

Education Working Paper Archive

**Teacher Licensure Tests:
Their Relationship to Mathematics Teachers' Academic
Competence and Student Achievement in Mathematics**

September 5, 2007

Sandra Stotsky
University of Arkansas

Introduction

Many educators choose to believe that learning to read and write is as natural as learning to listen and speak, even though scientifically based research does not support their belief. However, most educators (as well as the public) believe that most students must be taught mathematics to learn it. Moreover, there is a body of research evidence that attests to the positive relationship between students' mathematics achievement and their teachers' mathematics knowledge (e.g., Monk, 1994). Teachers who know more mathematics than their peers have students who learn more mathematics than their peers (e.g., Boots, 2007). Thus, state and federal officials, as well as the general public, are rightly concerned about the academic qualifications of those who teach mathematics (and science) in the public schools, especially since there has been a steady decline for decades in the number of mathematics and science majors or minors choosing secondary school teaching careers. There has also been a steady decline in the number of high-achieving women seeking to become elementary teachers or teachers of other subjects (Hoxby & Leigh, 2005).

About two decades ago, in an effort to ensure that their teachers had an adequate grasp of the field of their license before they began teaching, states began to require the passing of a subject matter licensure test for entry into the profession. Licensure tests—typically tests assessing the basic substantive knowledge needed for professional practice—are the major objective measure of quality control used by most professions for entry into the profession. It is well known, for example, that as a result of the recommendations in the Flexner Report in 1910, major substantive changes were made to the content of medical education in the United States (and Canada). It is less well known that quality controls took two forms: selective admissions (rare before the 1920s) and tougher state licensing exams (Hampel, 2005). By 1935, only 66 of the 160 medical schools in this country in 1904 had survived.

By default, licensure tests have determined what new teachers in elementary, middle, and high school need to know in mathematics in order to teach the subject. They have also influenced how new teachers taught mathematics if they or other required tests contained pedagogical items. However, we lack a critical summary of the research on the content, value, and uses of teacher licensure tests. A small but growing number of studies have examined the content or value of teacher licensure tests and their relationship to student achievement. What do they tell us about the overall content of licensure tests for prospective teachers, whether the tests assess their basic skills or their knowledge of the discipline? Do researchers take content coverage, item difficulty, and cut scores into account in assessing the value of teacher licensure tests? What is their predictive value for student achievement? Is this a meaningful statistic? The purpose of this paper is to indicate what we can learn from these studies, especially those that examine the content or use of teacher tests assessing mathematics knowledge, and to highlight a number of questions that warrant research if these tests are to serve the same function that licensure tests serve other professions.

Definition of Key Terms

- A **certificate** is professional recognition by a **non-governmental agency** or association bestowed on an individual who has met predetermined qualifications specified by the agency or association. E.g., the National Board for Professional Teaching Standards awards a certificate, not a license, to those who complete its certification program. Teachers' licenses have often been called "certificates," but they are in fact licenses.
- A **license** is a permit granted by a **governmental agency** to an individual who has met specified requirements. A teacher training institution cannot grant a license

(often called a certificate because the process for obtaining a license has traditionally been called certification). At the most, the institution may recommend a candidate for licensure for fulfilling those specified training requirements it has provided. Teachers must obtain a license to teach in a public school in every state, although they need not complete an “approved program,” traditional or alternative, in order to obtain a license. In a growing number of states, it is possible to obtain an initial or temporary license by passing just the state’s required licensure tests and a criminal background check. A license provides some measure of quality control and consumer protection at the individual level.

- **Program approval** is a process in which a peer review team determines whether a teacher training program meets state requirements and can be legally empowered to make recommendations for licensure. The program must be approved, ultimately, by a state agency. A state agency can accept a recommendation to approve a program from a non-governmental agency that has undertaken the review, such as the National Council for Accreditation of Teacher Education (NCATE) or Teacher Education Accreditation Council (TEAC). Program approval, or accreditation as it is often called, provides consumer protection and quality control at the institutional level in pre-service preparation.
- **Licensure tests** are tests that have been developed and/or approved by a state agency to assess prospective teachers’ basic qualifications for licensure in that state. As with licensure tests for other professions, they are not under the control of the preparation programs themselves. They are not intended to serve as achievement, intelligence, or diagnostic tests although they sample from many (presumably relevant) domains, and compensatory scoring is used to arrive at a raw score, which is then translated into a scaled score. As with most other licensure tests, the test-taker either passes or fails; teachers who fail may retake a test as many times as they wish (and many do). Most teacher licensure tests are relatively short and inexpensive, compared with licensure tests for entry into other professions.

Background Information

Why the focus on teacher licensure tests now. There are two major reasons for teacher licensure tests: to protect the public and to make teacher training programs accountable for the initial academic competence of those who complete their programs. While licensure tests in some professions have not always been motivated by an interest in making their professional training programs accountable for graduates’ performance on these tests, this was clearly the reason for the development of these tests for prospective teachers.

A provision in Title II in the 1998 reauthorization of the Higher Education Act compelled all states to require licensure tests for new teachers. Each state henceforth had to report annually on the pass rates on tests of its own choosing for each cohort of prospective teachers completing training programs in the state’s own teacher training institutions. Today, most states require teaching candidates to take at least two tests; one assesses the candidate’s reading, writing, and arithmetic skills, the other assesses the content knowledge needed for teaching the field of the license at the grade levels it covers.

When teacher licensure tests are taken. These two licensure tests are taken at different junctures in teacher preparation, typically **not** at the completion of the program, unlike most professional licensure tests. Because most states do not mandate when their teacher tests are to be taken, a

growing number of teacher training institutions use the state-required skills test to screen admission **into** their licensure programs. States using Educational Testing Service tests require PRAXIS I for this purpose. States contracting with National Evaluation Systems use a skills test developed by NES for this purpose. The test of prospective teachers' subject matter knowledge, a test that often includes pedagogical items, is increasingly being used to screen admission **into** student teaching in undergraduate licensure programs. Both tests are usually required for admission **into** post-baccalaureate programs for the initial license.

