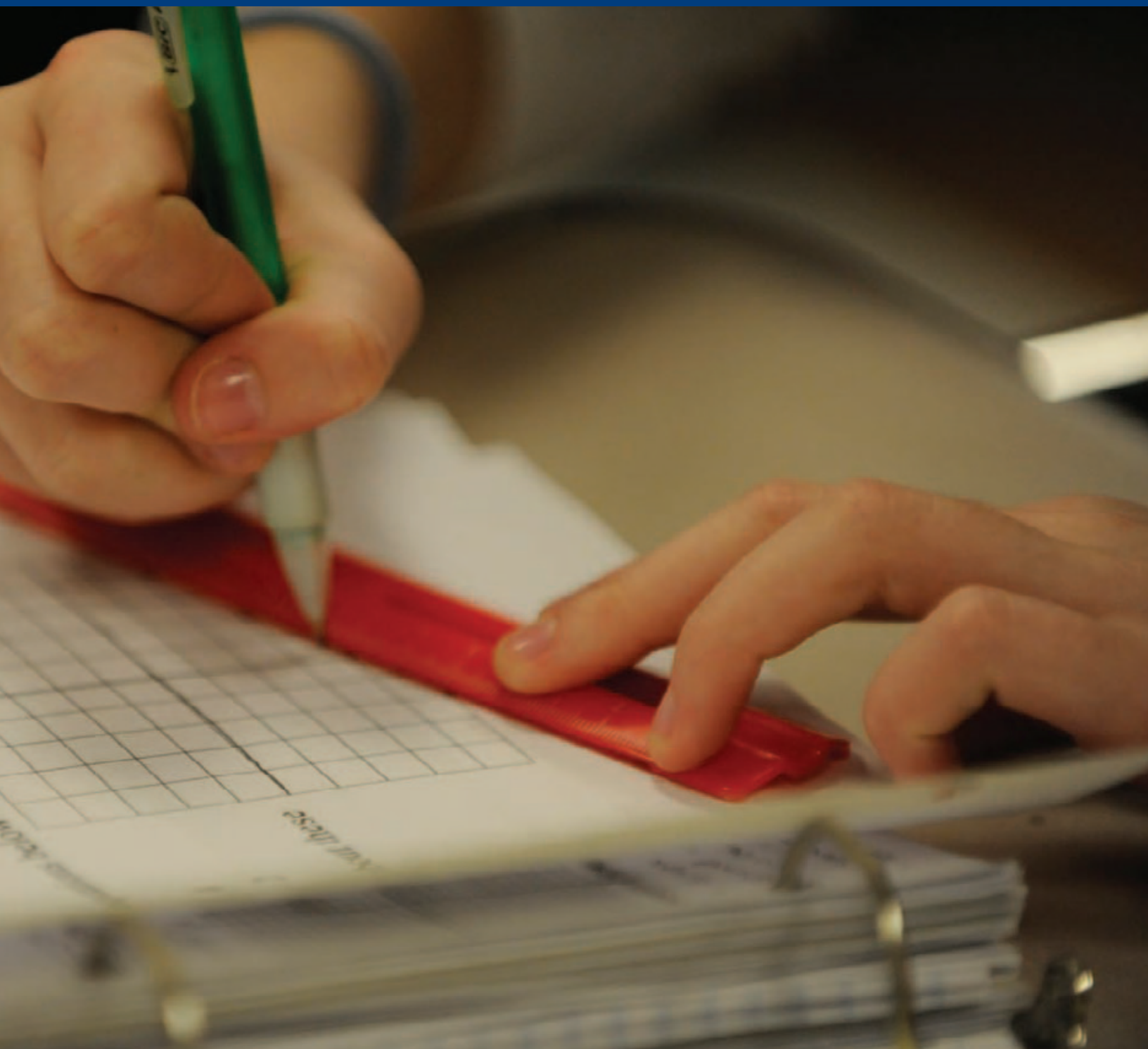
U.S. DEPARTMENT OF EDUCATION

Recent Trends in Mean Scores and Characteristics of Test-Takers on *Praxis II* Licensure Tests



Recent Trends in Mean Scores and Characteristics of Test-Takers on *Praxis II* Licensure Tests

Prepared by:

Drew H. Gitomer

Yi Qi

Educational Testing Service (ETS)

Principal Investigator

Drew H. Gitomer, ETS

Prepared for:

U.S. Department of Education
Office of Planning, Evaluation and Policy Development
Policy and Program Studies Service

2010

This report was prepared for the U.S. Department of Education under Contract Number ED-08-PO-0730 with the Educational Testing Service. Elizabeth Eisner served as contracting officer's representative for this report. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. Department of Education
Arne Duncan
Secretary

Office of Planning, Evaluation and Policy Development
Carmel Martin
Assistant Secretary

Policy and Program Studies Service
Alan Ginsburg
Director

April 2010

The United States Department of Education asked the Educational Testing Service (ETS) to conduct this study because ETS was in a unique position to conduct these analyses at the lowest cost to the government. ETS maintains all of the *Praxis II* testing data from all states that administer the *Praxis II* exams. ETS has permission from the states to conduct analyses on these data as long as the analyses involve the aggregation of the test scores from multiple states and do not identify the names of individual states. In order for another research firm to obtain the data and conduct a comparable analysis, another firm would have had to seek and obtain permission from each state to use that state's Praxis data. Moreover, given the extent of this study, any other research firm would have been required to obtain approval for the collection of data from the Office of Management and Budget (OMB). This process of obtaining OMB approval, including the period for public comment on the information requests, would take at least 120 days, and involves the additional cost of preparing an extensive OMB clearance package.

CONTENTS

EXHIBITS

Executive Summary

II. Methodology

III. Findings

A. Test 0012—Elementary Education: Pedagogy

B. Test 0014—Elementary Education Content Knowledge

APPENDIX A

APPENDIX B

APPENDIX C

ACKNOWLEDGMENTS

We are grateful to a number of colleagues who assisted us with various aspects of the project. Elizabeth Eisner, the Department of Education contracting officer, and was a tremendous support throughout the course of the project. We also thank the following reviewers for their thoughtful comments that helped us clarify our ideas and improve the report—Carmel Martin, assistant secretary, Office of Planning, Evaluation and Policy Development (OPEPD); Judy Wurtzel, deputy assistant secretary, OPEPD; Alan Ginsburg, director, Policy and Program Studies Service (PPSS); David Goodwin, former director, Program and Analytic Studies, PPSS; Daphne Kaplan, a PPSS team leader; and these reviewers from other ED offices: Betsy Warner (IES); Aaron Young (Budget Service/OPEPD); Libby Witt (OESE); Richard Mellman (OGC); and Nettie Laster (OII).

We appreciate the support of a number of ETS colleagues. Sailesh Vezzu helped create the dataset we used for data analyses. Sharon Temmer supported our requests for detailed information about the *Praxis* assessments. Evelyn Fisch provided outstanding administrative and editorial support.

Of course, any errors or limitations of the paper are the authors' responsibility alone.

EXECUTIVE SUMMARY

For well over a century, there have been significant concerns about the academic quality of those who choose teaching as a profession (see Lageman, 2000; National Educational Association 1895). More recently, a number of federal, state, and local policies instituted during the last 10–15 years were designed to ensure that teachers have appropriate backgrounds in the content they teach. These policies have included increased accountability of teacher preparation programs, expansion of teacher licensure testing to ensure that all teachers demonstrate subject matter competence, increased requirements for entry into teacher preparation programs, strengthened requirements for program accreditation, and increased support for development and expansion of alternate route programs.

This study concerns the 'highly qualified teacher' provisions of the *Elementary and Secondary Education Act* (*ESEA*, 2002), as reauthorized, and other policies at the federal, state and local levels, which have aimed to elevate the content knowledge of teachers. This examination of *Praxis II* score trends was not meant to serve as an evaluation of *ESEA* or of any of the other policies instituted at the state or local levels; the study served simply as an effort to examine trends in a systemic outcome that all of these policies have aimed to influence—the content knowledge of prospective teachers. More specifically, this report answers the following questions:

1. Have the demographic and/or performance characteristics of individuals who took each test changed across the years included? Total sample size, proportions by gender and race, percent passing, and experience breakdowns are reported.[1]

2. Do mean *Praxis*[2] scores differ by gender, race, and experience? Additionally, do the mean passing scores for those who pass differ substantially from those who do not pass as well as from the mean passing scores of all candidates considered together?

3a. Across years, is there any evidence of a trend in *Praxis* scores for those who pass? How do these trends compare with the trends of those who do not pass?

3b. Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

This study examines changes in teacher licensure scores from 1999 to 2006. We focus specifically on tests in the *Praxis II* series because these tests are the most widely used assessments across multiple states for purposes of measuring content knowledge for initial teacher licensure. The purpose of this study is to identify trends in *Praxis* scores on a select number of tests across recent years and across as many states as possible. While teachers in states using *Praxis* are not fully representative of the national population of teachers, data from

[1] We exclude results that have insufficient sample size as defined by cell sizes under 15. For several models we combine the two experience levels (one-to-three years and greater than three years) in order to obtain a sufficient sample. Specifics are reported in the text.

[2] ETS and *PRAXIS* II are registered trademarks of Educational Testing Service (ETS). *PRAXIS* is a trademark of ETS.

Praxis testing provides the most complete available picture of licensure test performance for the population of those entering teaching.

The study includes licensure candidates who took the following tests (including *Praxis* test code number) in multiple states:

- Elementary Education: Pedagogy (0012),
- Elementary Education Content Knowledge (0014),
- English Language, Literature, and Composition: Content Knowledge (0041),
- English Language, Literature, and Composition: Pedagogy (0043),
- Mathematics: Content Knowledge (0061),
- Mathematics: Pedagogy (0065),
- Middle School Mathematics: Content Knowledge (0069),
- General Science, Part 1: Content Knowledge (0431), and
- Earth and Space Sciences: Content Knowledge (0571)

The study focuses on trends in mean scores for those who pass the *Praxis* tests, as these are individuals who are eligible to enter teaching. Analyses are disaggregated by passing status, gender, race, and whether or not the test candidate has prior teaching experience.

The adoption of tests by states varies widely. States have adopted different tests at different times and frequently applied different passing standards. The study attempted to compare test results across years by including a stable set of states within each analysis and applying appropriate state- and year-specific passing standards for each test. Thus, different combinations of states and years are used in each of the analyses. Because states have been transitioning to new testing requirements, some analyses only include three years of data, while others include up to eight years. Some analyses include six states, while others include 18 states. Trends within individual states are not included; all analyses aggregate test data across a minimum of six states.

Multiple regression models were used to analyze trends in *Praxis* scores across years. One set of models examined only changes in scores across years. A second set of models also controlled for gender, race, and prior teaching experience.

Key Findings:

- **For the *Praxis* tests examined in this study, there is little change in mean test scores observed over time.** While the study identifies many significant trend effects, the magnitude of these effects is relatively small. Because the samples include thousands of individuals, even small differences can be statistically significant, though substantively unimportant.

  It is also true that passing rates for many of the tests have decreased over time. Part of the trend may be due to the increase in passing scores established by states during the course of the study. However, that may not be the only reason and should be explored further. There is a need to understand who is taking the tests, including the nature of their preparation and academic qualifications. Further work needs to explore the extent to which failure rate changes can be explained by the increase in passing scores.

- **Those who pass *Praxis* tests have scores substantially higher than those who do not.** Individuals who pass these tests have mean and median scores that are approximately two standard deviations higher than those who fail. Standard deviations are based on all test-takers from the respective samples. This pattern is consistent across all tests examined in this study (see Exhibit ES.1). This finding has two practical implications. First, licensure tests are filtering out individuals who attain very low scores on tests of content knowledge. Second, it is unlikely that many of these low scoring individuals will achieve a passing score simply through taking the test multiple times without learning more of the content that is measured on the test. The magnitude of these differences far exceeds any score variations that might be attributed to measurement error. The mean score differences between those who pass and those who fail has remained stable across year as illustrated in the example from the Mathematics: Content Knowledge from 1999–2006 (see Exhibit ES.2).

Exhibit ES.1
Mean *Praxis* Scores (median) and Standard Deviation by Test by Passing Status Across Years

Test	Passing Mean (Median)	Failing Mean (Median)	Standard Deviation
Elementary Education: Pedagogy	158 (157)	135 (136)	10.0
Elementary Education: Content	166 (165)	132 (133)	16.8
English Language, Literature, and Composition: Content Knowledge	180 (180)	151 (152)	14.7
English Language, Literature, and Composition: Pedagogy	158 (155)	135 (135)	13.7
Mathematics: Content Knowledge	153 (151)	117 (117)	20.9
Mathematics: Pedagogy	152 (150)	116 (115)	20.7
Middle School Mathematics: Content Knowledge	169 (167)	140 (140)	17.7
General Science, Part 1: Content Knowledge	169 (168)	139 (140)	15.5
Earth and Space Sciences: Content Knowledge	169 (169)	136 (138)	17.5

Exhibit Reads: The mean and median scores for individuals who pass the *Praxis* tests are substantially higher than scores for those who fail the *Praxis* tests.

Exhibit ES.2
Mean *Praxis* Scores of Test-takers by Passing Status—Mathematics: Content Knowledge

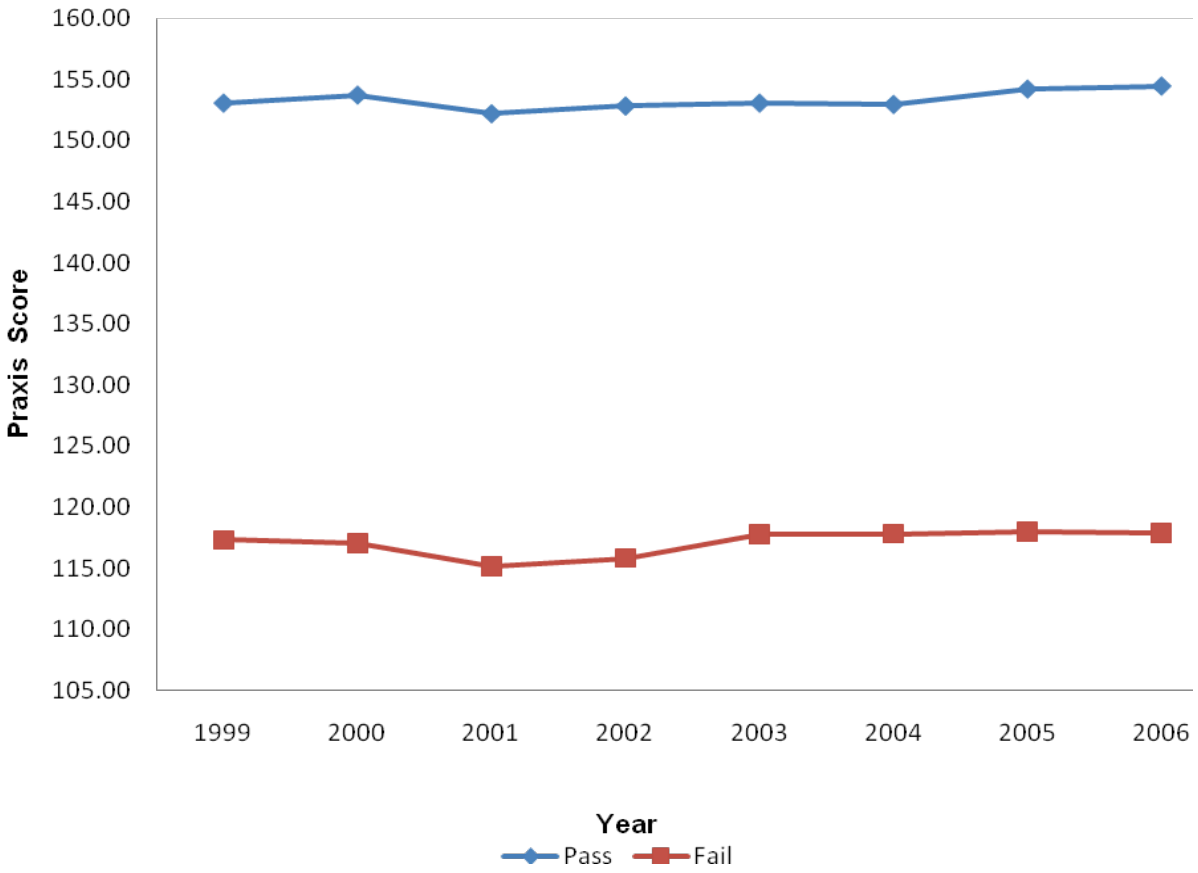


Exhibit Reads: Mean scores for those who pass the Mathematics: Content Knowledge test are substantially higher than for those who do not.

This very large disparity in scores between those who pass and those who do not means that there are *not* large numbers of candidates hovering near a passing standard such that if people simply took a test one or two more times their status might change without their actually learning more of the content that is tested. The large number of individuals who fall substantially below the passing standards are evident from examining the distribution in the example below, taken from the Mathematics: Content Knowledge test (see Exhibit ES.3). In fact 75 percent of all those who do not pass the test have scores at least one standard error of measurement below the median passing score across states.

Exhibit ES.3
Distribution of Scores in 2006 for Mathematics: Content Knowledge

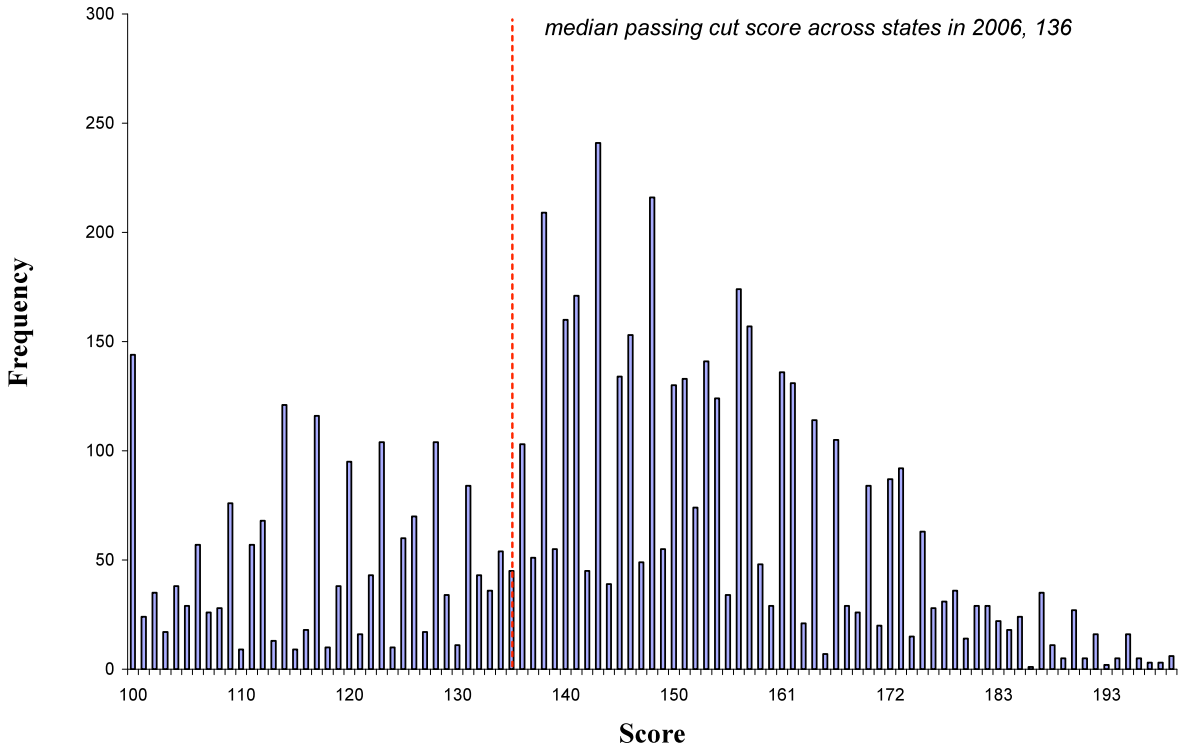


Exhibit Reads: Many test-takers have scores substantially below state passing standards.

