing the number of teachers of color should be a central goal of all teacher educators. I maintain this goal even as I question the precise benefits of a more culturally "unified" teaching and student population (see Téllez, 1999). The goal of a culturally unified teacher and student population, however, is harder to achieve when tests of the kind evolving from SB 2042 (Teacher Performance Assessment) are used in high-stakes teacher licensure.

From the testing companies' perspective, all widely-used tests for preservice teachers, while biased, are not biased in ways that are socially unacceptable. For

their part, the standard regime in reducing bias via the use of item review committees, analysis of differential item functioning, and content analyses is a fair method for reducing bias to acceptable levels (Allan, Nassif, & Eliot, 1988). I believe that, for the most part, this method works well in reducing bias. However, the results of testing in many contexts reveal enduring differences along ethnic and class lines.

Tests of teacher certification reveal some of those same differences. For instance, in Texas, the result of the teacher education program accountability program (which is based on a paper-and-pencil test of pedagogical knowledge) shows that the universities with a high concentration of African-American students are much more likely to be placed on the “accredited—under review” list than those with a greater proportion of European-American students (<http://www.sbec.state.tx.us/edprep/edprep.htm>).

At the individual student level, the data suggests preservice teachers of color fail paper-and-pencil tests of teacher certification at greater rates than their European-American counterparts (Hood & Parker, 1989). We have known of these differential failure rates for a very long time, but the commonly-used tests of teacher licensure have been upheld by both the psychometric community and the courts. How have these tests, which are typically very high stakes, held up to bias review when it is clear that so many students of color are failing?

I believe the answer lies in a phenomenon I call *accretionary bias*. Test bias review and analysis at the individual item can demonstrate that each single item on a test is unbiased. However, examining individual item functioning by ethnicity occludes the possibility that each of the test items contains a proportionally small level of bias, which, over the full test, *accrues*, resulting in biased results at the test level, even as individual items are deemed to be acceptable. Accretionary bias, I argue, can happen in three ways.

First, paper-and-pencil tests of pedagogy, like those used in Texas and those proposed by the Educational Testing Services for use in California, tend to measure students’ analytical and reading comprehension skills rather than their knowledge of teaching knowledge and skills. In the Texas test, for example, examinees are required to answer several complex multiple choice items.² Research has demonstrated that such items tend to favor those examinees who possess strong general analytical abilities and are not effective in measuring achievement (Kolstad, Briggs, & Bryant, 1983). And even a cursory look at ETS’ proposed Teaching Performance Assessment Prototype suggests that examinees will be required to comprehend a great deal of text very quickly in constructing a response.

I am not suggesting that preservice teachers of color have few analytical abilities or are poor readers, only that when we “mix” a test of teacher certification with these abilities, we reduce the validity of the test. I am also not suggesting that these abilities are not important. Certainly, strong reading comprehension skills are important for teachers, and analytical talents are likely associated with quality

Three Themes on Standards in Teacher Education

teaching. But these skills, should we choose to value them, should be assessed in other formats. When conflated with a test of pedagogical knowledge, they tend to indicate lower scores for those of color.

Second, research demonstrates a relation between test anxiety and performance. The general findings from more than 1000 studies on test anxiety reveal that anxiety generally diminishes test performance. Germane to the discussion here, Phillips (1978) and Samuda (1975) have reported that test anxiety is differentially higher for African-American students, although such an effect appears to diminish as students move from elementary school to university level. Still, high levels of anxiety among preservice teachers of color are bound to affect the performance of students for whom a passing score is the only “ticket” to becoming a teacher.

Finally, the third form of accretionary bias is found in the content of the paper-and-pencil tests of pedagogy. In spite of the test developer’s general attention to diversity and questions focused directly on minority students, the overall pedagogy featured is one which is more commonly found in suburban schools among European-American students. For instance, the scenarios in the tests feature teaching in which teachers work *with* students, students are encouraged to ask questions, and the development of open-ended questioning practices by teachers — and correspondingly “open” responses by students. Assessing a teaching candidates’ knowledge of a constructivist, collaborative, and critical instructional orientation instruction is, of course, part of the test objectives. However, many teacher education students of color attended schools in which the instruction was neither constructivist nor collaborative. Indeed, much of what preservice teachers of color have experienced as “good teaching” in their own educational careers may run counter to the teaching practices considered good by the test. It is not surprising, then, when preservice teachers of color fail to see the fine divisions among alternative responses: the instruction they are reading about is widely different than the education they received, and the contradiction can be confusing.