Testing companies. At present, two large, private testing companies develop the teacher tests used by the states. Educational Testing Service provides licensure tests for about 35 states, chiefly states with small populations, and National Evaluation Systems contracts to provide tailor-made tests for over 12 states, chiefly the most populous states. Well over 50% of U.S. teachers are licensed in NES states (Mitchell & Barth, 1999). American Board for Certification of Teacher Excellence is a new player on the scene, now developing subject tests embedded in a short preparation program that lead to initial licensure for prospective teachers who do not wish to enroll in a traditional preparation program. ABCTE's subject tests have a pre-determined high cut score for the initial license and, in addition, a pre-determined higher pass score that leads to certification for master teacher status for experienced teachers who seek an alternative to the certification program sponsored by the National Board for Professional Teaching Standards. So far, ABCTE's Passport to Teaching program has been accepted as alternative route to an initial license in seven states, and about 4000 teachers are now in the pipeline, according to its website.

Types of educator licenses. Across states, a bewildering variety of licensure tests are used for entry into the teaching profession. They cover differing spans of grade levels and contain different numbers and types of test items in the field of the license. These differences reflect what educators and educational organizations in a state have decided meet children's or schools' "needs" and/or the forces of supply and demand. A state may choose to require one or more license-specific tests for those who teach and/or supervise arithmetic or mathematics as:

1. Pre-school, kindergarten, and/or primary grade teachers (e.g., PreK-2, PreK-3, K-3, K-2)
2. Elementary teachers (e.g., K-6, 1-6, K-5, 1-5, or K-8)
3. Elementary mathematics teachers (e.g., 1-6)
4. Special education teachers (for one educational level, for all educational levels, and/or for moderately or severely disabled students)
5. English as a Second Language or bilingual education teachers
6. Middle school mathematics teachers (e.g., 5-8, 6-8, or 5-9)
7. High school mathematics teachers (e.g., 9-12, 8-12, 6-12, or 5-12)
8. Mathematics coaches
9. Mathematics curriculum supervisors/directors (e.g., K-8 or 7-12 or K-12)

Relationship to state K-12 standards. The tests that a state requires for prospective mathematics teachers may or may not be closely related to its own K-12 mathematics standards. Much depends on whether the test is an off-the-shelf ETS test or one tailored to the state's K-12 mathematics standards by NES. Even in NES states, the committee developing or reviewing test items may not necessarily hew closely to the state's K-12 mathematics standards.

Cut or pass scores.. Each ETS state determines its own cut or pass score, which may differ from that of another state using the same ETS test. NES states also determine their own cut or pass score. ABCTE has a pre-determined pass score, no matter in what state test-takers live. Pass scores are determined by a differing group of people in each state. There are no data across states on how many test items need to be correct for a passing score on each of the different tests that

states require. Test formats differ across tests and testing companies. In some but not all NES states, tests may have about 80 multiple-choice items and two short essay questions. ETS tests tend to have mainly multiple-choice items, but ETS does offer tests with essay questions as well. For licensing prospective elementary teachers, ETS provides an array of tests and test formats. Compensatory scoring is used for most ETS and NES tests. As a result, it is possible in theory for a test-taker to get most mathematics test items wrong on the multi-subject test taken by most prospective elementary teachers but still pass the test.

Difficulty level of tests for the teacher of mathematics. There are no data on the difficulty level of mathematics tests across testing companies and licensure tests. Like most licensure tests, teacher licensure tests must be secure, and items are not released on a frequent basis, if at all. (Practice tests may be provided, however.) So, researchers need to make special arrangements to view a test and cannot report on details in currently used test items, just on sample items selected by ETS for its tests or by a department of education or other state agency for NES tests.

Information on psychometric qualities of teacher licensure tests. Information on the psychometric qualities of the teacher tests provided by ETS is provided directly by ETS to interested researchers. However, that is not the case with comparable information on tests developed by NES. Information on the psychometric qualities of these tests is the property of the states with which NES has a contract, and the agency in charge of the contract in each state must give permission to make that information available to a researcher. Information can be obtained, but the relevant agency must be asked, first, and it may ignore or turn down the request for reasons of its own.

Research on Teacher Licensure Tests

Studies of test content. Two studies have examined the content of a number of tests, not just sample questions. Ruth Mitchell and Patte Barth (1999) examined licensure tests chiefly in English and mathematics. They concluded that current tests are for the most part academically weak. Two-thirds of the mathematics items on PRAXIS I, a test of teachers' basic skills, were judged to be at the middle school level, with fewer items on algebra and geometry than on the 1996 grade 8 NAEP mathematics test. Overall, they judged the tests of teachers' skills at the "8th to 10th (sometimes 7th) grade level."

They judged the overall content of the subject tests they examined about the same as in "high-level high school courses," with a "few under-used exceptions." They found most mathematics licensure tests dominated by "simple recall" in multiple-choice items. They judged secondary mathematics tests to be at the 10th to 11th grade level. They judged those required for elementary licensure as a whole "at about the tenth grade level." According to their analysis, licensing tests fail to ask for a deep knowledge of the key concepts connected to the field of the license.

Mitchell and Barth view the system as "designed to prevent false negative judgments (about either candidates or the institutions that produce them)." "Underlying the whole process," they believe, "is the assumption that teachers only need to know the content that is expected of their students, and maybe just a little bit more." This assumption is made more explicit by the low pass scores states tend to set, such that passing a licensing exam "can mean nothing more than a high school diploma." They urge a loosening of the "stranglehold that litigation and psychometrics have on developing licensing examinations" so that they can become "instruments that signify high professional standards."

In a 2006 report for the National Council for Accreditation of Teacher Education, Diana Rigden examined the contents of five tests provided by ETS for licensing elementary teachers, as well as the information NES provides on the reading tests it has developed for three states. She wanted to see if these eight tests address the knowledge base for effective reading instruction. Rigden found that only one of the ETS tests (PRAXIS 0201), a dedicated reading test required only in Tennessee (and for which test-takers get credit simply by taking it), and the three NES tests have items that address the five components of scientifically based reading instruction. She observed that PRAXIS 0011, a multiple-choice test commonly used for elementary licensure in ETS states, “is not a good measure of a teacher candidate’s knowledge of the five components of effective reading instruction.” It should be noted that 35% of its test items address reading instruction. Rigden’s study does not provide information on how ETS’s commonly-used multi-subject tests for elementary licensure (such as PRAXIS 0011) address the mathematics knowledge needed for teaching mathematics in grades 1-6.