- **It is critical to separate those who pass from those who do not in order to make sense of trends.** The study clarifies why it is so important to distinguish the sample of passers from the population of all test-takers. Different inferences are made about trends depending on which sample is used. We contend that looking at passers is the most policy relevant focus, for these are the individuals who are actually meeting the teaching qualifications established by states. For two reasons, the legitimate policy focus is on the qualifications of prospective teachers, not prospective test-takers.

  First, there is no prerequisite for taking a *Praxis* test. Any interested individual can take a *Praxis* test, regardless of academic or preparation history. Second, many states have increased passing standards so that the same candidates with lower scores would no longer satisfy state requirements. Increasing the passing standard not only elevates the mean score for those who pass, it also has the consequence of increasing failure rates if the same hypothetical population of individuals with the same qualifications were to take the test across years. This tradeoff is an explicit policy decision by states to improve the academic quality of the teaching force. By looking separately at those who pass and those who do not, changes in average scores can be better understood, particularly for those who qualify to enter the teaching profession.

Indeed, when looking at only those who pass the test, trends are either positive or stable over time for the majority of tests. Interestingly, trends are also often positive for those who do not pass. Yet, if all candidates are included, the majority of tests show a decline. At first blush, it appears counterintuitive that those who pass and those who fail show a positive trend, yet when combined there is a negative trend.

In fact, this curious finding is an instance of a well-known statistical phenomenon known as *Simpson's Paradox,* in which subgroups can each show the same pattern, yet when combined, a different pattern is observed because the relative proportion of individuals in each subgroup changes. For certain tests, the proportion of those taking the tests who failed increased relative to those who passed. Thus, the increasing proportion of candidates who fail the test can result in the overall means decreasing at the same time that the mean scores for both those who pass and those who fail either increase or remain stable.

One example of this phenomenon was observed for the Mathematics Content test in considering candidates new to teaching and those with prior experience (see Exhibit ES.4). The top two lines in the figure show that scores for those who passed increased slightly across years, as did scores for those who failed (the bottom two lines). However, when all test-takers are included (the middle two lines) there is an observed decrease, particularly for those test-takers with prior experience. This is a result of the substantial increases in failure rates, displayed at the bottom of the figure. In the full report, we further disaggregate the experience factor by presenting separately results of those with one-to-three years of experience and those with more than three years' experience.

Exhibit ES.4
Mean *Praxis* scores of Test-takers by Passing Status and Teaching Experience—Mathematics: Content Knowledge

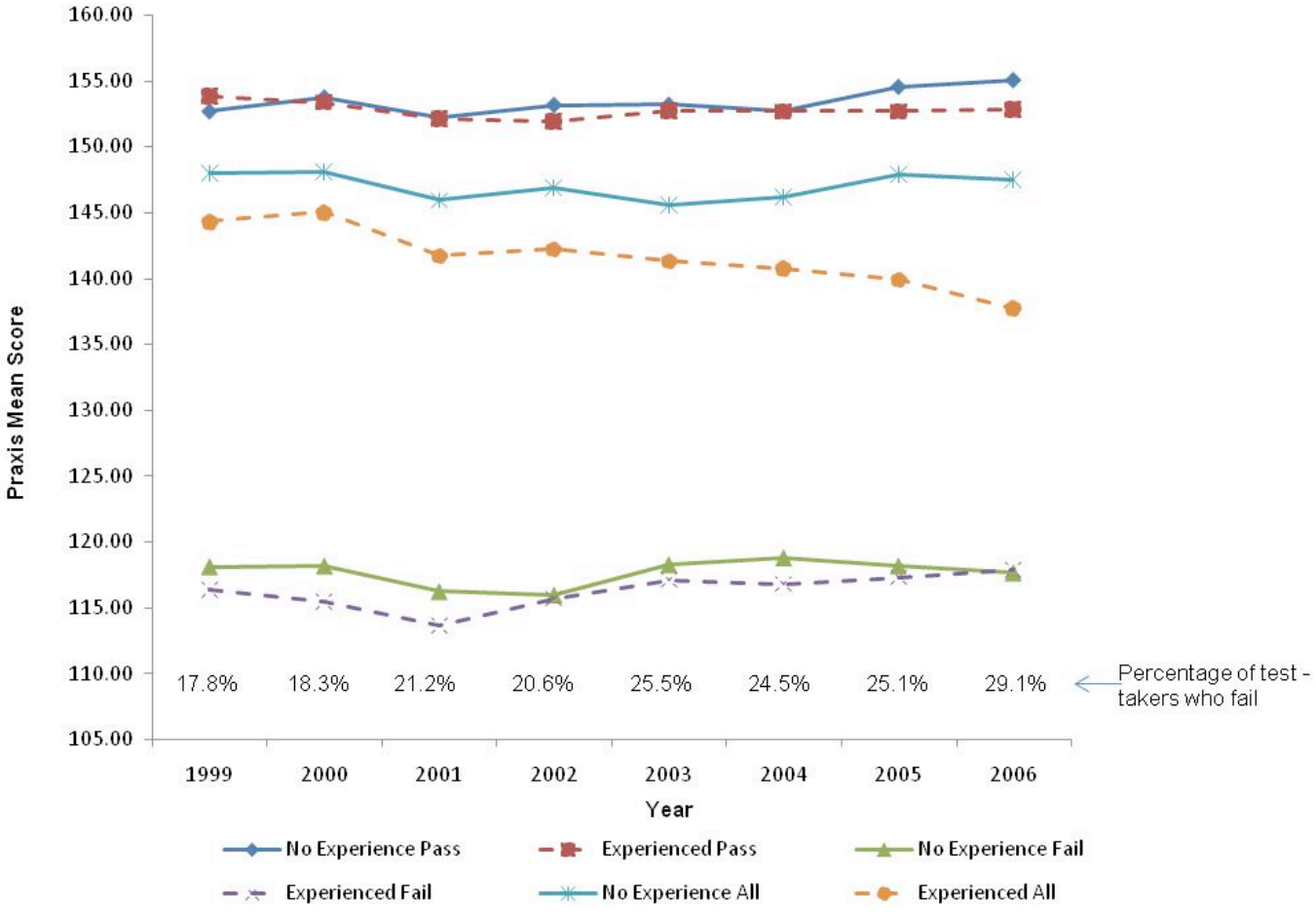


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Failure rates increase substantially across testing years.

- **Score trends are inconsistent across tests.** Trends vary across the nine tests included in this study. For those who pass, three tests show positive performance trends, two show declines, and the remaining four do not have a trend in any particular direction (see Exhibit ES.5).

Exhibit ES.5
Study Summary of Trends in Scores for Those Who Pass

Test	Number Years	Direction of Trend Across Years
0012 Elementary Education: Pedagogy	6	Negative
0014 Elementary Education: Content Knowledge	3	No Trend
0041 English Language, Literature, and Composition: Content Knowledge	8	No Trend
0043 English Language, Literature, and Composition: Pedagogy	8	No Trend
0061 Mathematics: Content Knowledge	8	Positive
0065 Mathematics: Pedagogy	8	Positive
0069 Middle School Mathematics: Content Knowledge	3	No Trend
0431 General Science, Part 1: Content Knowledge	8	Positive
0571 Earth and Space Sciences: Content Knowledge	5	Negative

Exhibit Reads: **Tests vary in the number of years included in the analysis and in trends over years.**

- **There has been an increase in the number of individuals taking the *Praxis II* tests and the increase is seen both among individuals with teaching experience and without teaching experience.** For many of the tests included in the analyses, the size of the sample of test-takers nearly doubled over the different range of years included. At the same time that a substantial increase in test-takers was occurring, there was very little movement in scores. There is no clear rationale for why this pattern of data exists. It appears that the increase in candidates has not materially affected the overall preparation of individuals to succeed on the respective tests.

- **Race and ethnicity differences are consistent across tests.** Across all *tests,* white candidates who pass have higher mean scores than African-American candidates.

- **Race and ethnicity differences are not disaggregated by gender because of small sample sizes in some of the cells.**

- **Gender differences vary by test.** Female candidates who pass have higher mean scores than male candidates on two of the three pedagogy tests included in the analysis. Male

candidates who pass have higher mean scores on the six content knowledge tests analyzed. Gender differences tend to be relatively small.

- **For most of the states examined in this study, passing standards are related to mean passing scores and passing rates in predictable ways.** Given relatively sparse data, there does appear to be a general pattern that states with higher passing standards have candidates who have higher mean scores. Passing rates are typically lower as well, though there are notable exceptions.

- **Experience effects vary across tests**. Test-takers who pass and have prior teaching experience have slightly higher scores than those without experience for five tests. Those without experience have higher scores for both mathematics content tests. Experience differences tend to be relatively small.

- **Experience factors are relatively small except for the Middle School Mathematics test.** For this test, those new to the field of teaching score substantially higher than more experienced test-takers. This finding is consistent with that of Gitomer (2007), which showed that the middle school test was being taken by a large number of elementary trained teachers who were teaching in middle school and needed to be subject matter certified under the Highly Qualified Teacher provision of *ESEA*. It is very possible that this group of teachers was not as well prepared in mathematics as newly prepared math teachers who are targeting middle school for a career path.

- **The large, positive trends in SAT scores by *Praxis II* candidates observed in prior research are not echoed here in similarly large positive trends in *Praxis II* scores.** This may be due to the more limited number of years included in this study as well as several other factors. Gitomer (2007), primarily on the basis of SAT scores, concluded that the academic quality of teachers had improved a substantial amount over recent years. This begs the question of whether there are reasons to explain the current study's more modest findings compared with the prior study. We believe there are a number of critical differences between these studies that might account for the somewhat different findings.

  Perhaps the most important distinction concerns the cohorts that were sampled. Gitomer (2007) compared *Praxis* candidates from 1994–97 to those from 2002–05, which actually contains data that vary by 11 years (1994–2005). For the present study, no data are included prior to 1999. It was during the late 1990s that concerted efforts were made to include testing as part of policies to address teacher quality issues. As examples, the *Higher Education Act* was reauthorized and included requirements for teacher education institutions to report their students' licensure test performance. The National Council for the Accreditation of Teacher Education (NCATE) developed new standards for accreditation that focused much more on student performance in teacher education programs. Other state and institutional efforts also were being put in place at this time.

  Thus, if these policies had an effect, it is possible that they were more visible using a lens that begins in 1994. It is plausible that any changes brought about by this confluence of policies were in place by 1999. Indeed, if we look at just the five tests in this study for which there are eight years of data, we obtain a more positive view than one which looks at all nine tests, including those using a briefer lens. Of the five tests with data going back to 1999, three show positive trends and two show no trend.

It is possible that more substantial changes would be evident if we could include the years 1994–98. Unfortunately, this is just speculative and there is no way of examining this hypothesis. However, given when the policy activity occurred, changing the window of examination may explain the variation in findings.

Of course, the SAT and *Praxis* tests are designed to measure different things and though correlated, are independent measures. Thus, there is no reason to believe that scores on the two measures would show trends in lockstep. Further, these measures may be more or less sensitive to policy changes and instructional interventions. On the one hand, if greater emphasis is placed on entrance into teacher education, then we might expect an increase in college admissions scores, particularly if those scores are part of the admissions equation. If *Praxis* scores reflect an undergraduate curriculum that has stayed fairly consistent across years, we might expect less change in those scores. Again, this is speculative, but it creates hypotheses that might be worth further study.

Finally, we do not understand how the increased presence of alternate route candidates influences trends in test scores. It is possible that many of these test-takers bring in relatively strong academic skills as measured by SAT scores. However, they may have a wide range of course work as background which may or may not prepare them well for the *Praxis* tests. The current system of data collection does not capture the nature of teacher preparation routes, so this issue would need to be studied separately.

Taken together, for a limited set of *Praxis* tests across a number of recent years, we see little or no change in scores. Policies that gave significant attention to teacher licensure test performance, if they did have an impact as suggested by Gitomer (2007), may have already had their effect prior to the years studied in this analysis.

I. INTRODUCTION

For well over a century there have been significant concerns about the academic quality of those who choose teaching as a profession (Carnegie Forum on Education and the Economy, 1986; Coleman, et al., 1966; Greenwald, Hedges, and Laine, 1966; Kerr, 1983; Koerner, 1963; Lagemann, 2000; National Educational Association, 1895). These concerns have focused on both general verbal and mathematical abilities as well as specific content knowledge.

A number of policies have been instituted during the last 10–15 years designed to ensure that teachers have an appropriate background in the content they teach. These policies, reviewed in Gitomer (2007) have included:

- Increased accountability of teacher education programs through the 1998 reauthorization of the *Higher Education Act* (and *Amendments to the Higher Education Act of 1965*, *1998*) that required programs to report on licensure test results of their program completers.

- An expansion of teacher licensure testing to ensure that all teachers demonstrated competence to teach particular subject matter as specified in the Highly Qualified Teacher provisions of the federal *Elementary and Secondary Education Act* (*ESEA*, 2002), as reauthorized.

- Increased requirements for entry into teacher education programs by setting minimum grade-point averages (GPAs) and other admissions criteria (e.g., state of Pennsylvania, 2007).

- Strengthened requirements for program accreditation by the professional accrediting bodies, National Council for Accreditation of Teacher Education (NCATE, 2006) and the Teacher Education Accreditation Council (TEAC, n.d.).

- The rapid development and expansion of alternate route programs, many of which were designed to lure academically strong candidates who were not attracted to teaching if it required entry into a formal teacher education program (Feistritzer, 2007).

In view of this range of initiatives intended to improve the academic quality of teachers, Gitomer (2007) compared the academic profile of prospective teachers during the years 2002–05 with similar individuals from the years 1994–97 (see Gitomer, Latham, and Ziomek, 1999, for the original study). Specifically, we compared SAT scores for individuals who took a teacher licensure test (*Praxis II*) across these two cohorts. SAT scores were considered to be a common, broad proxy of academic quality. In fact, both SAT math and verbal scores were significantly higher for the more recent cohort.

The present study is designed to shine a different lens on changes in the academic quality of prospective teachers. Specifically, this study examines changes in teacher licensure scores themselves over time. We focus specifically on tests in the *Praxis II* series. These tests are

developed and administered by Educational Testing Service (ETS)[3] and are used by a majority of states as a critical part of the teacher certification process.

Of course, not all states use the *Praxis* assessments for licensure. A number of states, including those with the largest population of teachers (e.g., Calif., N.Y., Texas, Mich., and Fla.), have state-specific licensure tests. Because these tests are unique to particular states, it is not possible to compare performance across these states. Thus, while the sample used in the current study cannot be said to be fully representative of all teacher candidates, it is by far, the most complete sample of teacher candidates available to compare teacher candidates across states.

This study concerns the Highly Qualified Teacher provisions of the *ESEA*, and other policies at the federal, state and local levels, which have aimed to elevate the content knowledge of teachers. This examination of the *Praxis II* score trends was not meant to serve as an evaluation of *ESEA* or of any of the other policies instituted at the state or local levels; the study served simply as an effort to examine trends in a systemic outcome that all of these policies have aimed to influence—the content knowledge of prospective teachers. More specifically, this report answers the following questions:

1. Have the demographics or characteristics of individuals who took each test changed across the years included? Total sample size, proportions by gender and race, percent passing, and experience breakdowns are reported.[4]

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores for those who pass differ substantially from those who do not pass as well as from all candidates considered together?

3a. Across years, is there any evidence of a trend in *Praxis* scores for those who pass? How do these trends compare with those who do not pass?

3b. Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

The purpose of this study is to identify trends in *Praxis* scores on a select number of tests across recent years across as many states as possible. While teachers in states using *Praxis* are not fully representative of the national population of teachers, *Praxis* scores provide the most complete available picture of the population of those entering teaching.

While the *Praxis II* tests are used across states, every state creates its own teacher certification testing program. States select the particular *Praxis II* tests they will use from more than 100 test titles that are part of the *Praxis* program. In addition, states set unique passing scores for each test after conducting standard-setting studies and state adoption procedures. Thus, a score that might be considered passing in one state may not be in other states.

[3] The authors of this study are employed by ETS. However, the findings contained in this report do not represent the views of ETS and no organizational endorsement is implied.

[4] We exclude results that have insufficient sample size as defined by cell sizes under 15. For several models we combine the two experience levels (one-to-three years and greater than three years) in order to obtain a sufficient sample. Specifics are reported in the text.

Our earlier studies (Gitomer, 2007; Gitomer, Latham, and Ziomek, 1999) highlighted a number of critical issues that have implications for the design of the proposed study. These issues have import for:

- The number of states and years to be included in the analysis.
- Whether to examine those who pass the tests separately from those who do not.
- Whether to examine those who have teaching experience from those who do not.

1. **States have adopted new licensure tests during the last decade**. The *Praxis* series replaced the National Teacher Examination (NTE) system that was introduced in the 1930s. The NTE tests focused almost exclusively on content and were relatively disconnected from modern theories of teaching and learning. They were based primarily on surveys of teacher education curricula. When the *Praxis* system was instituted in 1993, states began to adopt *Praxis* tests to replace the older NTE tests. ETS, the developer of both *Praxis* and NTE, moved all of the NTE tests under the *Praxis* umbrella during the transition phase. However, the complete transition took longer than expected, so that even in 1997, many states still were using NTE tests.

   The *Praxis* series introduced the concept of tests of content-specific pedagogy in addition to direct measures of content knowledge. These tests include questions about how to teach content, difficulties students might have, and ways to assess students effectively. Many states adopted both content and pedagogy tests as part of licensure requirements. Both types of tests are included in this study.