These three features of bias, each of which accrue over the span of the test, are not considered when standard techniques of bias identification are used.

The assessment of teaching skills remains an ideal (Grover, 1991), and paper-and-pencil tests of pedagogy are a blunt tool for assessing an extraordinarily complex human endeavor. But we should not allow our measurement imprecision to differentially and damagingly affect the very students who often have the most difficult journey to teacher licensure.

Conclusion

Teacher educators in California, Texas, and other states must renew the battle for control over the standards that drive our work and the assessment of our students’ knowledge. As I have pointed out, the political and legislative expediency of standards will challenge us, our efforts at self-regulation will become increas-

ingly difficult, and we may find our work to create a more culturally diverse teaching population undone by mandated tests. Nevertheless, we are bound to the cause of self-regulation. Our students almost universally come to regard their preparation in pedagogy as crucial to their development as teachers. If we are forced to abide by external standards which cheapen the complexity of teaching children and youth, our programs will amount to little more than year-long apologies for the curriculum and testing schemes mandated upon us.

Acknowledgement

I wish to thank Mark Dressman, University of Illinois, Champaign-Urbana, for conceptual and editorial assistance.

Notes

¹ Conant's words, taken from the 1963 report bearing his name:

I felt confident that I was an excellent teacher and I had developed my skill by experience, without the benefit of professors of education. I saw no reason why others could not do likewise, including those who graduated with honors in chemistry and who wished to teach in high school. (Conant, 1983, p. 1)

² Complex multiple choice items require the examinee to consider various combinations of responses while considering the same stem. Below is an example, taken from the Texas Professional Development test.

Following the discussion of a possible composting project, Mr. Rivera arranges for his students to take a guided tour of a greenhouse that uses composting. Which of the following are most likely to be key benefits of Mr. Rivera's use of this local resource?

- I. helping students understand and visualize procedures used in composting
- II. promoting students' sense of control with regard to their learning
- III. enhancing student motivation to pursue the project they have been discussing
- IV. encouraging students' conscious use of self-assessment during the learning process

- a) I and II only
- b) I and III only
- c) II and IV only
- d) III and IV only

References

- Allen, R.G. Nassif, P.M., Eliot, S.M. (1988). *Bias issues in teacher certification*. Hillsdale, NJ: Erlbaum.
- Conant, J.B. (1963). *The education of American teachers*. New York: McGraw-Hill.
- Dressman, M., & Tellez, K. (2000). *Accredited under review: An inside-outside account of high stakes education reform*. Paper presented at the annual meeting of the American

Three Themes on Standards in Teacher Education

- Education Research Association, New Orleans.
- Grover, B.W. (1991). The teacher assessment dilemma: What is versus what ought to be! *Journal of Personnel Evaluation in Education*, 5(2), 103-119.
- Hood, S., & Parker, L. (1989). Minority bias review panels and teacher testing for initial certification: A comparison of two states' efforts. *The Journal of Negro Education*, 58, 511-519.
- Kolstad, R.S., Briggs, L.D., & Bryant, B.B. (1983). Complex multiple-choice items fail to measure achievement. *Journal of Research and Development in Education*, 17, 7-11.
- Phillips, B.N. (1978). *School stress and anxiety: Theory, research, and intervention*. New York: Human Sciences Press.
- Roth, R.A. (1996). Standards for certification, licensure, and accreditation. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education*, 2nd Edition (pp. 242-306). New York: Macmillan.
- Samuda, R.J. (1975). *Psychological testing of American minorities*. New York: Harper & Row.
- State Board of Educator Certification accreditation rankings (n.d.) Retrieved October 31, 2002, from <http://www.sbec.state.tx.us/edprep/edprep.htm>.
- Télliez, K. (1999). Mexican-American preservice teachers and the intransigency of the elementary school curriculum. *Teaching and Teacher Education*, 15, 555-570.