Studies of relationships between teacher tests and student achievement. Five studies have examined the relationship between teachers’ scores on licensure tests and K-12 student achievement, but in different ways and with different measurement instruments. They address mathematics only with respect to whatever mathematics is on the licensure tests that had been taken by the teachers in their studies. None provides a systematic analysis of exactly what these licensure tests actually assess.

In a 2007 study, Joshua Boot compared value-added results for 78 Tennessee middle-school teachers (65 held secondary mathematics licenses, the others held apprentice or interim licenses) who took both ABCTE’s secondary mathematics test and its test of professional teaching knowledge. Students of the teachers with scores on the ABCTE secondary mathematics tests that were one standard deviation above the study mean showed greater gain in mathematics achievement than students of the teachers with scores that were one standard deviation below the study mean.

In a 2006 study, Joshua Boot compared value-added results for 55 Tennessee teachers in self-contained elementary classrooms who took both ABCTE’s multiple subject test for elementary teachers and its test of professional teaching knowledge. Students of the 13 teachers who passed both ABCTE tests had significantly greater overall improvement in achievement than students of the other teachers (who failed to pass one or both of the ABCTE tests), exceeding one year’s progress in all subjects. In other words, teachers who met ABCTE’s requirements for elementary education certification produced greater academic achievement in their students, especially in mathematics, than teachers who did not. It should be noted that Tennessee’s elementary teachers must pass several PRAXIS tests to obtain licensure.

In a 2006 study to be published, Dan Goldhaber analyzed the relationship between the quintile status of almost 24,000 North Carolina teachers on two PRAXIS tests they had taken for elementary licensure (PRAXIS 0011 and PRAXIS 0012) over a 10-year period (1994-2004) and the scores of their 701,000 students in grades 4 to 6 on state tests in mathematics and reading. His results “generally support the hypothesis that licensure tests are predictive of teacher effectiveness, especially in teaching mathematics.” However, Goldhaber expressed concern about “false positives” and “false negatives” in relation to the cut-score that a state uses and considers current teacher tests a “weak signal” of teacher quality.

In a 2006 study, Charles Clotfelter, Helen Ladd, and Jacob Vigdor examined the relationship between the assessment of teacher effectiveness and how students are matched to teachers, using 3842 grade 5 teachers in 1160 elementary schools in North Carolina in 2000-2001. Using scores

on the two PRAXIS tests these teachers had taken for elementary licensure (PRAXIS 0011 and PRAXIS 0012) and on the state's student tests in mathematics and reading, the study found that the positive correlations between teacher qualifications (as measured by experience and licensure test scores) and student achievement were explained largely by the match between students and teachers across schools. For the typical student, the researchers found that the benefit from having a highly experienced teacher was approximately one-tenth of a standard deviation on reading and math tests scores; with respect to licensure test scores, a one-standard-deviation increase in scores increased predicted student achievement in math by 1 to 2 percent of a standard deviation. Experience was clearly more important than licensure test score but only for socio-economically higher and more able students in mathematics.

In a 1998 report for the Pennsylvania State Board of Education, Robert Strauss examined teacher preparation and selection in Pennsylvania. According to his report, Pennsylvania awards 20,000 new elementary teaching licenses each year while less than 2,000 new elementary teachers are hired in Pennsylvania annually; i.e., its training institutions prepare far more elementary teachers than the state needs. The passing scores on the ETS tests it uses, set by panels of Pennsylvania teachers, are very low: about 90% of the test-takers pass the tests after answering from 25% to 60% of the questions correctly. (In comparison, 48% in Pennsylvania annually pass its law boards, and only 18% its CPA exams.) Graduates from some teacher training institutions answer only from 20% to 40% of the questions, while graduates from other institutions correctly answer from 50% to 75%. Strauss finds no statistically significant relationship between hiring decisions and teacher test scores; i.e., local schools do not necessarily hire the most academically qualified teachers. However, "where districts utilize more professional personnel procedures in their recruitment of teachers, student achievement is generally higher. Where more emphasis is given to matters of residency and non-academic matters, student achievement is lower." To improve preparation and selection of teachers in Pennsylvania, he recommends (among other things) higher passing scores on teacher licensure tests, more stringent program approval standards that specify content majors (especially for secondary school teachers), and state-specified admissions standards for teacher preparation institutions.

Related studies. Several other studies have examined other relationships, other aspects of licensure tests, or the information and examples that testing companies provide for their tests.

In a chapter from *Handbook on the Assessment of Teachers* (Hill, Sleep, Davis, & Ball, 2006), the authors provide a wealth of information on the history of teacher assessment, efforts to measure teachers' knowledge in mathematics, methods for measuring professional mathematical knowledge, and contemporary approaches to testing teachers for licensure. The authors offer many useful observations about teacher tests and teacher testing after examining sample items for mathematics tests developed by ETS, NES, and ABCTE. They recommend that subject matter licensure tests should assess "instructionally relevant" mathematics—the mathematical knowledge that teachers use—and that these tests should have predictive validity for teachers' classroom performance.

In a 2006 study of the relationship between about 10,000 "certified," "uncertified," and "alternatively certified" teachers of reading and mathematics in grades 4 to 8 and their students' scores on state tests in mathematics and reading in New York City's schools over a six-year period, Thomas Kane, Jonah Rockoff, and Douglas Staiger found that teachers from traditional training programs were generally no more or less effective than teachers from alternative (or no) programs (including a large number from Teach For America). More variation in effectiveness could be found within each status group than among them. Finding little predictive value for teacher effectiveness from licensure status, the report recommends using value-added data on

student achievement in the first two years of teaching to judge teacher effectiveness. It seems to imply that licensure could be abandoned for initial hiring but does not suggest what criteria might be used for that purpose. Indeed, it warns against relying on academic background for initial hiring. The study cannot address the usefulness of licensure tests in screening teaching candidates for their academic competence since most of the teachers in the study had to take and pass New York State's tests, regardless of entry route. (New York is a NES state.)

A National Research Council volume titled *Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality* (Mitchell, Robinson, Plake, & Knowles, 2001) is expressly concerned with the general content validity of teacher tests and their technical, or psychometric, qualities. Although this volume seems to suggest that there is little or no predictive ability from current teacher licensure tests for student achievement, its nine chapters provide no information on (1) the quality of the content of any licensure test, (2) the difficulty level of the items in any test, and (3) the distribution of the content covered on any test.