   In addition to this transition, other states have newly adopted *Praxis* as their certification testing program during recent years as well. The fact that the *Praxis* tests for given certification areas have been adopted over a number of years by various states means that there is no single set of years we can use as a window for comparing trends. This is because the general trend is that with each advancing year from 1997 forward, an increasing number of states are likely to have adopted a particular *Praxis* test. Therefore, for each test examined, there is a tradeoff between number of years available to examine and number of states that can be included. We describe the process we used to optimize these two competing constraints.

2. **The academic profile of those who pass the *Praxis* tests and those who do not varies substantially.** Those who pass have higher SAT scores, higher grades, and are more likely to have stronger content course work (e.g., Gitomer, 2007; Gitomer, Latham, and Ziomek, 1999). This forces the critical question of how to focus the analysis of score trends. We argue that it makes sense to distinguish those who pass from those who do not. There are two important reasons for doing this.

   First and foremost, it is important to know the academic standards that have been met for those who are eligible to teach, not simply for anyone who takes the test. The fact is that those who score low are not eligible to teach and it is also true that there are no prerequisite requirements for taking any *Praxis* exam. For example, there is likely some subset of individuals who, toward the end of their college career, decide that it might be worth a small investment to take the *Praxis* exam, viewing teaching as a possible career option. The critical point is that we want to understand the trajectory in academic achievement of those

who are prospective teachers, defined by their passing certification requirements. It is not clear what the policy implications are of simply looking at those who just take the test.

Second, a number of states have increased requirements by raising the passing score requirements. For example, from 1994 to 2006, seven states raised passing scores for English Language, Literature, and Composition: Content Knowledge (0041), and five states raised passing scores for Mathematics: Content Knowledge (0061). During this period, eight states raised requirements for more than one of the tests included in this study. Thus, precisely because they have raised the bar, any increase in average scores of those who pass may be reflective of explicit changes in policy. And even if some individuals need to take the test multiple times, it is only when they pass (and have higher scores) that they would be considered a prospective teacher. So, the policy changes may, in fact, result in some types of behavior (e.g., additional study) that actually raise the academic quality of the pool.

While the most germane policy question is to examine trends for those who pass *Praxis*, in order to understand the full picture of performance we also report analyses of candidates who fail as well as of all test-takers (combining those who pass with those who do not).

3. **Since the 2002 reauthorization of *ESEA*, there has been a very large increase in the number of teachers taking *Praxis II* who report having had prior teaching experience.** There have always been *Praxis* test-takers who have had prior teaching experience. Some might have been teaching on emergency certification, while others might want an additional certification, and others may have moved from non-*Praxis* states to a state that required *Praxis*. However, likely due to *ESEA* and HQT, the proportion of *Praxis* test-takers who were experienced nearly doubled between 1994–97 and 2002–05 (Gitomer, 2007). This increase is probably due to the large number of teachers who did not meet HQT qualifications and needed to be deemed *highly qualified* in particular subject areas or needed to stop teaching under emergency certification provisions. Gitomer (2007) also reported that the experienced teachers taking *Praxis* had, on average, weaker academic profiles than those newly entering into teaching.

   Thus, it is important to disaggregate our analyses by experience level. If the policy question is whether the profile of new teachers is changing, our analysis must disentangle results by experience level of teachers, particularly within these years of HQT implementation, in which the certification of already practicing teachers is likely to skew the data.

   It is important to recognize that, going forward, there should be a decrease in the number of experienced teachers taking licensure tests. Under *ESEA*, all current teachers should have satisfied the HQT requirements. Thus, the disaggregation by experience level is a necessary step in examining this particular cohort of teachers, during a time when the HQT provisions were having an impact on the *highly qualified* status of experienced teachers.

This study of *Praxis* score trends focus on nine tests as requested by the Department of Education. Detail descriptions of each of the tests can be found on the ets.org Web site: http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=23932d3631df4010VgnVCM10000022f95190RCRD&vgnextchannel=89a25ee3d74f4010VgnVCM10000022f95190RCRD.

The study includes licensure candidates who took the following tests (including *Praxis* test code number) in multiple states:

- Elementary Education: Pedagogy (0012),
- Elementary Education Content Knowledge (0014),
- English Language, Literature, and Composition: Content Knowledge (0041),
- English Language, Literature, and Composition: Pedagogy (0043),
- Mathematics: Content Knowledge (0061),
- Mathematics: Pedagogy (0065),
- Middle School Mathematics: Content Knowledge (0069),
- General Science, Part 1: Content Knowledge (0431), and
- Earth and Space Sciences: Content Knowledge (0571).

II. METHODOLOGY

SELECTING STATES AND YEARS

The initial step in this study required a decision as to which states and years to include in the analysis. Our goal was to be able to make valid comparisons by selecting the states that have a stable number of test-takers throughout the years. We began by determining the number of test-takers by state and year for each of the tests of interest. In Appendix A we identify the year in which the state established a passing standard as well as the first year in which the number of test-takers is relatively stable going forward.

The adoption of tests by states varies widely. States have adopted different tests at different times and frequently applied different passing standards. The study attempted to compare test results across years by including a stable set of states within each analysis and applying appropriate state- and year-specific passing standards for each test. Thus, different combinations of states and years are used in each of the analyses. Because states have been transitioning to new testing requirements, some analyses only include three years of data, while others include up to eight years. Some analyses include six states, while others include 18 states. Trends within individual states are not included; all analyses aggregate test data across a minimum of six states. A number of states, including those with the largest population of teachers (e.g., Calif., N.Y., Texas, Mich., and Fla.), have state-specific licensure tests. Because these tests are unique to particular states, it is not possible to compare performance across these states. Thus, while the sample used in the current study cannot be said to be fully representative of all teacher candidates, it is by far, the most complete sample of teacher candidates available to compare teacher candidates across states.

The sample for each analysis thus includes all individuals who took the particular *Praxis II* assessment in one of the included states during the time period considered. (See Exhibit 1.) For example, the analysis of the Elementary Education: Pedagogy assessment (0012) includes the scores for all individuals who took this test between 2000 and 2005 in the states of Connecticut, the District of Columbia, Georgia, Hawaii, Maryland, Nevada, and South Carolina (seven states).

We selected states and years for the study based on the principle of maximizing both number of states and years included. So, for example in 2000, only three states had adopted the Elementary Education Content Knowledge Test (0014). By 2007, 26 states and localities (including D.C. and Virgin Islands) had adopted the test. Including more years of data decreases the number of states who had participated throughout and thus makes it difficult to make fair comparisons. Therefore, we examined the history of each state's use of *Praxis* tests that appears in Appendix A and arrived at a set of selection decisions (see Exhibit 1). By examining Appendix A, we can use the example of Mathematics: Content Knowledge (0061) to understand the decision process and its implications. If the time period of "2001–06" is selected, two more states, Indiana and South Carolina, could be included. However, because a relatively large number of states are already available for the years 1999–2006, we felt it was advantageous to include the additional two years of data and forego the inclusion of these two states.

The outcome is that different year ranges are used for different tests, and different states are included as well. Within each test though, comparisons are valid in that the same states are involved.

The outcome of this process is that the earliest test year included was 1999. Prior to that year, very few states had made the transition to the *Praxis* tests now in use. This means that the time period explored in this study does not extend as far back as Gitomer (2007) which compared the academic profile of teacher licensure candidates dating back to 1994.

During this time period, testing volumes increased substantially. Further, states did not increase at the same rates. Across states, there were different population shifts, demands for teachers, and outputs of institutions preparing teachers. There is a legitimate question as to how to treat different increases across states in relation to overall trends for *Praxis* tests.

Another legitimate question is how to address different shifts in the population across states. Imagine that State X doubles its output of teacher candidates while State Y has a constant output over a certain number of years. Further, assume that State X candidates have average higher scores than those in State Y. An overall upward trend across states in mean scores might be observed not because the average scores in State X or State Y increased, but because more teacher candidates were being produced out of State X.

One way to approach this issue is to simply ignore these different increases. If a greater number of highly qualified individuals are entering the national teacher pool, it could be argued that it does not matter if these individuals come from different states—the point is that teachers can find employment across state lines and it would be desirable if more candidates came from states that produced more qualified candidates and less desirable if the reverse occurred.

A counterargument would be that licensure is essentially a local process and therefore, we need to understand whether findings are a function of consistent changes across states, or perhaps an artifact of changes in the relative size of states' contributions to the overall sample. In this case, a weighting process would be in order.

In this paper, we examine characteristics of the test-taking population, but also conduct analyses to understand the extent to which there is variation in the proportion of this group coming from particular states over time. Specifically, for each test and range of years, we determine the proportion of our sample that comes from each included state. We then examine the change in the proportion from the first year to the last. For the most part, the relative proportions are stable. Approximately three-quarters of all states included in the analyses do not vary by more than 2 percent in their contribution to the overall sample. Very rarely is the change more than 5 percent. Given these results and the aforementioned argument, we report only data aggregated and unweighted by state.

Exhibit 1
Selection of States and Year Ranges

Praxis Test	Years	Selected States
0012 Elementary Education: Pedagogy	2000–05	CT, DC, GA, HI, MD, NV, SC (7 states)
0014 Elementary Education Content Knowledge	2004–06	AK, AL, CO, ID, KY, LA, MD, MN, MS, NJ, OH, RI, TN, TX, VA, VT, WA, WI (18 states)
0041 English Language, Literature, and Composition: Content Knowledge	1999–2006	AR, CT, DC, GA, HI, IN, KY, LA, MD, MO, NH, NJ, OH, OR, PA, SC, TN, VA, WV (19 states)
0043 English Language, Literature, and Composition: Pedagogy	1999–2006	AR, DC, HI, LA, MD, NV, TN (7 states)
0061 Mathematics: Content Knowledge	1999–2006	AR, CT, DC, GA, HI, KY, MD, MO, MS, NH, NJ, OH, OR, PA, TN, VA, WV (17 states)
0065 Mathematics: Pedagogy	1999–2006	AR, DC, HI, MD, NV, TN (6 states)
0069 Middle School Mathematics: Content Knowledge	2004–06	AK, AL, CT, DE, GA, IN, KY, LA, MD, MO, MN, MS, NC, NH, NJ, OH, OR, PA, SC, TN, VA WA, WV (23 states)
0431 General Science, Part 1: Content Knowledge	1999–2006	DC, HI, NJ, NV, OR, TN (6 states)
0571 Earth and Space Sciences: Content Knowledge	2002–06	AR, CT, IN, MD, MO, NH, NJ, OH, PA, TN, VA, VT (12 states)

Exhibit Reads: The number of states and years included in each analysis vary by test.

DETERMINING PASSING STATUS

Passing scores that define the Pass-Fail status of an individual on a test are established by the individual states through a formal standard-setting process. In this study, we consider passing status only on the basis of individual tests, although recognizing that in many cases, candidates need to pass multiple tests in order to meet state certification requirements.

Appendix B presents the passing scores in effect for each state by year. In a number of cases, states have changed the required passing score, most frequently increasing the score needed to be considered passing. **Bolded cells** indicate cases in which a state raised the passing score of a test and *italicized* cells indicate cases when a state lowered the passing score of a test.

All tests are scaled on a 100–200 point scale. However, there is no way to make any sensible comparison of scores across test titles because the content of the tests and the characteristics of the test-taking samples are different.

Of course, we do not know for certain which state passing standard should be applied to any candidate. An individual may be testing for a license in any number of states. Therefore, we chose to apply the same rules we applied in Gitomer (2007) and Gitomer, Latham, and Ziomek (1999). We made the assumption that the best estimate of where a candidate was seeking licensure was the state in which the most recent test was taken. The rationale for this strategy is based on the finding that teachers are highly likely to wind up teaching very close to where they grew up and went to college. We applied the passing score in effect to create a Pass-Fail status for each test-taker. If the candidate took the test multiple times, the most recent test score value was used, regardless of passing status.

CANDIDATE DATA

Data were taken from ETS *Praxis* data files from 1994–2006. In addition to the test results, data were collected from the Background Information Questionnaire (Appendix C) that the *Praxis* candidates complete as part of the test registration process. We also describe any additional collapsing of data fields that we executed in order to make comparisons that included sufficient sample sizes.

Specific variables of interest include:

- **Gender.**

- **Race and Ethnicity**—While the questionnaire includes ten categories, the only groups that had a sufficient number of candidates to analyze year to year differences were African-American and white candidates. This was true even after collapsing subgroups together (e.g., the three groups whose native language is Spanish). Even for African-Americans, there were a number of tests for which there were insufficient numbers on an annual basis to include in the analysis.

- **Test date—**This field indicates the date on which subjects take their *Praxis II* tests. We defined "academic year" as beginning in Aug. 1 and ending July 31. In this analysis, year reflects the twelve months beginning Aug. 1 (e.g., Jan. 2003 is considered part of the 2002 year).

- **Test Score—**This field indicates the scores the subjects received on the test. For this study, we are including all *Praxis* test-takers who had reportable scores.

- **Passing Status—**Test scores were coded as pass or fail based on the passing score established by a state at the time the test was taken.

- **Test Code**—This field indicates the test codes assigned to the *Praxis II* tests.

- **State—**This field indicates the state in which a *Praxis II* test was taken, which we believe to be the best estimate of the state for which the test-taker is seeking licensure.

- **Teaching experience—**This field indicates the test-takers' teaching experience. In the BIQ, test-takers are asked to identify their teaching status as one of the following:
    1. Planning to enroll or currently enrolled in a teacher education program
    2. Recently graduated and expect to begin teaching in the near future

3. One-to-three years teaching experience
4. More than three years teaching experience
5. Not planning to teach at this time

We collapsed the five teaching experience statuses into three categories. The following responses were collapsed into the *No Experience Category:*

- Not planning to teach at this time
- Planning to enroll or enrolled in teacher education program
- Recent grad and will begin teaching in near future

Separate categories were established for amount of experience so that we could differentiate those who were fairly new to teaching (one-to-three years) to those who were more experienced. Thus the categories were:

- One-to-three years teaching experience
- More than three years teaching experience

While we report experience levels in some analyses, the regression models only consider two discrete conditions—some experience vs. no experience. Sample sizes were often insufficient to include finer distinctions in the model.

III. FINDINGS

Analyses are organized by test. We conducted analyses to answer the following questions:

1. **Have the demographic or performance characteristics of individuals who took each test changed across the years included? Total sample size, proportions by gender and race, percent passing, and experience breakdowns are reported.**[5]

2. **Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?**

3a. **Across years, is there any evidence of a trend in *Praxis* scores for those who pass? How do these trends compare with those who do not pass? How do these trends compare with those who do not pass?**

A key question of the study is whether there are statistically significant non-zero changes in test scores across years. To do this we conducted a series of regression analyses using different models.

The most basic model is a simple linear model in which mean test scores are regressed on Year. This model is simply:

$$y = \beta_0 + \beta x + \varepsilon$$

where y is mean test score and x is testing year.

3b. **Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?**

More complex regression models include the variables of race, gender, and experience. The purpose of this analysis is to determine whether any potential changes in the proportion of individuals in these categories may be accounting for any trends in test scores. These models take the general form of

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

where x_1, x_2, x_3, and x_4 represent year, gender, race, and experience level, respectively.

All results marked with a * are significant at the $\underline{p} < .01$ level. In cases where the significance level is $.01 > \underline{p} < .05$, the value is marked with a + and the exact $\underline{p}$ level is reported.

[5] We exclude results that have insufficient sample size as defined by cell sizes under 15. For several models we combine the two experience levels (one-to-three years and greater than three years) in order to obtain a sufficient sample. Specifics are reported in the text.

Simpson's Paradox

For a number of tests, we present what at first blush appears as a set of conflicting and non-intuitive findings. Specifically, there are a number of cases in which we see positive or neutral trends for individuals who pass the test and similar patterns for those who fail. However, when all candidates are considered together, we find a negative trend. In fact, this curious finding is an instance of a well-known statistical phenomenon known as *Simpson's Paradox* (Blyth, 1972) in which subgroups can each show the same pattern, yet when combined, a different pattern is observed because the relative proportion of individuals in each subgroup changes.

For certain tests, the proportion of those taking the tests and failing increased relative to those who passed. Thus, the increasing proportion of candidates who fail the test can result in the overall means decreasing at the same time that the means for both those who pass and those who fail either increase or remain stable. What we see for a number of tests is that even though scores are stable for both those who pass and those who do not, the proportion of candidates who fail increases across years. That proportional increase leads to a decrease for all candidates even when there is no decrease in subgroups.

A. TEST 0012—ELEMENTARY EDUCATION: PEDAGOGY

Background Information on the Test:

This test consists of four constructed-response prompts around different teaching situations that ask candidates to demonstrate their understanding of curriculum, instruction, and assessment.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

The analysis included tests from 2000 through 2005. Over the six years, gender and race distributions across test-takers are relatively consistent, even as the number of candidates almost doubled. Passing rates for this test are extremely high and consistent across years as well (see Exhibit 2). We do not include 2006 because Georgia discontinued the use of the test after 2005. Because relatively few states use this test, and because a fairly large number of candidates were from Georgia, we opted to include only those years in which Georgia continued to use this test.

Exhibit 2
Test-taker Characteristics by Year—Elementary Education: Pedagogy

	2000	2001	2002	2003	2004	2005
Sample Size	7,732	8,231	8,459	10,600	10,564	13,105
Percent Male	10.0%	9.9%	10.6%	9.7%	10.1%	9.5%
Percent African-American	12.9%	14.0%	13.1%	11.6%	12.2%	13.8%
Percent White	78.9%	77.0%	76.7%	78.5%	78.7%	77.2%
Percent Passed	95.8%	95.2%	95.8%	96.2%	95.1%	97.2%
Percent No Experience	57.0%	54.3%	56.9%	58.6%	59.1%	53.9%
Percent 1–3 Years of Experience	23.1%	25.1%	24.2%	23.4%	20.7%	20.2%
Percent More than 3 Years of Experience	19.8%	20.7%	19.0%	18.0%	20.3%	26.0%

Exhibit Reads: While the number of test-takers increased, demographic characteristics are relatively stable across years while passing rates are very high.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. When analyzing mean scores by gender, passing status, and experience, the two levels of experience are combined due to small sample sizes in particular cells.

Female candidates and white candidates who pass have higher scores than male candidates and African-American candidates, respectively. Individuals without experience appear to have slightly higher scores than those individuals with experience (see Exhibits 3, 4 and 5).

Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 158 (157) and 135 (136), respectively (sd = 10.0). Further, the proportion passing is very high and the mean score for those who pass is substantially higher than the median passing score across states (see Appendix B and Exhibit 6).

Exhibit 3
Mean *Praxis* Scores (SD) of Those Who Pass by Year—Elementary Education: Pedagogy

	2000	2001	2002	2003	2004	2005
All	160.10 (10.0)	158.72 (9.53)	158.09 (8.80)	157.08 (8.68)	155.99 (8.39)	157.43 (8.00)
Female	160.41 (10.03)	159.03 (9.54)	158.37 (8.83)	157.39 (8.69)	156.24 (8.41)	157.69 (8.00)
Male	157.21 (9.20)	155.64 (8.79)	155.64 (8.09)	154.12 (7.90)	153.79 (7.86)	154.74 (7.52)
African-American	154.28 (9.09)	153.65 (9.04)	153.39 (8.26)	153.33 (7.99)	152.85 (7.45)	153.96 (7.51)
White	161.09 (9.82)	159.74 (9.25)	159.02 (8.59)	157.81 (8.63)	156.58 (8.39)	158.08 (7.87)
No Experience	160.81 (9.61)	159.55 (9.22)	159.00 (8.40)	157.52 (8.58)	156.22 (8.34)	157.89 (7.82)
1–3 Years of Experience	159.34 (10.33)	158.08 (9.75)	156.97 (9.02)	156.69 (8.68)	155.74 (8.31)	157.14 (7.98)
More than 3 Years of Experience	158.97 (10.51)	157.15 (9.72)	156.85 (9.43)	156.03 (8.84)	155.48 (8.55)	156.51 (8.22)

Exhibit Reads: There is a decline in mean scores across years as well as consistent gender, race and ethnicity differences

Exhibit 4
Mean *Praxis* Scores of Test-takers by Gender, Passing Status and Teaching Experience—Elementary Education: Pedagogy

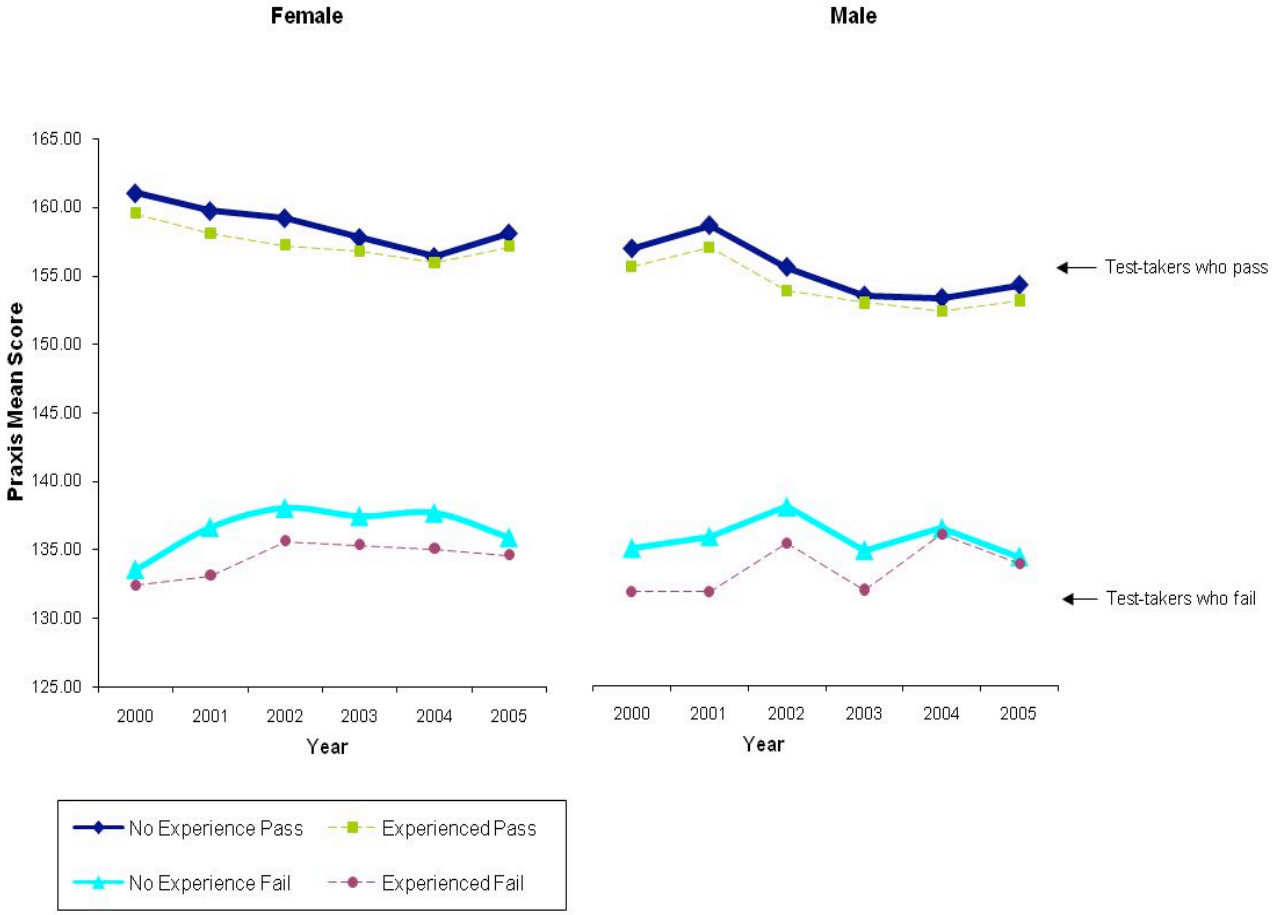


Exhibit Reads: For those who pass, scores for female candidate are higher than for male candidates. Mean scores for those who pass are much higher than for those who do not.

Exhibit 5
Mean *Praxis* Scores of Test-takers by Race and Ethnicity, Passing Status and Teaching Experience—Elementary Education: Pedagogy

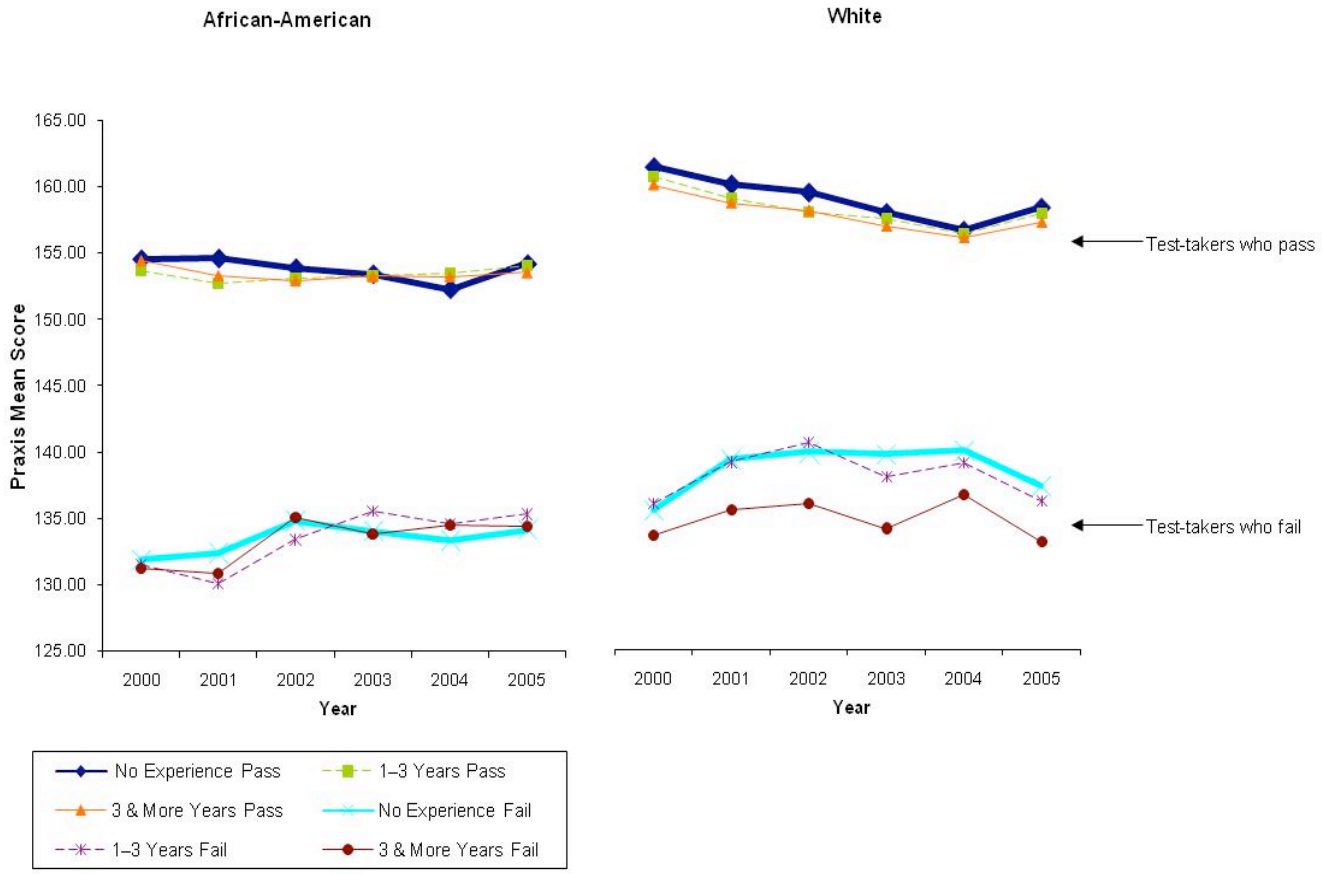


Exhibit Reads: Scores for white candidates are higher than for African-American candidates. Mean scores for those who pass are much higher than for those who do not.

Exhibit 6
Distribution of Scores in 2005 for Elementary Education: Pedagogy

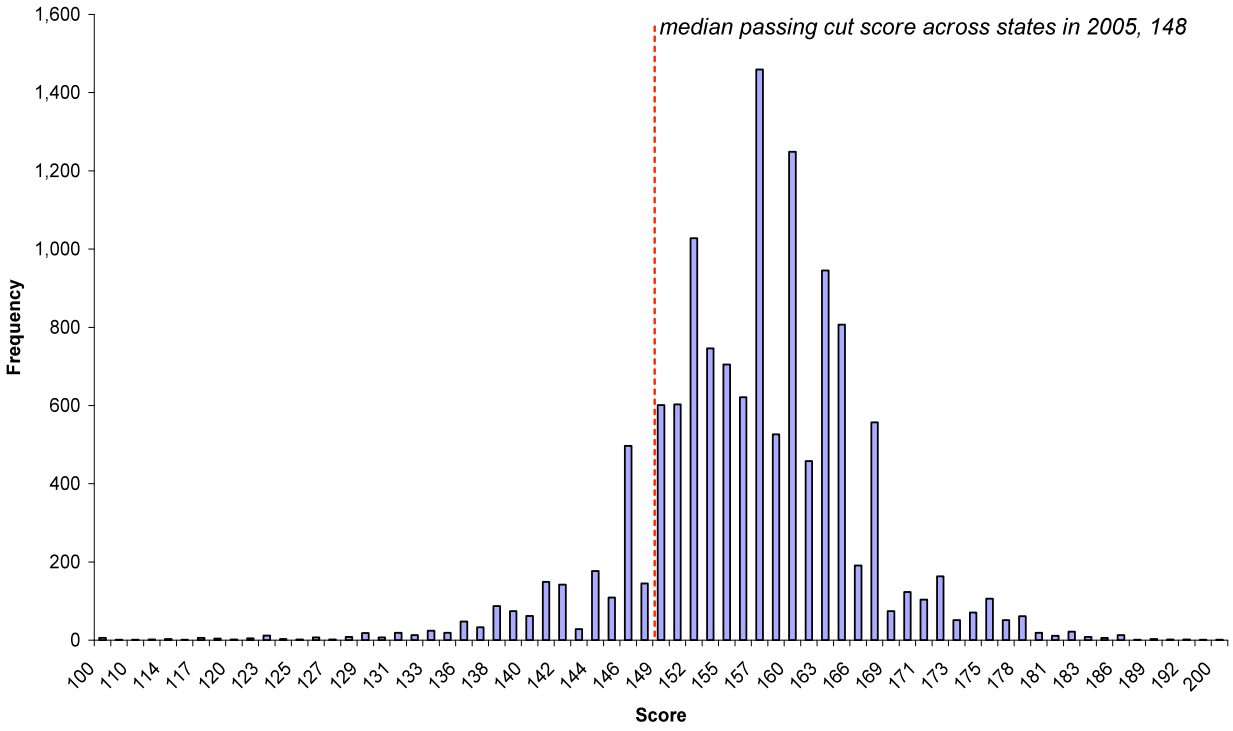


Exhibit Reads: The majority of candidates pass the test with scores well above the median passing score across states.

Mean passing scores for all test-takers are disaggregated by experience. This analysis reinforces the large difference in scores between those who pass and those who fail. Additionally, there appears to be a small decrease in scores across years as well as a generally higher mean score for those with no experience. The fact that the trends for all candidates are only slightly lower than for passers is due to the very high passing rate for this test (see Exhibit 7).

Exhibit 7
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—Elementary Education: Pedagogy

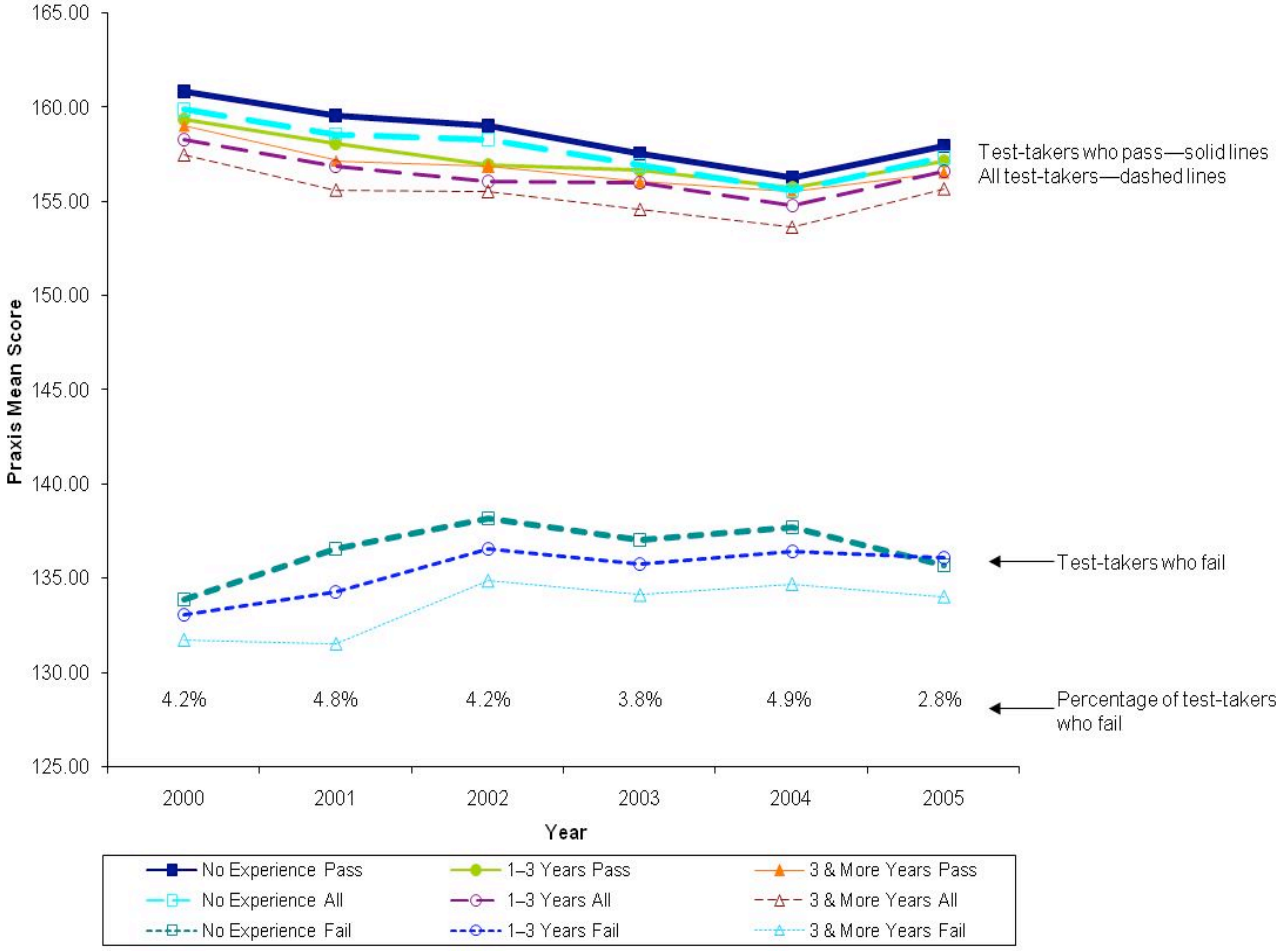


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Candidates with no teaching experience score higher than those with experience.

3. Across years, is there any evidence of a trend in *Praxis* scores for those who pass? How do these trends compare with those who do not pass?

Trends are analyzed using two regression models. The simple model examines mean score changes over year. The comprehensive model includes gender, race, and experience variables. In the comprehensive model we consider only two levels of experience—experience vs. no experience. For each of these models, we first consider only candidates who pass and then all candidates.

The analysis confirms that there was a small, but significant decline in scores between 2000 and 2005 for the seven states included in the analysis. Additionally, there were statistically significant race, gender, and experience differences (see Exhibit 8). The pattern across years is stable across simple and more comprehensive models.

As can be seen in the simple model, the proportion of variance accounted for by testing year is very small. Statistical significance is achieved due to the very large sample of candidates.

Exhibit 8
Regression Results—Elementary Education: Pedagogy

Variable	Std Wts.— Simple Model Passers	Std Wts.— Simple Model All	Std Wts.— Comprehensive Model Passers	Std Wts.— Comprehensive Model All
Test Year	-.112*	-.084*	-.118*	-.093*
Gender			.087*	.101*
Race and Ethnicity			.179*	.244*
Experience			-.04*	-.042*
Adjusted R Squared	.013*	.007*	.058*	.085*

Exhibit Reads: There is a small decline in scores across testing years for those who pass as well as for all candidates.

B. TEST 0014—ELEMENTARY EDUCATION CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 120 multiple-choice items, evenly distributed around understanding of content important for elementary education in reading and language arts, mathematics, social studies, and science.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Data from 18 states are included. However, because the adoption of this test was relatively recent, only three years of data are analyzed (see Exhibit 9). Across the three years, there are no obvious demographic trends and passing rates hover around 90 percent while there is also an increase in the absolute number of test-takers. There appears to be a small increase in the proportion of African-Americans taking the tests, but additional years of data would be needed to determine if this is a stable trend.

Exhibit 9
Test-taker Characteristics—Elementary Education Content Knowledge

	2004	2005	2006
Sample Size	28,199	34,395	33,979
Percent Male	11.3%	11.2%	11.1%
Percent African-American	9.7%	9.9%	11.3%
Percent White	81.6%	81.5%	79.3%
Percent Passed	91.3%	92.3%	89.6%
Percent No Experience	73.0%	73.7%	75.3%
Percent 1–3 Years of Experience	16.4%	15.4%	13.8%
Percent More than 3 Years of Experience	10.5%	11.0%	11.0%

Exhibit Reads: Demographic characteristics are relatively stable while passing rates are high, although it is difficult to detect trends with only three years of data.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Male candidates who pass 0014 appear to have somewhat higher passing scores than female candidates who pass. There are no apparent or consistent patterns associated with experience. White candidates appear to have higher passing scores than African-American candidates (see exhibits 10, 11 and 12).