In a Working Paper issued in 2003 by the National Bureau of Economic Research (later published in 2004), Joshua Angrist and Jonathan Guryan address the question of whether teacher testing raises teacher quality. Using Schools and Staffing Survey data to estimate the effect of state teacher testing requirements on teacher wages and teacher quality as measured by educational background, they concluded that teacher testing increases teacher wages with no corresponding increase in quality. The study did not address the quality and difficulty level of teacher tests or the relationship of teacher tests to student achievement.

Based on an analysis of the information that testing companies provide about their licensure tests for prospective elementary teachers, Sandra Stotsky (2006a) concluded that most of these tests do not assess adequately (if at all) research-based knowledge of reading instruction. For example, PRAXIS 0011 (used in Goldhaber's 2006 study) seems to devote no more than 7% of its items to three crucial components of beginning reading instruction (development of phonemic awareness, phonics, and vocabulary); PRAXIS 0012 (also used in Goldhaber's 2006 study) seems to devote no more than 1% to these three components. Rigden's examination of the content of several ETS and NES licensure tests for elementary teachers in a 2006 report for NCATE supports Stotsky's conclusions.

Stotsky also examined the information on a relatively new set of PRAXIS tests called Principles of Learning and Teaching. This set of tests (one for each of four different educational levels) is designed to assess "what a beginning teacher should know about teaching and learning." According to the ETS website, 18 states require these tests in addition to a subject test. To judge by the sample constructed-response and multiple-choice questions and sample responses and question answers for these tests, this set of tests strongly promotes constructivist pedagogies and discredits alternative pedagogies.

What We Can Learn from Studies on the Predictive Value of Teacher Licensure Tests

So far, extremely little, and for several reasons. First, we do not know what most licensure tests of subject matter knowledge assess and, therefore, what passing them means. The licensure tests taken by teachers in North Carolina and Pennsylvania were predictive of teacher effectiveness (Goldhaber, 2006; Strauss, 1998), and in Tennessee an academically stronger test predicted more effective teachers than did a weaker test (Boot, 2006). However, none of these studies examined what these tests measure in reading or mathematics. We do know from Stotsky's and Rigden's research that one of the tests used in these three states—PRAXIS 0011—does not address the elements of sound reading instruction adequately. So, it is not at all clear what we learn about the quality of a new elementary teacher of reading from a passing score on this test.

Second, when researchers do learn what tests of subject matter knowledge assess, most are judged so academically weak that passing them has little academic significance. According to Mitchell and Barth's 1999 study, the overall difficulty of the mathematics content on tests for prospective high school mathematics teachers is no higher than upper high school level, and is at a lower level on tests for prospective elementary teachers, before a state sets a cut score. Items on the ETS test for assessing prospective teachers' own reading, writing, and mathematics skills (PRAXIS I) were judged to be chiefly at the middle school level, before a state determines its cut-off score. The chief problem that Mitchell and Barth see is not false negatives—teachers who fail their licensure tests but might have become effective teachers—but false positives—teachers who pass their licensure tests but do not become effective teachers. At best, these tests may screen out only grossly incompetent candidates, especially for the elementary license.

Third, we do not know what student achievement on the state mathematics tests in Tennessee, North Carolina, and Pennsylvania means since none of the studies provided information on the content of the state's student tests. These tests may not indicate a very high level of mathematics achievement if NAEP state test results are used as the common yardstick. While Pennsylvania's rating was about mid-way in rank order, North Carolina and Tennessee received two of the four lowest ratings in a study comparing the difference between the percentage of a state's students who scored at the proficiency level on the 2003 NAEP tests and the percentage of students who scored at the proficiency level on the state's own tests (Peterson & Hess, 2005). In other words, there is a huge difference between the percentage of students judged by Tennessee and North Carolina tests as proficient in reading and math and the percentage of students judged by NAEP tests as proficient in reading and math. These states had much higher percentages of "proficient" students on their own tests than on the NAEP tests (e.g., Tennessee's reading test judged 87% of its eighth graders as "proficient," while NAEP's test found only 27% "proficient"). The problem could be the cut score a state uses for its student tests and/or a test design that does not permit measurement of high achievement levels.

One would expect in theory (all other things being equal) effective mathematics teachers to come from the pool of teachers who enter the profession with an adequate grasp of mathematics for the range of grade levels and students they are licensed to teach. But if states use mathematics tests that cannot discriminate between prospective teachers with an adequate grasp of the subject and those without one, and if state tests of student achievement also cannot discriminate meaningfully among students with higher levels of mathematics achievement, it is not clear that we can gain much useful information about the effectiveness of mathematics teachers from an examination of the relationship between such teacher tests and such student tests.

There is yet another reason that we may not be able to learn much from existing research on this issue: licensure tests are not intended or constructed to predict teacher effectiveness. Subject matter tests are constructed to discriminate between those test-takers who are and are not "just acceptably qualified individuals" (i.e., just at the level of subject matter knowledge required for entry-level teaching in the field). Raw scores on a licensure test might have a loose relationship with student achievement, but those who don't pass a pass/fail test don't get a license to teach. Several independent experts on the construction of teacher licensure tests (i.e., they are not employed by companies developing teacher tests), with experience in reviewing the specifications and procedures for their development and validation, see their purpose only as distinguishing between candidates with and without minimum levels of skills and knowledge necessary for an entry-level teaching position (Mehrens, Klein, & Gabrys, 2002). According to this perspective, one should not expect teacher performance on licensure tests to be related to student outcomes when these tests are not intended to be a predictor of teacher effectiveness,

While students of more mathematically knowledgeable teachers tend to have higher mathematics achievement or gain more in mathematics than students of less mathematically knowledgeable teachers, as judged by SAT or ACT scores or various indices of mathematics coursework or knowledge, current licensure tests cannot necessarily identify a mathematically competent test-taker with respect to the grades covered by the license. For example, if the cut score on a multi-subject licensure test for elementary teachers is so low that test-takers can pass it despite failing many if not most of the mathematics test items, the test *ipso facto* cannot identify mathematically competent teachers. And, if objective value-added results for experienced teachers in a state cannot be calculated (e.g., because student scores cannot be traced to specific teachers) and teacher effectiveness is determined by evaluations of classroom performance by local supervisors, high teacher ratings based on very subjective measures may well keep in place academically impoverished licensure tests featuring non-research-based pedagogies.