Again, the variation in mean scores between those who pass and those who do not is very large, approximately two standard deviations. The mean (and median) scores for those who pass and those who fail are 166 (165) and 132 (133), respectively (sd = 16.8). Even though the proportion of those passing is very high, it appears that many who do not pass have markedly less content knowledge as measured by the test and would be unlikely to pass the test on another attempt simply by chance. The standard error of measurement (SEM) for this test is 5.9.[6] More than three-quarters of candidates (76 percent) who do not pass the assessment score at least one SEM (a score of 138) below the median passing score. Thus, their scores are well below the state standards that have been established (see Exhibit 13**).** Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 10
Mean *Praxis* Scores (SD) of Those Who Pass by Year—
Elementary Education Content Knowledge

	2004	2005	2006
All	165.76 (13.82)	165.74 (13.97)	165.59 (14.15)
Female	165.08 (13.60)	165.04 (13.71)	164.86 (13.88)
Male	170.35 (14.49)	170.55 (14.74)	170.64 (14.94)
African-American	156.56 (10.99)	155.69 (11.30)	154.64 (11.57)
White	166.51 (13.68)	166.82 (13.73)	166.88 (13.87)
No Experience	165.69 (13.61)	165.71 (13.70)	165.61 (13.93)
1–3 Years of Experience	165.43 (14.04)	165.43 (14.16)	165.16 (14.47)
More than 3 Years of Experience	165.81 (14.76)	166.61 (15.19)	166.15 (15.29)

Exhibit Reads: Scores appear to be stable across testing years. There are consistent gender, race and ethnicity differences

[6] All reported Standard Errors of Measurement (SEM) data are based on data from 2005. These measures are very stable across years, so we chose to base the analysis on a single year's estimate.

Exhibit 11
Mean *Praxis* Scores of Test-takers by Gender, Passing Status and Teaching Experience—Elementary Education: Content Knowledge

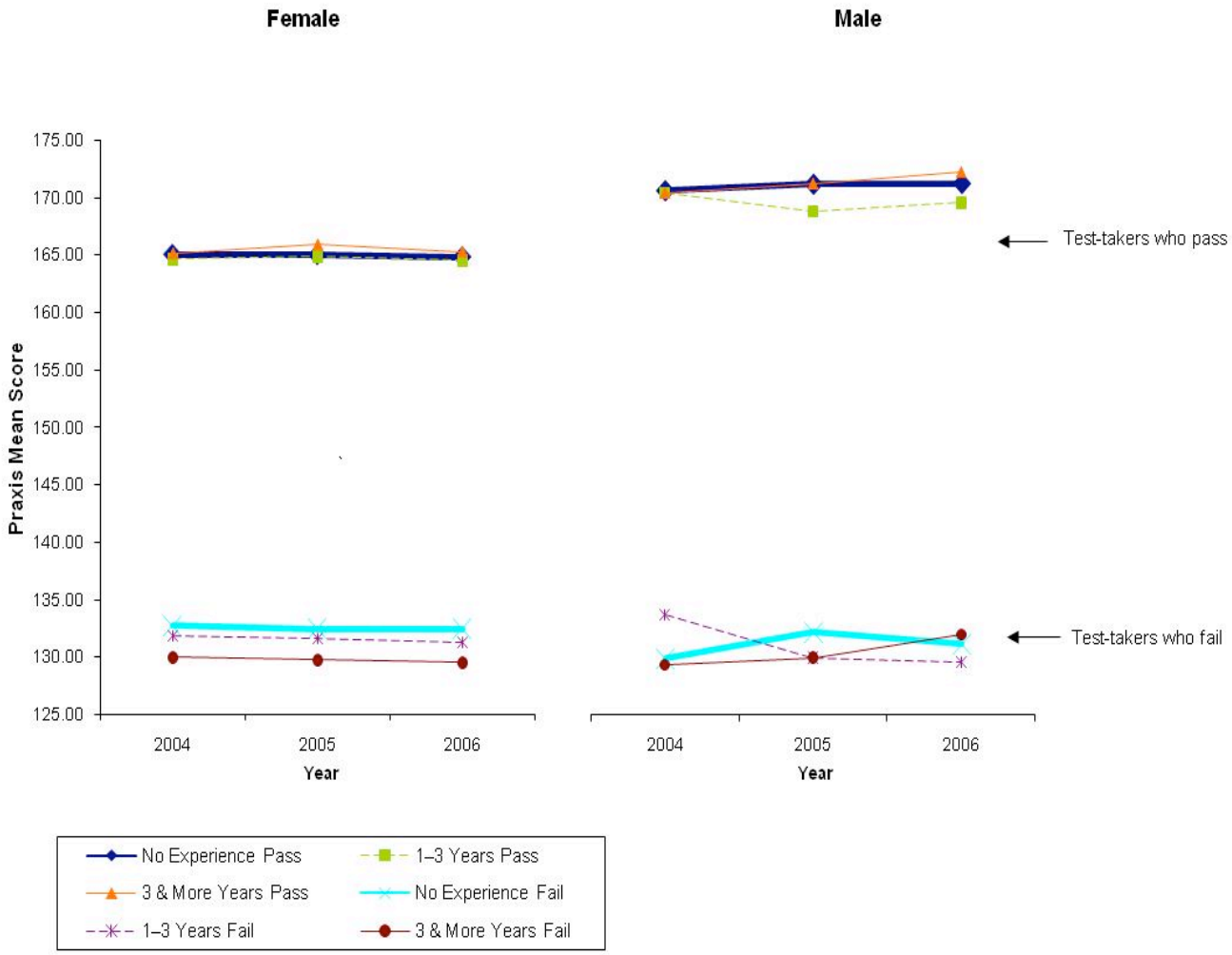


Exhibit Reads: For those who pass, scores for male candidate are higher than for female candidates.

Exhibit 12
Mean *Praxis* Scores of Test-takers by Race, Ethnicity, Passing Status and Teaching Experience—Elementary Education: Content Knowledge

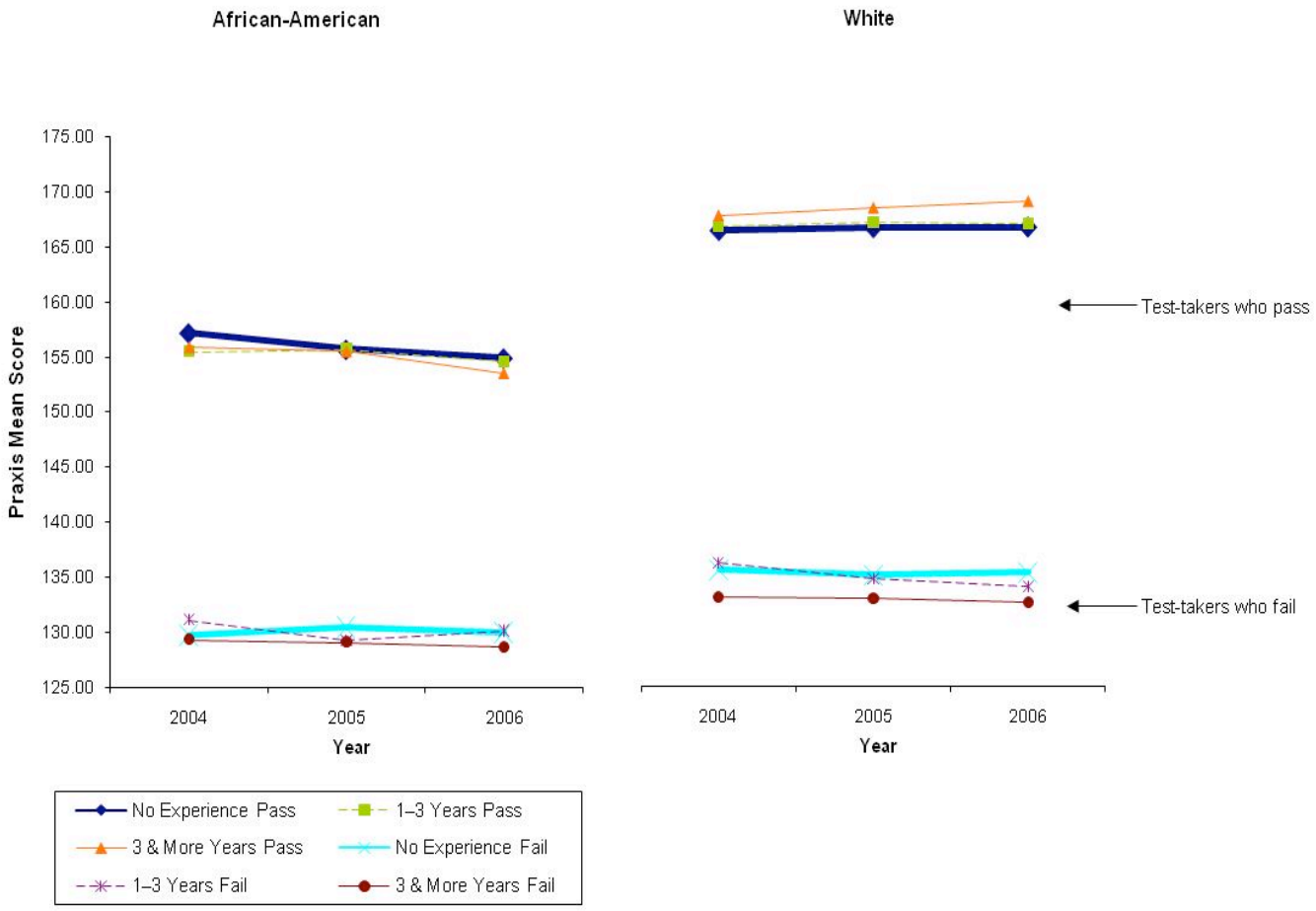


Exhibit Reads: Scores for white candidates are higher than for African-American candidates

Mean passing scores for all test-takers are disaggregated by experience. The analysis reinforces the large difference in scores between those who pass and those who fail, with scores for those who pass well above the median passing standards for states. For those who fail, scores are well below the median passing standards for states (see Appendix B). Interestingly, while scores for those who pass, regardless of experience, seem to be very stable, there does seem to be a slight downward trend when all candidates are included. For passers, those who have the most experience appear to have slightly higher mean scores. For those who fail, those with the most experience also appear to have the lowest scores (see Exhibit 14).

Exhibit 13
Distribution of Scores in 2006 for Elementary Education: Content Knowledge

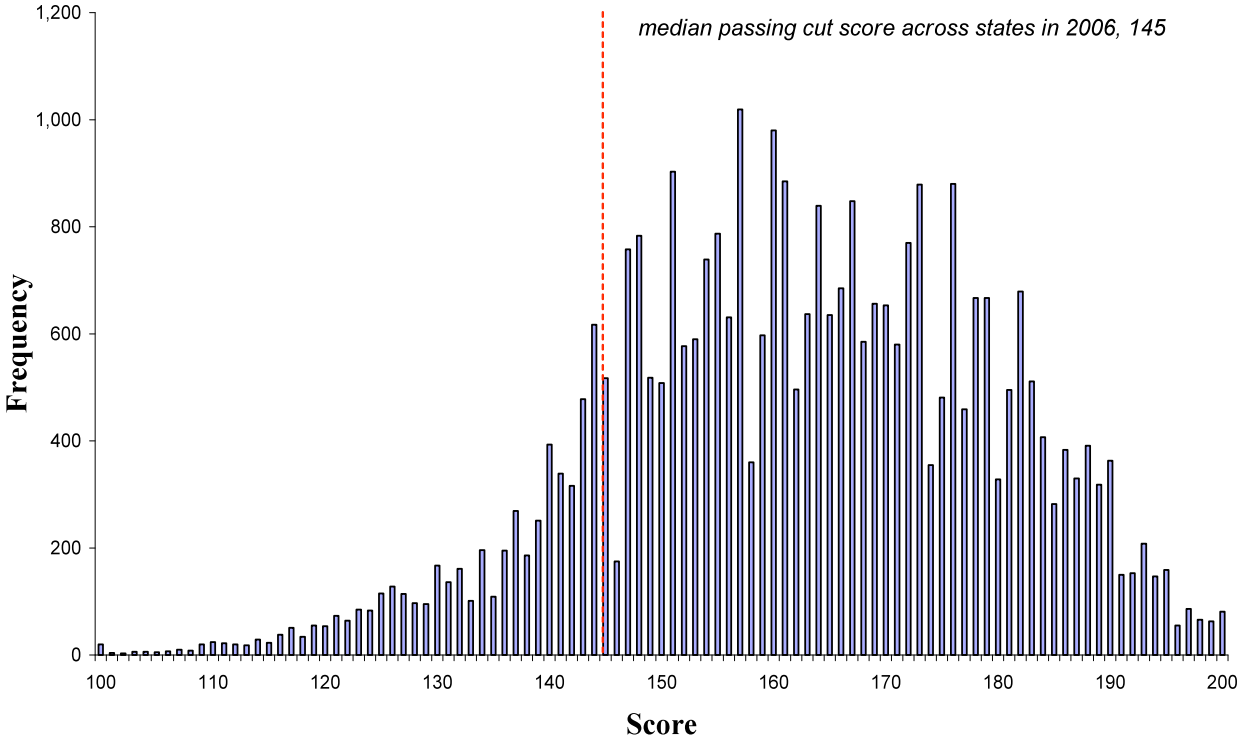


Exhibit Reads: Of the small proportion of test-takers who do not pass, many have scores substantially below state passing standards.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

In the regression analyses, mean scores do not vary much across years for those who pass. However, there is a very slight decline when all candidates are considered. Though the trend is discernible, it is extremely small and explains a very small portion of the score variation.

The comprehensive analysis confirms that male candidates who pass have slightly higher scores than females and white candidates who pass score higher than African-American candidates. There is also a main effect of experience with those having the most experience also having the highest scores.

A slightly different picture emerges when all candidates are considered. Race and ethnicity effects are larger and the experience effect is reversed due to the fact that, for those who fail, candidates with the most experience have the lowest scores (see Exhibit 15). The trend effect remains non-significant.

Exhibit 14
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—
Elementary Education: Content Knowledge

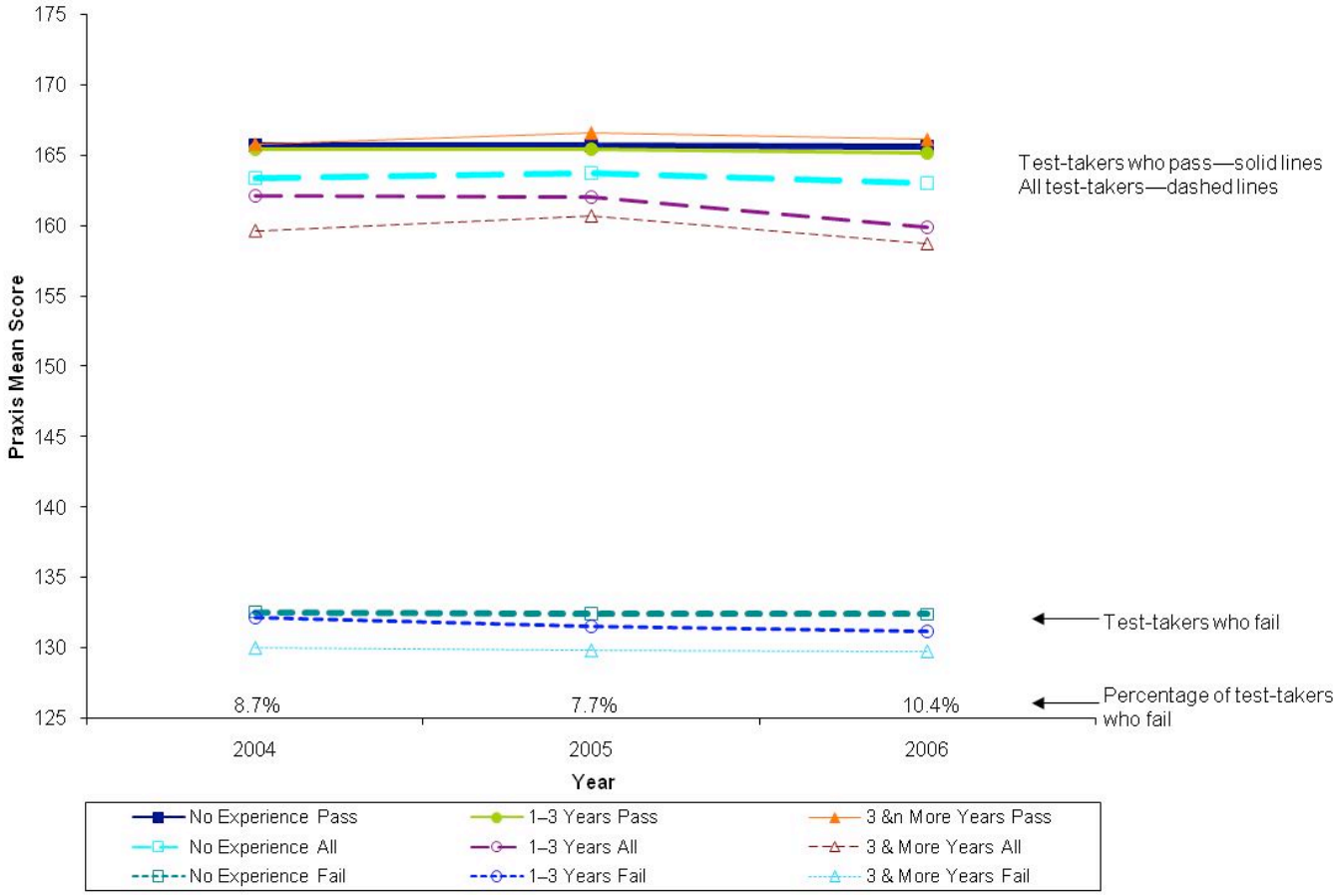


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Experience differences vary by passing status.

Exhibit 15
Regression Results—Elementary Education: Content Knowledge

Variable	Std Wts.— Simple Model Passers	Std Wts.— Simple Model All	Std Wts.— Comprehensive Model Passers	Std Wts.— Comprehensive Model All
Test Year	-.005 ns	-.02*	.006 ns	-.006 ns
Gender			-.130*	-.117*
Race and Ethnicity			.224*	.374*
Experience			.022*	-.013*
Adjusted R Squared	.000 ns	.000*	.065*	.152*

Exhibit Reads: There is no change in scores across testing years for those who pass. There is a slight decline in scores for the simple model when all candidates are included

For the simple model, the finding of different trend lines for passers as compared to all candidates, albeit very small, has import for a number of analyses of other tests that will follow. There are two possible reasons that there these trends are different. One reason would be that there is a decline for those who fail so that, on average, there is a mean decline over time. However, as supported in a separate analysis of just those who fail, such is not the case (see also Exhibit 14). Those who fail also have scores that remain consistent across the years.

In fact, this curious finding is an instance of a well-known statistical phenomenon known as Simpson's Paradox. What we see for the Elementary Education: Content Knowledge test is that even though scores are stable for both those who pass and those who do not, the proportion of candidates who fail increases across years. That proportional increase leads to a decrease for all candidates while there is no decrease in subgroups.

C. TEST 0041—ENGLISH LANGUAGE, LITERATURE, AND COMPOSITION: CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 120 multiple-choice items, with the heaviest emphasis on literature and understanding text. Other major areas are composition and rhetoric and language and linguistics.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Data from 19 states are included. Across the eight years included in the analyses, the proportion of African-American candidates increases a small amount and the percent passing declines about 4 percentage points (see Exhibit 16). The number of test-takers almost doubles during this time period.

Exhibit 16
Test-taker Characteristics—English Language, Literature, and Composition: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
Sample Size	5,109	5,673	6,807	7,776	8,360	8,961	9,278	9,179
Percent Male	24.2%	24.8%	25.1%	24.8%	23.0%	23.7%	23.5%	24.0%
Percent African-American	8.2%	8.3%	8.5%	9.3%	8.5%	9.5%	9.3%	9.9%
Percent White	85.8%	85.0%	84.4%	83.0%	84.5%	82.9%	83.5%	82.1%
Percent Passed	90.6%	89.4%	91.1%	90.7%	88.6%	87.9%	86.6%	86.1%
Percent No Experience	68.7%	68.6%	66.3%	65.2%	66.2%	63.8%	64.4%	67.4%
Percent 1–3 Years of Experience	16.2%	17.0%	18.7%	19.3%	19.1%	19.2%	19.0%	18.1%
Percent More than 3 Years of Experience	15.1%	14.5%	15.0%	15.5%	14.7%	17.0%	16.6%	14.5%

Exhibit Reads: Demographic characteristics are relatively stable across years while passing rates show a decline. There appears to be a modest increase in the proportion of African-American test-takers.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Male candidates who pass 0041 appear to have somewhat higher passing scores than female candidates who pass. White candidates have higher passing scores than African-American candidates (see Exhibits 17, 18 and 19).

Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 180 (180) and 151 (152), respectively (sd = 14.7). The standard error of measurement (SEM) for this test is 5. Almost two-thirds of candidates (64 percent) who do not pass the assessment score at least one SEM below (a score of 155) the median passing score. Thus, their scores are well below the state standards that have been established (see Exhibit 20**).** Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 17
Mean *Praxis* Scores (SD) of Those Who Pass by Year—English Language, Literature, and Composition: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
All	179.53 (11.60)	179.63 (11.67)	180.44 (11.49)	180.33 (11.66)	180.40 (11.09)	180.41 (11.13)	179.82 (10.90)	179.95 (11.31)
Female	179.13 (11.58)	179.35 (11.66)	180.05 (11.41)	179.93 (11.63)	180.03 (11.10)	179.87 (11.04)	179.32 (10.82)	179.41 (11.25)
Male	180.63 (11.54)	180.50 (11.69)	181.57 (11.62)	181.54 (11.64)	181.63 (10.98))	181.87 (11.25)	181.03 (11.01)	181.42 (11.37)
African-American	171.21 (9.70)	171.63 (10.69)	172.43 (10.32)	170.90 (9.74)	173.04 (9.45)	171.82 (8.99)	171.90 (9.08)	171.14 (9.21)
White	179.99 (11.54)	180.03 (11.52)	180.96 (11.32)	181.00 (11.50)	180.86 (11.07)	180.86 (11.02)	180.15 (10.83)	180.59 (11.19)
No Experience	178.95 (11.37)	178.95 (11.21)	180.16 (11.29)	179.82 (11.47)	180.29 (10.90)	180.25 (10.95)	179.47 (10.79)	180.09 (11.10)
1–3 Years of Experience	178.82 (11.55)	179.35 (12.07)	179.55 (11.34)	180.26 (11.68)	179.65 (11.07)	179.47 (11.13)	178.93 (10.82)	178.89 (11.66)
More than 3 Years of Experience	182.26 (12.24)	182.62 (12.56)	182.50 (12.28)	182.16 (12.33)	181.61 (12.03)	181.14 (11.92)	180.67 (11.51)	179.79 (12.11)

Exhibit Reads: Scores appear to be stable across testing years, but there are consistent gender, race and ethnicity differences

Exhibit 18
Mean *Praxis* Scores of Test-takers by Gender, Passing Status and Teaching Experience—English Language, Literature, and Composition: Content Knowledge

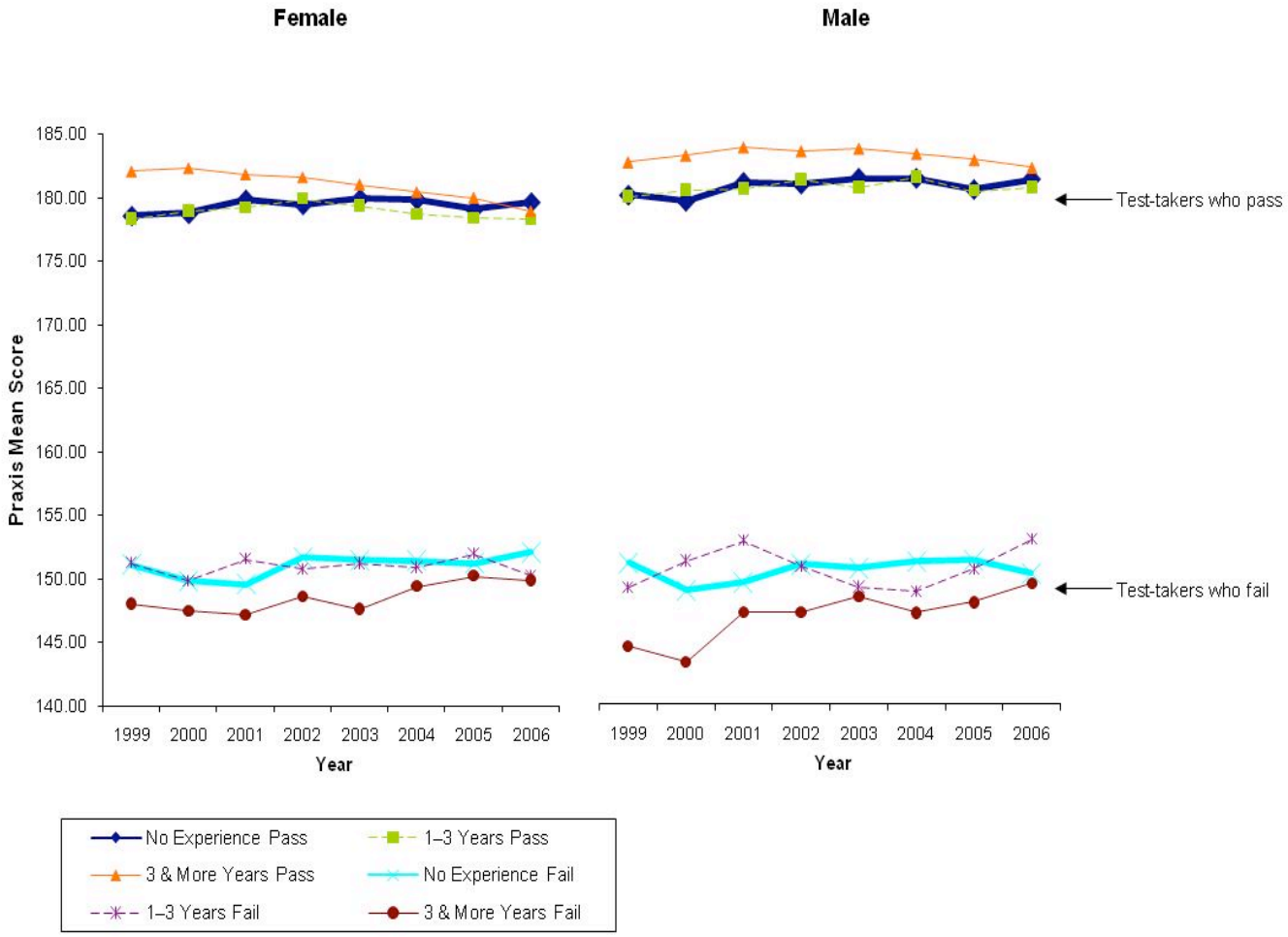


Exhibit Reads: For those who pass, scores for male candidates are higher than for female candidates.

Exhibit 19
Mean *Praxis* Scores of Test-takers by Race, Ethnicity, Passing Status and Teaching Experience—English Language, Literature, and Composition: Content Knowledge

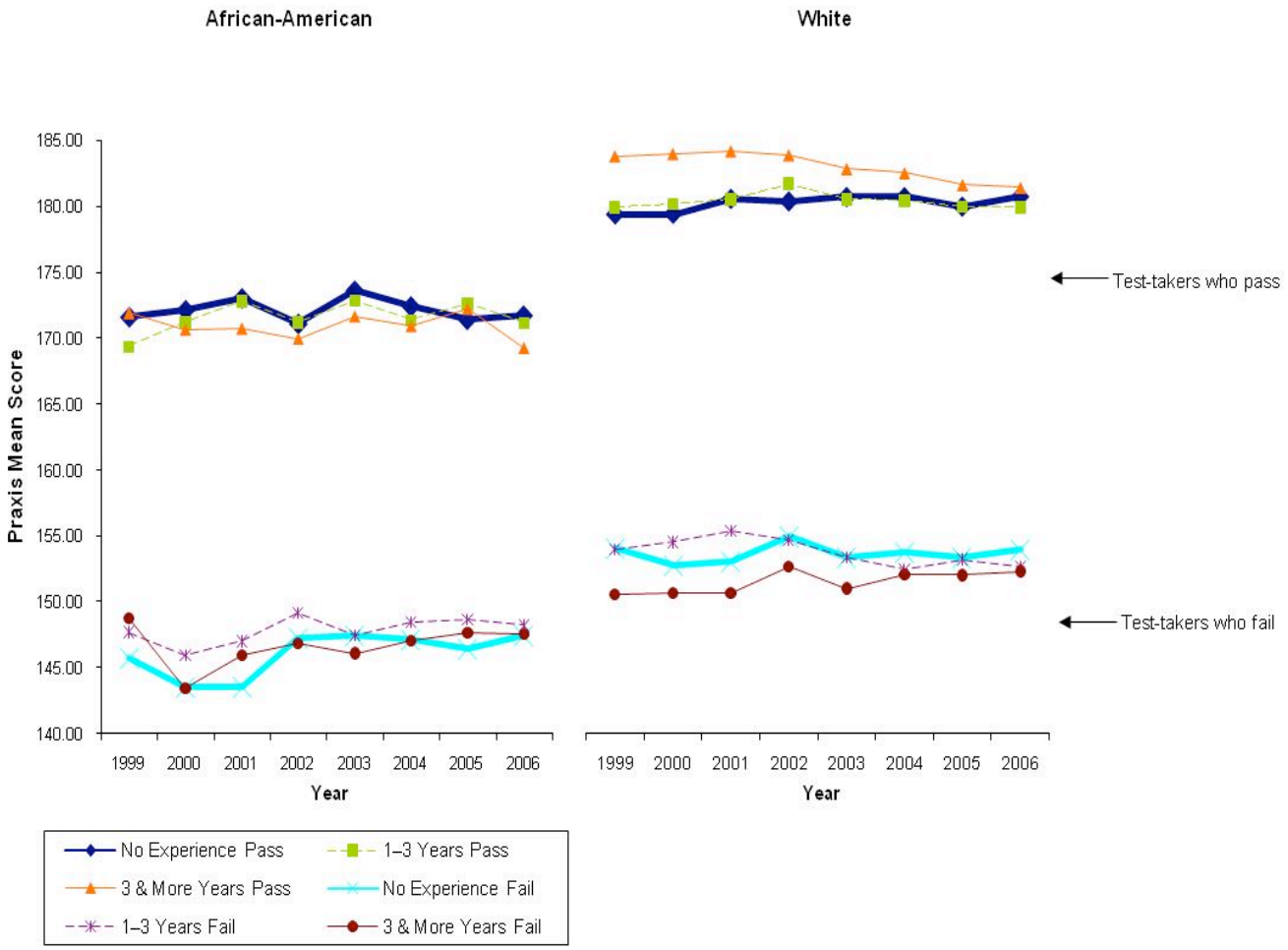


Exhibit Reads: Scores for white candidates are higher than for African-American candidates

Exhibit 20
Distribution of Scores in 2006 for English Language, Literature, and Composition: Content Knowledge

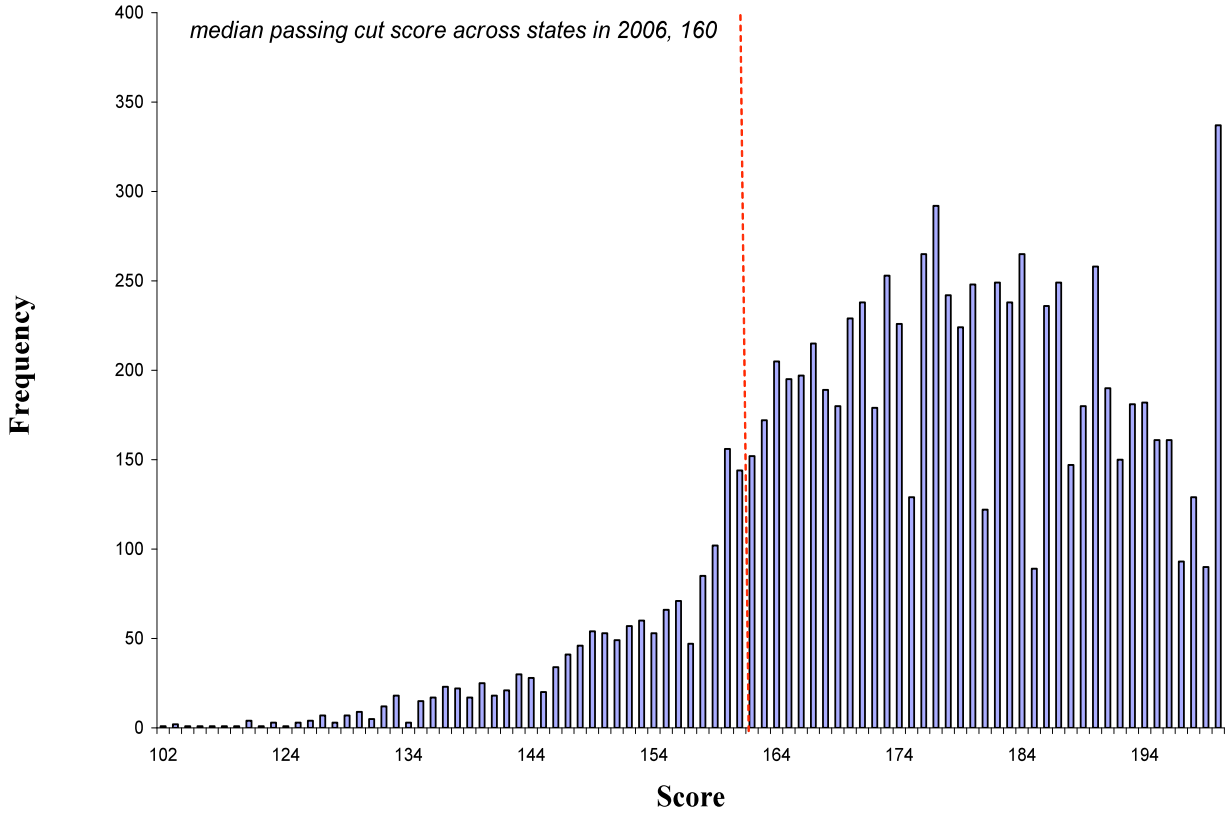


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Mean passing scores for all test-takers are disaggregated by experience. The analysis reinforces the large difference in scores between those who pass and those who fail. As with test 0014, while scores for those who pass, regardless of experience, seem to be very stable, there does seem to be a slight downward trend when all candidates are included. For passers, those who have the most experience appear to have slightly higher mean scores. For those who fail, those with the most experience also appear to have the lowest scores (see Exhibit 21).

Exhibit 21
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—English Language, Literature, and Composition: Content Knowledge

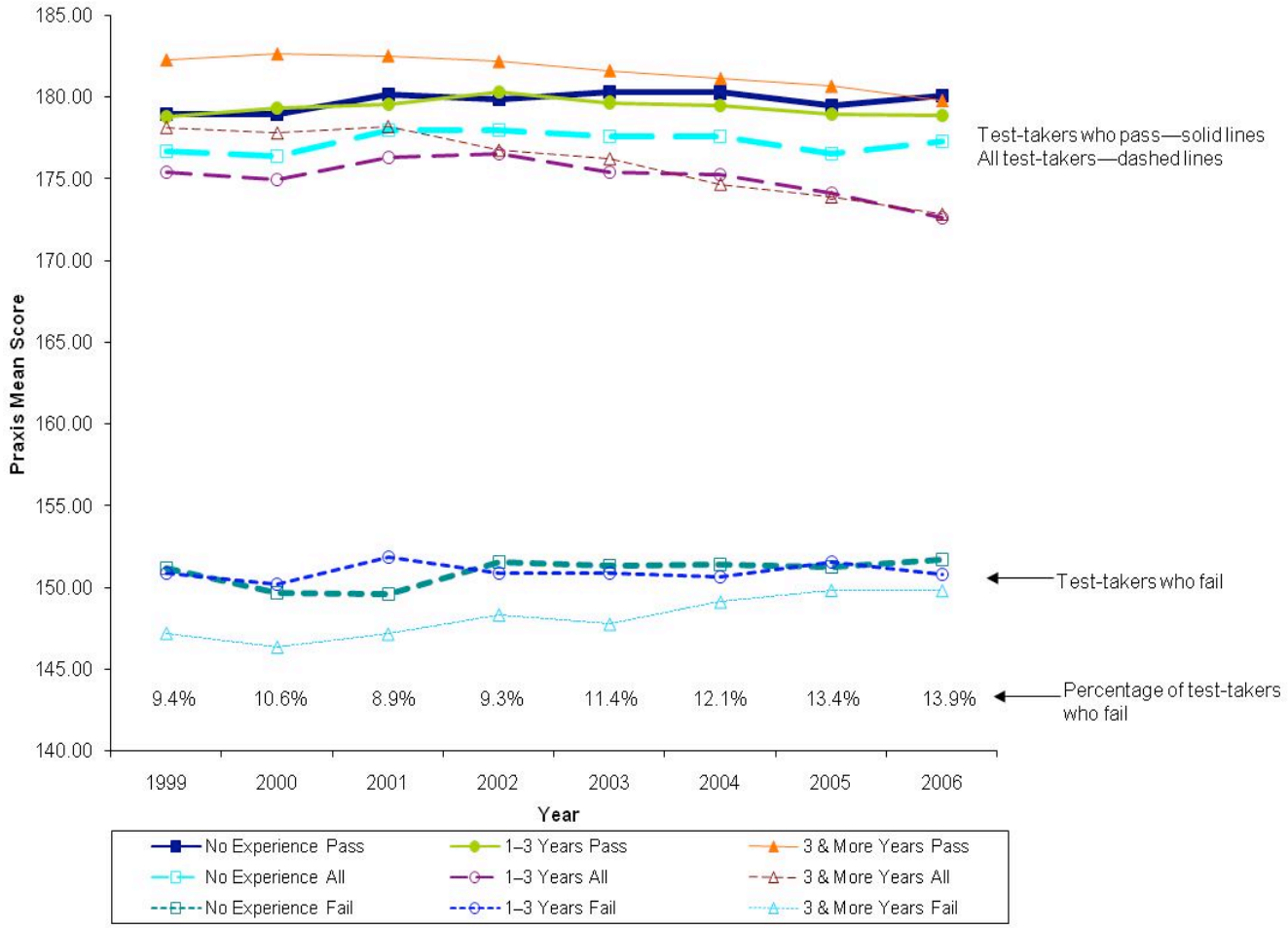


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Experience differences vary by passing status. There is a declining trend across testing years for all candidates.