A case in point is the use of PRAXIS 0011 and PRAXIS 0012 for predictive purposes. Both are described as measuring knowledge and skills for teaching reading, mathematics, and other subjects in the elementary grades. In Goldhaber's study, higher quintile status on these two tests predicted higher student achievement in reading and math in grades 4 and 6 on state tests. Yet, oddly, Rigden's analysis of the content of PRAXIS 0011, supported by Stotsky's analysis of ETS-provided information on both tests, found few or no questions on research-based reading instruction in its reading section (35% of the test—mathematics test items are 20% of the test). What does one make of this anomaly? Is the research on reading instruction seriously flawed? Are the student tests seriously flawed as measures of reading achievement? Do the tests assess a candidate's ability to parrot a pedagogical party line in reading, so to speak? Perhaps other explanations are possible for this puzzling finding (a relationship between student achievement and higher scores on licensure tests that assess non-research-based reading instructional knowledge). Clotfelter, Ladd, and Vigdor's study suggests that the relationship between higher quintile status on these two licensure tests and higher student achievement in *mathematics* in North Carolina might reflect a third factor: the higher socio-economic status of the students in the classrooms of teachers with higher test scores. Surely, a state should hesitate to alter its cut scores on these tests as a way to get stronger teachers.

Even if a test assesses a teaching candidate's mathematics knowledge base more adequately (i.e., the mathematics items account for much more than 20% of the test), it is not clear how to set the bar higher and how high. It is not easy to raise a bar. When standard-setters rate each test item in a standard-setting session, they ask: Is this a test item that an entry-level teacher will know? If standard-setters come either from higher education institutions that admit weak candidates into their licensure programs or from school systems that place less value on academic achievement than on other qualities in a prospective teacher, they may say no for a relatively difficult item even if it is related to basic knowledge in its field. There are fewer consequences for standard-setters' home institutions if a bar is set lower than higher.

It may well be necessary to alter the paradigm used for standard-setting on teacher licensure tests to ensure that new teachers come into their *first* teaching positions with enough discipline-based knowledge to enable them to teach the most advanced students they may encounter in the highest grades covered by their license. Perhaps standards-setters for a mathematics licensure test should be primarily teachers determined to be effective for the whole range of students typically found at the highest grade levels covered by the license (with effectiveness determined by value-added results from mathematics tests judged to be academically strong). And perhaps the question they should ask as they rate items on a licensure test for their grade levels is: Does this item reflect knowledge that an entry-level teacher would have gained from academic coursework preparing

the test-taker to be an effective teacher for the whole range of students that may be in the grades covered by the license?

An Agenda for Further Research

It is tempting to infer that we could bring prospective teachers' knowledge of mathematics to an acceptable level before they enter the profession simply by increasing the quality and difficulty of their subject matter licensure tests. But, many details need to be explored and decided upon first because of the vast differences in (1) the mathematics needed for teaching mathematics at different educational levels, (2) the range of mathematical competence among the students who might be in a typical elementary, middle, or high school classroom, and (3) the demands of the mathematics textbooks and other curriculum materials that teachers may be required to use, especially in the elementary and middle school. Some of the details can be informed by research; others require professional judgment.

What level of mathematics skills should be expected of prospective teachers? We have some information on the level of arithmetic skills demanded on the skills tests used across NES states and on current versions of PRAXIS 1. But we need more. Whether or not a skills test is used for admission to a teacher training program, passing such a test is required for licensure in all states. What arithmetic skills should be expected on these tests for those who want to teach preschool or kindergarten, given that most states do not require them to take a licensure test assessing subject matter knowledge? If the required test does assess knowledge of arithmetic, it would not assess much of this knowledge because the test would also assess subject matter knowledge in other areas commonly taught by early childhood teachers in self-contained classrooms. And should test-takers be allowed to use a calculator on skills tests? These judgments need to be made before these tests are academically upgraded.

What level of mathematics content should be assessed at different educational levels? The mathematics knowledge assessed on licensure tests for elementary, middle, and high school teachers (and perhaps special education teachers) should differ to some extent from test to test. But the mathematics knowledge that is needed on each test depends to some extent on what the public expects the teacher to teach. The mathematics coursework needed by prospective grade 11 or 12 mathematics teachers, and the mathematics assessed on their licensure tests, depends to a large extent on whether the public wants grades 11 and 12 teachers to be able to teach advanced courses in mathematics, e.g., Advanced Placement calculus AB or BC. Similarly, the coursework needed by grade 8 mathematics teachers, as well as the mathematics assessed on their licensure tests, depends to a large extent on whether the public wants grade 8 teachers to be able to teach algebra 1. And if the public wants elementary students prepared for a formal algebra I course in grades 7/8 (where it is taught in countries with high percentages of high achieving mathematics students in high school), that decision would clearly affect the kind of undergraduate mathematics coursework and licensure tests that prospective elementary teachers should take. Decisions on educational policy necessarily precede decisions on the mathematics content of licensure tests.

Similar considerations need to be explored for the many special education teachers in K-6 or K-8. How much mathematics knowledge should the public expect, from coursework and on licensure tests, of them? Should expectations be similar for those who teach in K-6 or K-8, whether they teach mainstream or special education students? Most licensure tests for special education teachers assess **no** subject matter knowledge at all. Addressing these teachers' deficiencies in mathematics (as well as in other subjects) through professional development is a Sisyphean task.

Another alternative also warrants consideration: how much mathematics knowledge is needed by full-time mathematics teachers in a subject-divided day in the elementary school, from grade 3

on, if not in the primary grades as well. The use of full-time mathematics (and science) teachers in the middle to upper elementary grades is a common practice in many other countries, and it is a practice that is beginning to grow in this country. Such a staffing strategy drastically reduces the number of K-6 teachers needing extensive if not perennial professional development in mathematics. The mathematics knowledge to be expected of elementary mathematics teachers and assessed on licensure tests would be less than the knowledge required of prospective middle school teachers, but how much less is a matter of professional judgment.

To illustrate concretely these differences in expectations, Appendix A shows one model of the differences in course requirements for elementary, middle, and high school mathematics teachers. Coverage of the topics listed in Appendix A is required in Massachusetts's licensure regulations, and their licensure tests must address these topics. There may well be other models in other states or on other tests for these three levels of mathematics teaching. Appendix B shows the pass rates on test administrations from May 2005 to May 2006 for these three tests. It also shows that a regular stream of candidates has been taking the test for the Elementary Mathematics license. (It should be noted that test-takers are not allowed to use calculators on the Elementary or Middle School Mathematics test.) We do not know exactly what this licensure test assesses (beyond what is stated in its objectives), how it does so (beyond what is available to inspect in its sample questions), its overall difficulty level in relation to the difficulty level of the items on the Middle School Mathematics test, how licensed elementary mathematics teachers function in their schools (e.g., as teachers of elementary mathematics or as mathematics coaches working with other teachers), and their effectiveness in comparison to teachers in self-contained classrooms. All this would be useful information for a study to gather for this licensure program and licensure test in a few states where the program and test are offered.

What should test items for prospective mathematics teachers assess? There seems to be general agreement that licensure tests for prospective teachers of mathematics, regardless of educational level, should assess their mathematics knowledge. Some mathematics educators are suggesting that these tests should also assess what they call instructionally relevant mathematics knowledge. Whether the tests should do so depends first on exactly what this knowledge base consists of at different grade levels. It also depends on whether this knowledge base encompasses what the teacher of a highly heterogeneous mathematics class would need to draw on for teaching the advanced students who might be in the class. How to determine the reaches of the advanced mathematics knowledge teachers would need to draw on for teaching the advanced mathematics students they might encounter in their classroom, and to what extent items reflecting that level of mathematics knowledge should be represented on a licensure test are only two of the important questions that need to be explored.

Much may also depend on the extent to which instructionally relevant mathematics knowledge can be taught in pre-service courses and to prospective teachers with minimal student teaching experience. It may be closer to the knowledge that is developed by alert teachers in the course of their teaching experience, based on observations of their students, analysis of their work, the research they read, and conversations with colleagues.

What types of pedagogical test items are useful for prospective teachers? As little as we know about the content of subject tests for prospective mathematics teachers, even less is known about the content of the pedagogical tests they are increasingly being required to take, such as the relatively new series developed by ETS called Principles of Learning and Teaching. The only study to date of the available information for these tests (Stotsky, 2006a) found no diversity of learning theories or teaching methods illustrated in the sample questions offered. Research-based instructional strategies are not illustrated at all. Instead, sample questions promote constructivist

pedagogies or discredit alternative pedagogies or both. To what extent these pedagogical tests (and the mathematics methods courses prospective teachers now take) inhibit new teachers' use of the mathematics knowledge they have warrants research.

Does the type of textbook make a difference for a new mathematics teacher? There are no rigorous studies comparing the mathematical demands of current textbooks on elementary or middle school teachers with respect to student achievement and their academic backgrounds. Many observers have noted that "reform" textbooks seem to require much more mathematical understanding on the part of the teacher than the textbooks they replaced. However, to my knowledge, there is no research on whether this aspect of "reform" textbooks leads to a much greater intellectual burden than is warranted for today's K-8 teachers, who on average have weaker academic backgrounds than did their counterparts years ago.

What is the content of the mathematics methods courses prospective teachers take? We do not know. There no research on the content of mathematics methods courses similar to recent studies on the syllabi used in reading methods courses (Walsh, 2006; Steiner & Rozen, 2004).

What is the quality of student teaching placements for prospective mathematics teachers? Again, we seem to know little if anything on a systematic basis about student teaching placements for prospective mathematics teachers—the most important component of teacher preparation—and the evaluation criteria used. While all training programs assess the performance of their student teachers, each training institution uses its own criteria for its assessments, leaving no basis for comparisons across training institutions within a state (never mind across states) of new teachers' pedagogical knowledge and skills, even for the same license. The Massachusetts Department of Education's experience in working with teacher educators across the state to develop sets of common license-specific performance criteria for use by its teacher training institutions when evaluating their student teachers suggested not only that no common criteria are used anywhere but also that a student teacher's use of his or her content knowledge in any area—in mathematics, science, or history—is *unlikely to be evaluated at all* (Stotsky, 2006b). The license-specific performance criteria developed in Massachusetts can be seen in Appendix C.

Does professional development make a difference? There is no clear and consistent evidence that an increase in teachers' mathematics knowledge through professional development leads to an increase in their students' achievement. For example, the Massachusetts Department of Education found no relationship between teacher gains in mathematical knowledge and student gains in a two-year experimental study from 2000-2002 with 36 teachers in a dozen low-performing middle schools. Teachers gained from the math courses they took, but their students didn't. The Department concluded that increased teacher knowledge may not quickly or even necessarily lead to gains in student learning for extremely low-achieving middle school students.¹

¹ This project sought to explore the effectiveness of mathematics coaching, carefully defined, using six full-time mathematics coaches who worked with over 40 teachers in these schools. The coaches were trained and supervised throughout the project. As part of the second year of the study, 36 teachers took a Department-sponsored middle school mathematics course taught in four locations by three mathematics professors using both a common syllabus and a pre-post test that they had developed. The teachers showed gains on the pre-post test that was given them. However, the students of teachers who took the course showed no greater gains overall than the students of a comparison group of teachers who were not enrolled in the course. Increased teacher knowledge may not quickly or even necessarily lead to gains in student learning for extremely low-achieving middle school students. The final report on the Middle School Mathematics Initiative, dated December 2002, is available from the University of Massachusetts Donahue Institute.

On the other hand, the National Science Foundation released a report on January 29, 2007 on the results to date of its Math and Science Partnership Program. This brief report indicates student gains in the one table it provides. However, the report provides no information on the length of the programs it funded, what tests were used to assess students across the states participating in these partnerships, if there were control groups, what curricula were used in the schools and in the professional development programs, what tests were used to assess teacher gains, how much mathematical knowledge teachers of these students gained, what kind of knowledge, and what the relationships were between teacher gains and student gains.