3. Across years, is there any evidence of a trend in *Praxis* scores for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

In the regression analyses mean scores do not vary across years for those who pass. However, there is a slight decline when all candidates are considered. However, though the trend is discernible, it is extremely small and explains a very small portion of the score variation.

The comprehensive analysis confirms that male candidates who pass have slightly higher scores than females and white candidates who pass score higher than African-American candidates. Those having the most experience also had generally higher scores.

A slightly different picture emerges when all candidates are considered in the comprehensive analysis. Again, there is a small, but significant decline across years. Gender, race and ethnicity effects are larger and the experience effect disappears due to the fact that for those who fail, candidates with the most experience have the lowest scores (see Exhibit 22).

As with Elementary Education: Content Knowledge, a different interpretation is suggested depending on whether the analysis includes all candidates or just those who pass. While there is no discernible trend for those who pass, the decline for all candidates seems to again be an instance of Simpson's paradox. The decrease for all candidates is caused by the increased proportion of students who fail.

Exhibit 22
Regression Results—English Language, Literature, and Composition: Content Knowledge

Variable	Std Wts. –Simple Model Passers	Std Wts. –Simple Model All	Std Wts. – Comprehensive Model Passers	Std Wts. – Comprehensive Model All
Test Year	.004 ns	-.027*	.004 ns	-.024*
Gender			-.054*	-.042*
Race and Ethnicity			.189*	.355*
Experience			.044*	-.005 ns
Adjusted R Squared	.000 ns	.001*	.040*	.130*

Exhibit Reads: There is no change in scores across testing years for those who pass. There is a decline when all candidates are included.

D. TEST 0043—ENGLISH LANGUAGE, LITERATURE, AND COMPOSITION: PEDAGOGY

Background Information on the Test:

This test consists of two constructed response questions, one focused on teaching literature and the other responding to student writing.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Across eight years, data from seven states are included. Compared with the English test of content (0041), far fewer candidates take this test, although the number of test-takers increases substantially over this time period. No obvious trends in the demographics of test-takers are apparent (see Exhibit 23) other than a small increase in the proportion of males taking the test.

Exhibit 23
Test-taker Characteristics by Year—English Language, Literature, and Composition: Pedagogy

	1999	2000	2001	2002	2003	2004	2005	2006
Sample Size	836	982	1,064	1,283	1,310	1,480	1,399	1,453
Percent Male	19.5%	22.8%	19.0%	19.5%	20.2%	22.3%	21.9%	23.2%
Percent African-American	15.2%	14.5%	12.3%	14.7%	13.2%	15.1%	13.1%	14.4%
Percent White	75.9%	77.5%	78.6%	75.8%	76.0%	75.7%	77.3%	75.3%
Percent Passed	89.7%	88.2%	87.0%	88.4%	92.5%	90.7%	89.9%	87.1%
Percent No Experience	55.0%	54.9%	54.3%	55.0%	56.8%	54.7%	54.5%	53.6%
Percent 1–3 Years of Experience	25.5%	28.7%	25.2%	26.6%	26.0%	26.8%	26.1%	30.6%
Percent More than 3 Years of Experience	19.5%	16.4%	20.6%	18.4%	17.2%	18.6%	19.4%	15.8%

Exhibit Reads: Demographic characteristics are relatively stable across years, although there appears to be a modest increase in the proportion of male test-takers.

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Because of the small number of individuals in particular cells if data are

disaggregated by experience, gender, race and ethnicity, more detailed figures by gender, race and ethnicity are not presented.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Once again, there are large difference in scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 158 (155) and 135 (135), respectively (sd = 13.7). Scores for those who pass are well above the median passing standards for states. For those who fail, scores are well below the median passing standards for states (see Appendix B and Exhibit 25).

Mean passing scores for all test-takers are disaggregated by experience. Plots are shown for those who pass, those who fail, and all candidates, regardless of passing status. The percent of those who fail is presented at the bottom of the exhibit, increasing across test years. There is more year-to-year variation in scores, likely attributable to the smaller sample size, but there is no obvious trend for those who pass. On the other hand, scores for those who fail seem to improve across years. Of note, and in contrast to several of the tests already presented, there is no clear trend in passing rates across testing years (see Exhibit 26).

Exhibit 24
Mean *Praxis* Scores (SD) of Those Who Pass by Year—English Language, Literature, and Composition: Pedagogy

	1999	2000	2001	2002	2003	2004	2005	2006
All	158.15 (12.75)	156.86 (11.64)	157.08 (12.24)	156.88 (12.77)	157.82 (11.74)	158.42 (11.51)	156.88 (11.14)	157.64 (11.13)
Female	157.97 (12.94)	156.54 (11.70)	157.24 (12.33)	157.26 (12.93)	158.20 (11.79)	158.49 (11.63)	157.06 (11.31)	157.65 (11.15)
Male	158.90 (11.89)	158.09 (11.41)	156.39 (11.84)	155.25 (11.97)	156.24 (11.43)	157.88 (11.07)	155.74 (10.40)	157.85 (11.33)
African-American	153.76 (12.64)	152.55 (10.98)	153.14 (12.62)	152.34 (13.36)	153.63 (10.49)	154.17 (11.80)	153.22 (12.29)	154.44 (12.03)
White	158.41 (12.46)	157.36 (11.66)	157.57 (12.11)	157.81 (12.41)	158.42 (11.84)	159.03 (11.41)	157.52 (11.00)	158.12 (11.01)
No Experience	157.73 (12.66)	156.52 (11.29)	156.63 (12.07)	156.71 (12.98)	158.43 (11.62)	158.82 (11.62)	156.79 (11.18)	156.72 (11.01)
1–3 Years of Experience	159.12 (12.77)	156.95 (12.50)	158.98 (12.76)	157.57 (12.41)	156.35 (11.37)	158.30 (10.95)	157.14 (11.14)	159.01 (11.21)
More than 3 Years of Experience	157.17 (12.71)	157.63 (11.29)	156.44 (12.04)	155.75 (12.38)	157.97 (12.56)	157.37 (12.00)	157.21 (11.28)	158.10 (11.26)

Exhibit Reads: Scores appear to be stable across testing years. There are consistent race and ethnicity differences.

Exhibit 25
Distribution of Scores in 2006 for English Language, Literature, and Composition: Pedagogy

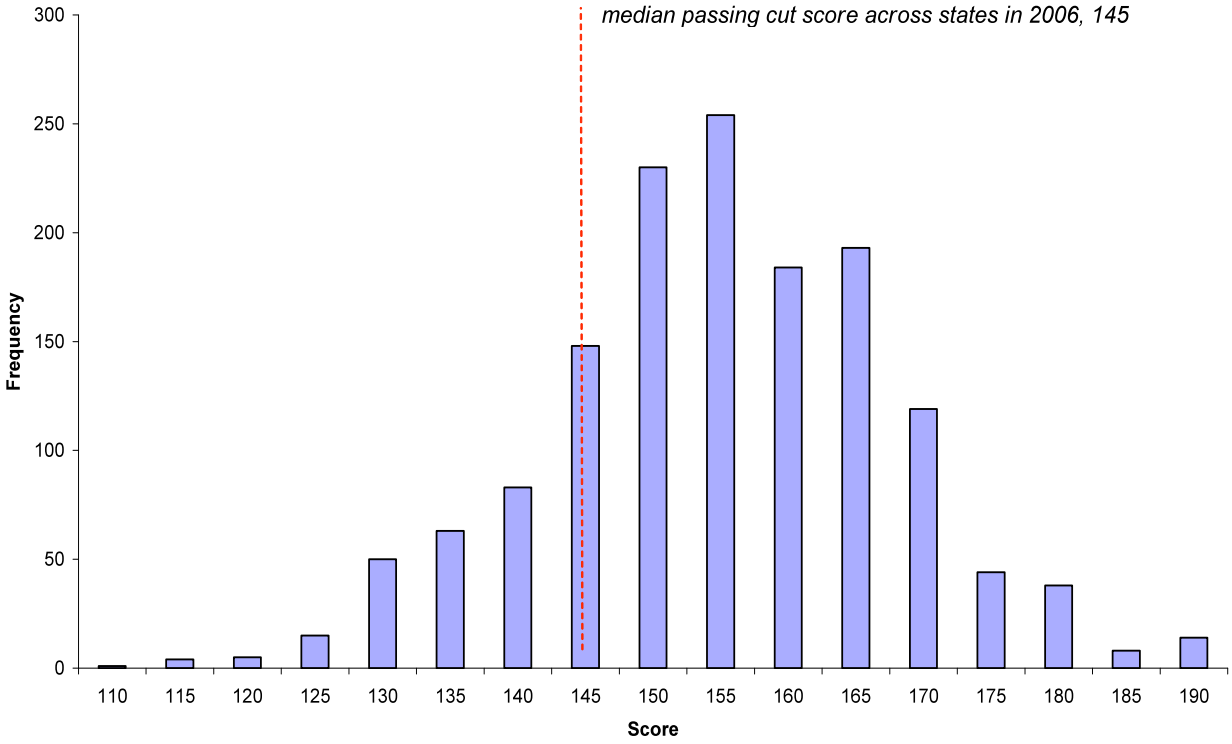


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Exhibit 26
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—English Language, Literature, and Composition: Pedagogy

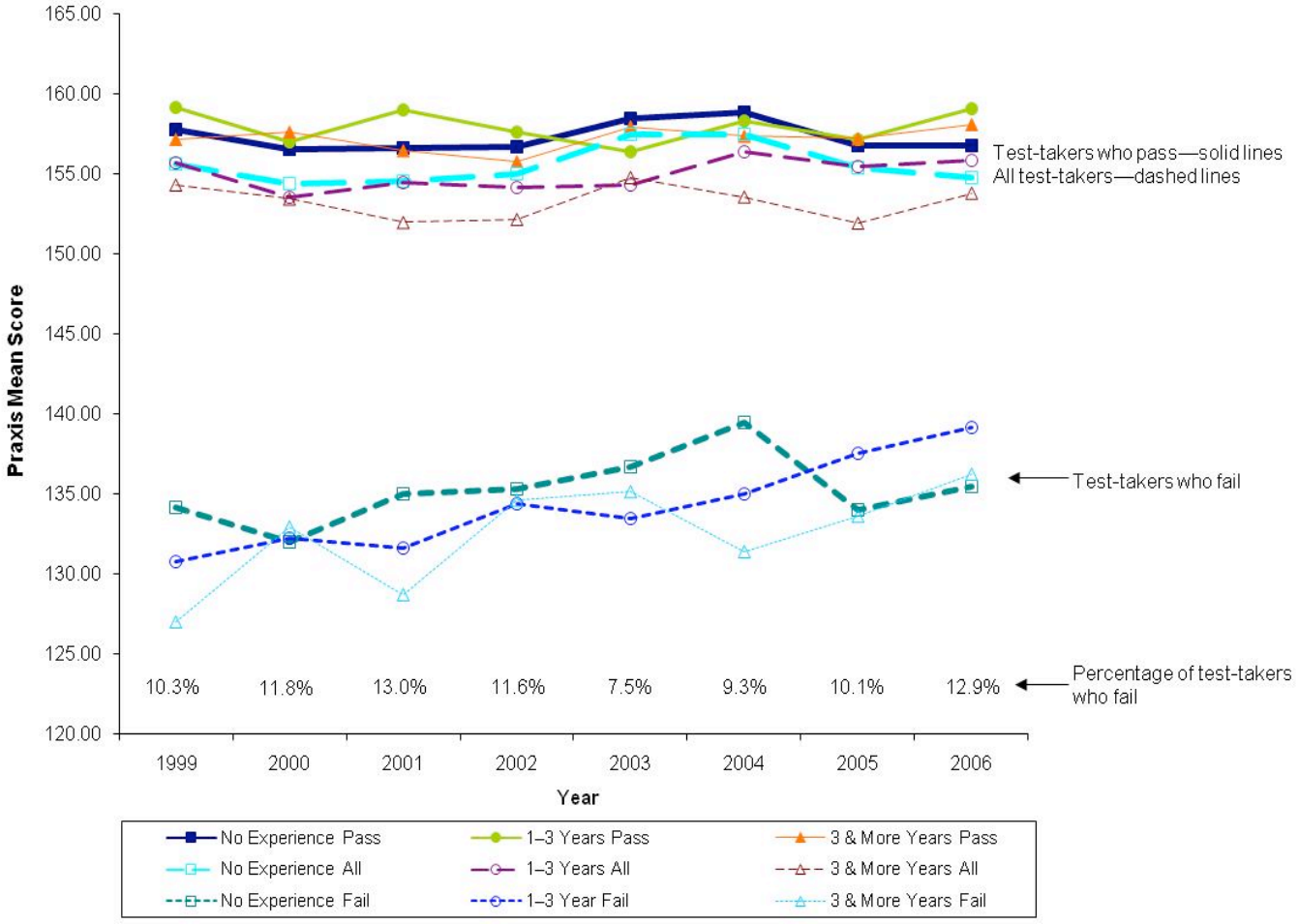


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Scores for those who fail increase across testing years.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

The regression analysis provides more detail about these trends. There is no relationship between testing years and test scores using the simple model. This holds for both passers and all test-takers (see Exhibit 27).

The comprehensive model for passers also does not show any trend in score changes across selected years. However female candidates perform slightly better than male candidates and white candidates score higher than African-American candidates. Those with experience have slightly higher scores as well.

In the comprehensive model for all candidates, there is an overall small positive trend in scores, likely due to the increase in scores for those who fail. Gender, race and ethnicity effects are larger in this model. The effect of experience is no longer significant.

The comprehensive analysis confirms that male candidates who pass have slightly higher scores than females, and white candidates who pass score higher than African-American candidates. There is also a main effect of experience with those having the most experience also having the highest scores.

Exhibit 27
Regression Results—English Language, Literature, and Composition: Pedagogy

Variable	Std Wts. – Simple Model Passers	Std Wts. – Simple Model All	Std Wts. – Comprehensive Model Passers	Std Wts. – Comprehensive Model All
Test Year	.006 ns	.019 ns	.013 ns	.026+
Gender			.031*	.062*
Race and Ethnicity			.141*	.212*
Experience			.036*	-.006 ns
Adjusted R Squared	.000 ns	.000 ns	.020*	.048*

+ p=.013

Exhibit Reads: There is no change in scores across testing years for those who pass as well as for all candidates in the simple model. There is a slight increase over years when all candidates are included in the comprehensive model.

E. TEST 0061—MATHEMATICS: CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 50 multiple-choice items that focus on knowledge and reasoning in topics relevant to teaching secondary mathematics: algebra and number theory, measurement, geometry, trigonometry, functions, calculus, data analysis and statistics, probability, matrix algebra and discrete mathematics.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Data from 17 states are included. Across the eight years included in the analyses, the proportion of African-American candidates increases substantially and the number of candidates more than doubles. Passing rates decline a large amount as well. The proportion of teachers with no experience shows a small but relatively consistent decline (see Exhibit 28).

Exhibit 28
Test-taker Characteristics—Mathematics: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
Sample Size	2,552	2,721	3,880	4,877	5,569	5,469	5,748	5,785
Percent Male	43.1%	40.9%	43.6%	47.2%	46.3%	44.1%	43.4%	43.7%
Percent African-American	7.5%	8.6%	10.6%	10.0%	11.0%	11.5%	11.5%	12.1%
Percent White	84.4%	82.1%	78.9%	78.1%	77.1%	77.8%	77.2%	77.3%
Percent Passed	82.2%	81.7%	78.8%	79.4%	74.5%	75.5%	74.9%	70.9%
Percent No Experience	65.5%	64.6%	63.3%	63.6%	63.4%	62.1%	61.9%	62.7%
Percent 1– Years of Experience	17.4%	17.3%	19.2%	19.2%	19.9%	21.4%	21.4%	20.3%
Percent More than 3 Years of Experience	17.1%	18.1%	17.5%	17.3%	16.7%	16.5%	16.7%	17.0%

Exhibit Reads: The proportion of African-American candidates has increased. Passing rates show a decline.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Male candidates who pass the Mathematics: Content Knowledge test appear to have somewhat higher passing scores than female candidates who pass. White candidates have higher passing scores than African-American candidates (see Exhibits 29, 30 and 31).