In yet another study, professional development did not seem to make a difference in teachers' knowledge. Hill, Rowan, & Ball (2005) showed that the students of primary grade teachers with higher levels of mathematics knowledge as measured by a multiple-choice test gained significantly more in mathematics over the course of the year than did the students of teachers with lower levels of mathematics knowledge. However, it is not known how the teachers with higher levels of mathematics knowledge gained their knowledge since all teachers in the study had already participated in many professional development workshops.

While most federal and state initiatives for professional development are now based on the assumption that increasing teachers' mathematics knowledge leads to greater gains in students' mathematics achievement, much may depend on the students' level of achievement in relation to their grade level and on the curriculum materials the teachers use (as we know from Reading First). In future studies of the effects of professional development programs in mathematics on students achievement, it would seem prudent to explore the possibility that their relationship to student achievement might be differentially influenced by the kind of mathematics textbooks that teachers use.

Concluding Remarks

International studies show that a much smaller percentage of American students are among the highest performers on mathematics tests compared with the percentage of their age peers in other nations. For example, Takahira et al (1998) found that "the most advanced mathematics students in the United States, about 5 percent of the total age cohort, performed similarly to 10 to 20 percent of the age cohort in most of the other countries" (p. 44).

Countries that are leaving us behind tend to have a few things in common. They decide at the national level what mathematics students should learn and when. They also appear to have higher and more uniform standards for the academic preparation of their teachers. As ETS noted in a 2003 report comparing our teacher education "pipeline" to the pipelines in seven other countries whose 1999 TIMSS scores in grade 8 mathematics or science were as good as or better than ours, other countries tend to "frontload" requirements. That is, they emphasize selection into and from preparation programs, rather than "backload" academic requirements after entry into the profession, as we do. Apprenticeships and other forms of clinical training may also be stricter in other professions in this country. In a comparison of teacher preparation with preparation in six other fields including accountants, architects, nurses, and lawyers (in a multi-state component of the bar exam), a study by the Finance Project noted that student teaching is markedly less structured and less supervised than the clinical training required in these other fields (Neville, Sherman, & Cohen, 2005).

Most American teachers of mathematics seem to require intensive and expensive professional development in the knowledge base for the subject they teach. However, there is no evidence that their understanding of mathematics can be increased more effectively **after** they begin teaching,

through professional development, than **before** they begin teaching, whether through regular mathematics coursework, specially designed mathematics courses, and/or mathematics methods coursework. Nor is there as yet clear evidence showing that an increase in teachers' mathematics knowledge through professional development leads to an increase in their students' mathematics achievement. While many current federal and state funding initiatives are predicated on the assumption that increasing teachers' mathematics knowledge leads to greater gains in students' mathematics achievement, much may depend on the students' level of achievement in relation to their grade level and the curriculum materials the teachers use.

As with the reform of medical education in the early 20th century, rigorous admission and exit requirements for teacher training programs may be the most significant steps that could be taken to upgrade the academic quality of new teachers of mathematics. In fact, the need for strong exit requirements for teachers was highlighted in a blue-ribbon panel report released by the Teaching Commission in 2004, chaired by Louis Gerstner, former Chief Executive Officer of IBM. One of the report's four recommendations was to strengthen current teacher tests by raising their passing scores and replacing "low-level basic competency tests with challenging exams that measure verbal ability and content knowledge at an appropriately high level." As with medical education, these steps need to be accompanied by profound changes in the training programs themselves, as outlined by Arthur Levine, former president of Teachers College, Columbia University, in his report *Educating School Teachers* (2006). These changes should be based on a valid body of information.

Research to address each of the above questions and issues would provide states with information to strengthen their licensure requirements and the licensure tests they develop or use for the many kinds of teachers who teach mathematics in some form from K-12. The information would also be useful in approving or accrediting traditional or alternative preparation programs.

References

Angrist, J. & Guryan, J. (2003). Does Teacher Testing Raise Teacher Quality: Evidence from State Certification Requirements. NBER Working Paper 9545. National Bureau of Economic Research.

Boot, J. (2006). Student Achievement and Passport to Teaching Certification in Elementary Education. Washington, D.C.: American Board for Certification of Teacher Excellence.

Boot, J. (April, 2007). Student Achievement and Passport to Teaching Certification in Mathematics.. Washington, D.C.: American Board for Certification of Teacher Excellence. http://www.abcte.org/files/math_2007_validity.pdf Retrieved April 22, 2007.

Clotfelter, C., Ladd, H., & Vigdor, J. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. Working Paper 11936. Cambridge: National Bureau of Economic Research. <<http://www.nber.org/papers/w11936>>

Goldhaber, D. (2006). Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness? University of Washington and the Urban Institute.

Hampel, R. (October 19, 2005). Doctoring Schools: The Medical Model and Teacher Training: A Historical Perspective. Education Week.

Hill, H.C., Rowan, B., & Ball, D. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. American Educational Research Journal, 42 (2), 371- 406.

Hill, H., Sleep, L., Lewis, J. & Ball, D. (2006). Assessing Teachers' Mathematical Knowledge: What Knowledge Matters and What Evidence Counts? In *Handbook on Teacher Assessment*.

- Hoxby, C. & Leigh, A. (2005). Wage Distortion: Why America's Top Women College Graduates Aren't Teaching. Education Next, Spring. <www.educationnext.org/20052/50.html>
- Kane, T., Rockoff, J., & Staiger, D. (2006). What does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. Cambridge: Harvard Graduate School of Education.
- Levine, A. (2006). Educating School Teachers. Washington, DC: The Education Schools Project.
- Mehrens, W., Klein, S., & Gabrys, R. (January 14, 2002). Report by the Technical Advisory Committee on the Massachusetts Tests for Educator Licensure. Submitted to the Massachusetts Department of Education and Commissioner of Education. <http://www.doe.mass.edu/mtel/news02/TAC_rep.pdf>
- Mitchell, K., Robinson, D., Plake, B., & Knowles, K. (2001). Testing Teacher Candidates: The Role of Licensure Tests in Improving Teacher Quality. Committee on Assessment and Teacher Quality. National Research Council. Washington, D.C.: National Academy Press.
- Mitchell, R. & Barth, P. (1999). How Teacher Licensing Tests Fall Short Thinking K-16. Education Trust, Volume 3, Issue 1.
- Monk, D. (1994). Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement. Economics of Education Review, 13 (2), 142. pp. 125-142.
- Neville, K., Sherman, R., & Cohen, C. (2005). Preparing and Training Professionals: Comparing Education to Six Other Fields. Washington, D.C.: The Finance Project.
- Peterson, P. & Hess, R. (2005). Johnnie Can Read...in Some States. Education Next.
- Strauss, R. (1998). Teacher Preparation and Selection in Pennsylvania. Research Report to the Pennsylvania State Board of Education.
- Rigden, D. (2006). Report on Licensure Alignment with the Essential Components of Effective Reading Instruction. Report commissioned by the National Council for Accreditation of Teacher Education. http://www.rften.org/content/Rigden_Report_9_7_06.pdf?PHPSESSID=8bad09b6a0936998ada5c4074c4394ad.
- Stotsky, S. (2006a). Why American students do not learn to read very well: The unintended consequences of Title II and teacher testing. Nonpartisan Education Review, 2(1). Retrieved September 8, 2006 from <http://nonpartisaneducation.org/Review/Articles/v2n1.pdf>.
- Stotsky, S. (2006b). Who should be accountable for what beginning teachers need to know? Journal of Teacher Education, 57 (3), 256-258. <http://JTE.sagepub.com/content/vol57/issue3>.
- Steiner, D. with Rozen, S. (2004). Preparing Tomorrow's Teachers: An Analysis of Syllabi from a Sample of America's Schools of Education. In F.M. Hess, A.J. Rotherham, & K. Walsh (Eds.). A Qualified Teacher in Every Classroom? Appraising Old Answers and New Ideas. Cambridge: Harvard Education Press.
- Takahira, S., Gonzales, P., Frase, M., & Salganik, L. (1998). Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in an International Context. U.S. Department of Education.
- The Teaching Commission. (2004). *Teaching at Risk: A Call to Action*. NY: CUNY Graduate Center, Manhattan,
- Walsh, K., Glaser, D., & Wilcox, D.D. (2006). What Education Schools Aren't Teaching about Reading, and What Elementary Teachers Aren't Learning. National Council on Teacher Quality.
- Wang, A., Coleman, A., Coley, R., & Phelps, R. (2003). Preparing Teachers Around the World. Policy Information Report. NJ: Educational Testing Service.

Appendix A: Topics for the Elementary, Middle, and High School Mathematics Licensure Tests in Massachusetts

- (a) The following topics will be addressed on a subject matter knowledge test for the 1-6 level:
1. Basic principles and concepts related to elementary school mathematics in the areas of number sense and numeration, patterns and functions, geometry and measurement, and data analysis.
 2. Algebra.
 3. Euclidean geometry
- (b) The following topics will be addressed on a subject matter knowledge test for the 5-8 level:
1. Algebra.
 2. Euclidean geometry.
 3. Trigonometry.
 4. Discrete/finite mathematics.
 5. Introductory calculus through integration.
 6. History of mathematics.
- (c) The topics set forth in (b) above and the following topics will be addressed on a subject matter knowledge test for the 8-12 level:
1. Abstract algebra.
 2. Number theory.
 3. Calculus through differential equations.
 4. Probability and statistics.
 5. Non-Euclidean and transformational geometries.
 6. Applied mathematics or mathematics modeling.

Source: *Massachusetts Regulations for Educator Licensure and Preparation Program Approval*. Massachusetts Department of Education, June 2003.

Appendix B: Pass Scores by Test Administration from May 2005-May 2006 on Three Mathematics Tests for Teacher Licensure in Massachusetts*

Test Administration: May 2006

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	207	60.4	98	43.9
Middle School Mathematics	200	53.0	104	36.5
Elementary Mathematics	45	53.3	11	63.6

Test Administration: March 2006

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	207	60.4	98	43.9
Middle School Mathematics	200	53.0	104	36.5
Elementary Mathematics	45	53.3	11	63.6

Education Working Paper Archive

Mathematics	214	67.3	81	38.3
Middle School Mathematics	187	54.0	116	38.8
Elementary Mathematics	52	65.4	6	33.3

Test Administration: November 2005

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	150	58.7	67	37.3
Middle School Mathematics	154	45.5	77	29.9
Elementary Mathematics	30	63.3	5	60.0

Test Administration: September 2005

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	85	58.8	28	39.3
Middle School Mathematics	66	57.6	40	30.0
Elementary Mathematics	11	45.5	1	0.0

Test Administration: July 2005

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	190	67.9	69	43.5
Middle School Mathematics	168	57.7	91	41.8
Elementary Mathematics	45	55.6	2	50.0

Test Administration: May 2005

Test Name	First-Time Test Takers		Test Retakers	
	N	% Passing	N	% Passing
Mathematics	207	68.1	60	41.7
Middle School Mathematics	219	57.1	104	51.0
Elementary Mathematics	51	72.5	3	66.7

*The Mathematics license covers grades 8-12. The Middle School Mathematics license covers grades 5-8
The Elementary Mathematics license covers grades 1-6.

Source: Massachusetts Department of Education (www.doe.mass.edu)

Appendix C: License-Specific Evaluation Questions for Prospective Mathematics Teachers in Massachusetts

1. Does the candidate appropriately balance activities for developing conceptual and procedural knowledge of mathematics?
2. Does the candidate use multiple representations of concepts such as numerals or diagrams, algebraic expressions or graphics, or matrices that model a method for solving a system of equations?
3. Are manipulatives and concrete representations used when appropriate?
4. Does the candidate help students to learn alternate methods of solving mathematics problems?
5. Are students' mathematical misconceptions identified and addressed?
6. Does the candidate model clear mathematical reasoning when helping students solve mathematics problems?
7. Does the candidate know how to teach the standard algorithms for arithmetical operations and teach them to students?
8. Does the candidate refer to the appropriate level of the state's mathematics standards to prepare a lesson?
9. Is the candidate's explanation of mathematical concepts accurate?
10. Does the candidate expect students to use accurate mathematical language to talk and write about mathematics?

Source: Guidelines for Preservice Performance Assessment. Massachusetts Department of Education, 2006. Retrieved from http://www.doe.mass.edu/edprep/ppa_guidelines.pdf