As with other tests, there are large difference in scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 153 (151) and 117 (117), respectively (sd = 20.9). Scores for those who pass are well above the median passing standards for states. The standard error of measurement (SEM) for this test is 8.6. Three quarters of candidates (74.7 percent) who do not pass the assessment score at least one SEM below (a score of 127) the median passing score. Thus, their scores are well below the state standards that have been established (see Exhibit 32**).** Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 29
Mean *Praxis* Scores (SD) of Those Who Pass by Year—Mathematics: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
All	153.08 (15.07)	153.75 (15.56)	152.19 (15.45)	152.83 (14.86)	153.08 (14.91)	152.97 (14.32)	154.24 (15.00)	154.49 (14.92)
Female	150.89 (14.10)	152.09 (14.82)	150.32 (14.20)	151.11 (13.42)	151.03 (13.89)	150.68 (12.88)	152.17 (13.68)	152.03 (13.63)
Male	155.91 (15.81)	156.14 (16.30)	154.52 (16.57)	154.66 (16.06)	155.26 (15.64)	155.36 (15.33)	156.37 (15.97)	157.14 (15.83)
African-American	144.51 (13.82)	147.35 (13.75)	145.11 (13.27)	144.86 (11.92)	145.64 (11.24)	146.21 (11.97)	146.11 (10.79)	147.43 (11.80)
White	153.46 (15.02)	154.20 (15.54)	152.69 (15.40)	153.57 (14.83)	153.32 (15.01)	153.06 (14.20)	154.46 (14.92)	154.74 (14.94)
No Experience	152.69 (14.29)	153.76 (15.14)	152.21 (15.23)	153.18 (14.67)	153.19 (14.51)	152.69 (13.81)	154.55 (14.89)	155.07 (14.81)
1–3 Years of Experience	152.29 (15.54)	151.16 (14.42)	150.44 (14.55)	151.77 (15.29)	152.09 (14.77)	151.86 (14.52)	152.88 (14.57)	152.43 (15.00)
More than 3 Years of Experience	155.43 (17.53)	155.50 (17.26)	153.87 (16.86)	151.98 (14.74)	153.43 (16.83)	153.78 (15.56)	152.30 (15.51)	153.36 (15.00)

Exhibit Reads: Scores appear to be relatively stable across testing years. There are consistent gender, race and ethnicity differences.

Exhibit 30
Mean *Praxis* Scores of Test-takers by Gender, Passing Status and Teaching Experience—Mathematics: Content Knowledge

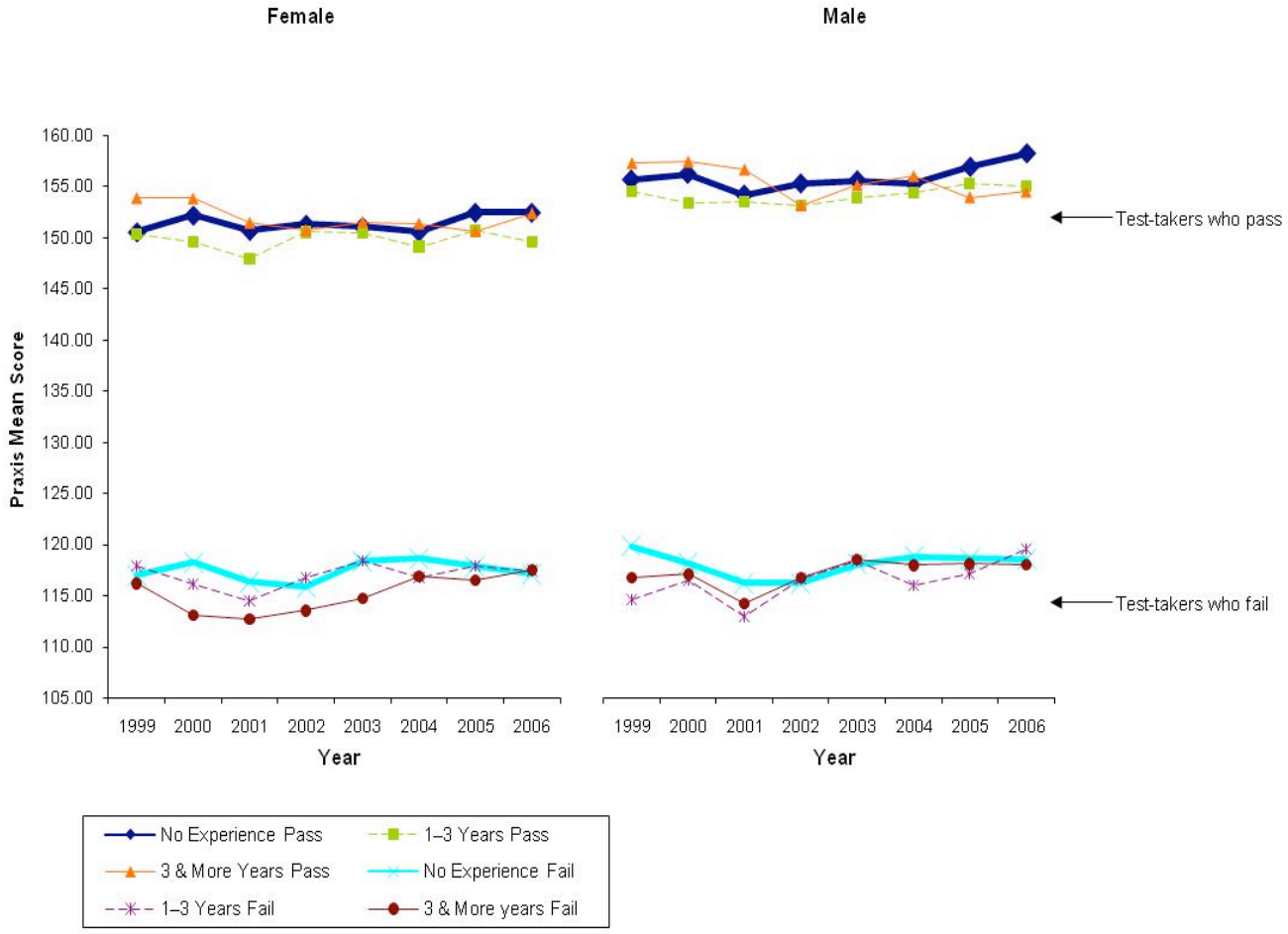


Exhibit Reads: Mean scores for those who pass are substantially higher than for those who do not.

Exhibit 31
Mean *Praxis* Scores of Test-takers by Race, Ethnicity, Passing Status and Teaching Experience—Mathematics: Content Knowledge

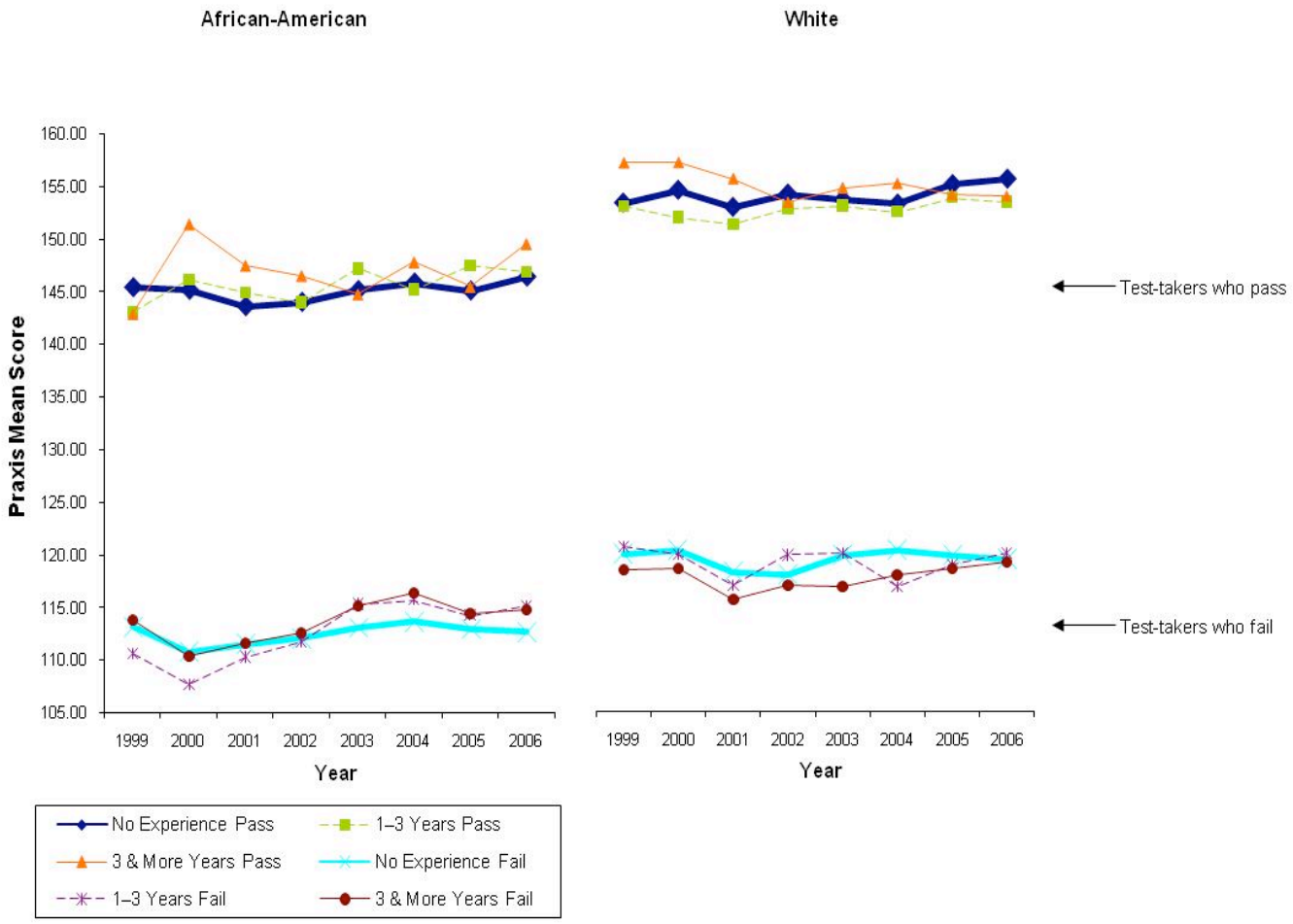


Exhibit Reads: Scores for white candidates are higher than for African-American candidates

Exhibit 32
Distribution of Scores in 2006 for Mathematics: Content Knowledge

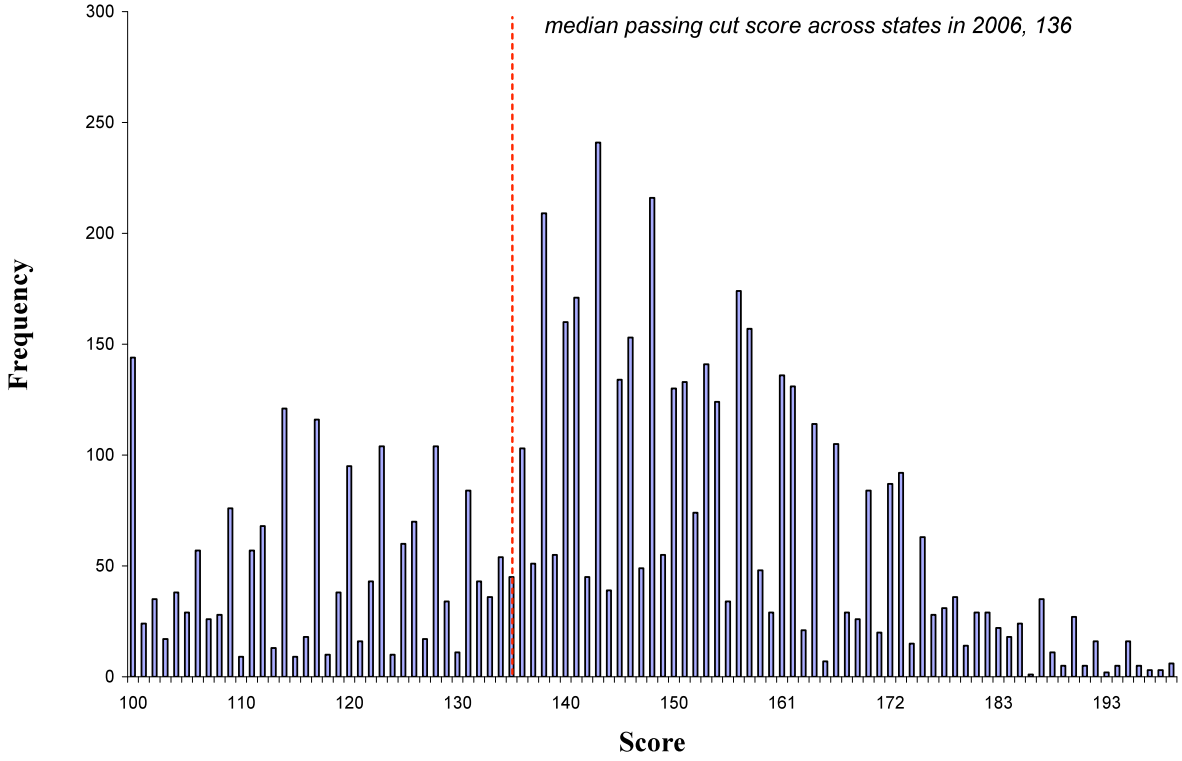


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Mean passing scores for all test-takers are disaggregated by experience. What is quite noticeable is the relatively large downward trend for all candidates, particularly teachers with experience. Such trends are not apparent for either those who pass or those who fail. However, the failure rates rise substantially over time, suggesting that the different pattern for all candidates compared with each of the subgroups (those who pass and those who fail) again is an instance of Simpson's paradox (see Exhibit 33).

Exhibit 33
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—Mathematics: Content Knowledge

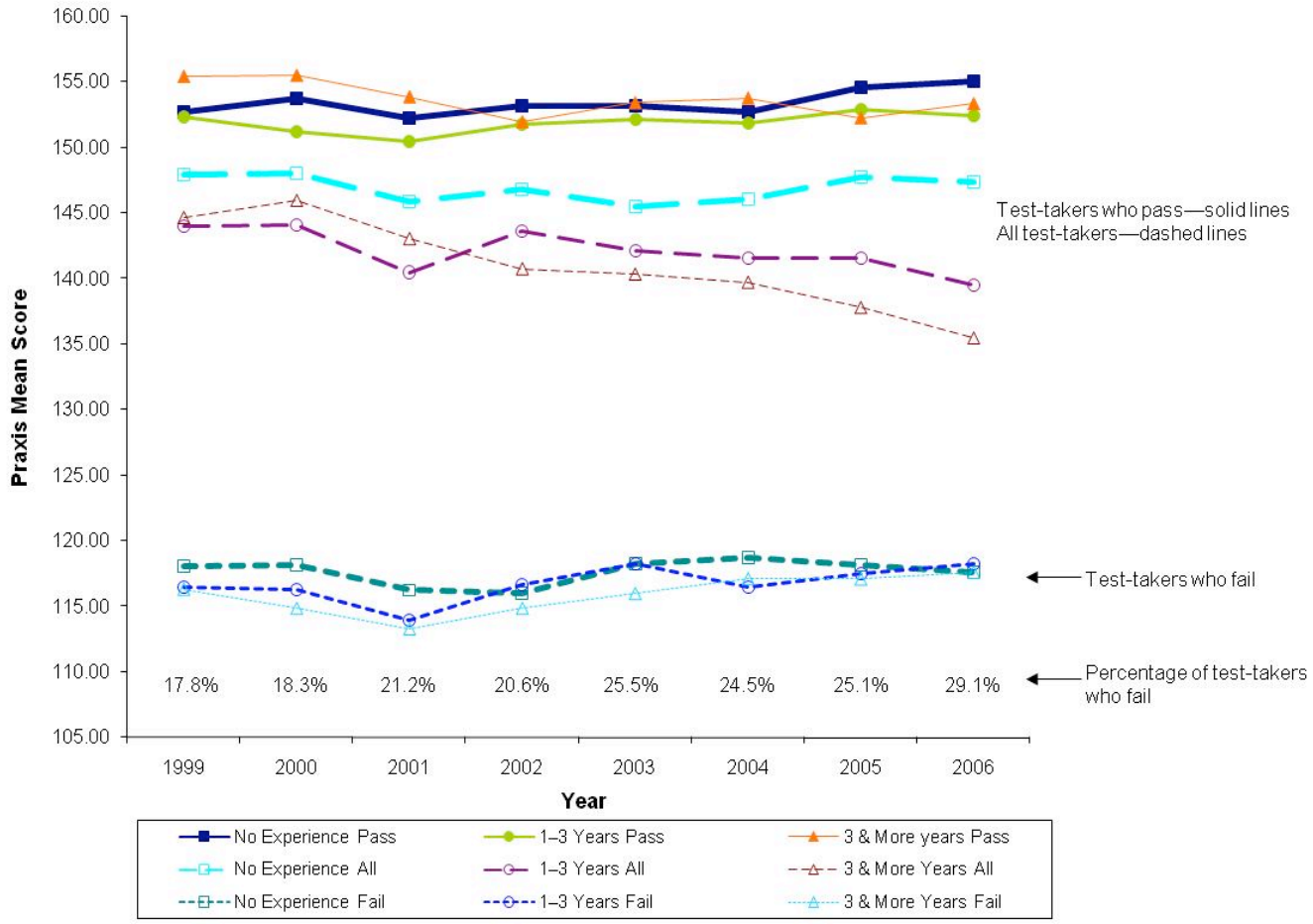


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Failure rates increase substantially across testing years.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

In the regression analyses, there is a small increase in scores for those who pass in the simple model. That trend is reversed if all candidates are considered (see Exhibit 34).

Using the comprehensive model, there are again divergent results. Scores increase for those who pass. When taking into account all candidates, the overall trend is negative. In all cases, the magnitude of these differences is small.

Males score higher than females and these results are relatively consistent regardless of the sample. The race and ethnicity difference is larger when all candidates are considered as African-Americans who fail also score lower than whites who fail. Experience differences are larger when all candidates are considered due to the inclusion of candidates who do not pass.

Exhibit 34
Regression Results for 0061 Mathematics: Content Knowledge

Variable	Std Wts. – Simple Model Passers	Std Wts. – Simple Model All	Std Wts. – Comprehensive Model Passers	Std Wts. – Comprehensive Model All
Test Year	.032*	-.032*	.020*	-.033*
Gender			-.149*	-.128*
Race and ethnicity			.132*	.300*
Experience			-.013+	-.087*
Adjusted R Squared	.001*	.001*	.040*	.124*

+ p = .039

Exhibit Reads: Scores increase across testing years for those who pass *Praxis.* Scores decrease when all candidates are included in the models.

F. TEST 0065—MATHEMATICS: PEDAGOGY

Background Information on the Test:

This test consists of three constructed-response essay questions, focused on planning, implementing and assessing instruction in mathematics, respectively

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Across eight years, data from six states are included. Compared with the mathematics test of content (0061), far fewer candidates take this test, although the number of test-takers more than doubles. There appears to be a general increase in the proportion of African-Americans who take the test. No other obvious trends in the demographics of test-takers are apparent although there is much more fluctuation than for tests with larger samples of test-takers (see Exhibit 35).

Exhibit 35
Test-taker Characteristics—Mathematics: Pedagogy

	1999	2000	2001	2002	2003	2004	2005	2006
Sample Size	341	412	478	555	595	774	731	797
Percent Male	42.2%	38.8%	40.2%	40.9%	41.0%	42.0%	42.5%	43.7%
Percent African-American	11.7%	9.7%	11.3%	12.8%	12.9%	12.1%	15.5%	14.9%
Percent White	75.0%	79.1%	76.4%	73.4%	75.1%	74.4%	71.6%	67.1%
Percent Passed	81.8%	82.3%	76.2%	77.5%	79.7%	85.5%	86.2%	82.1%
Percent No Experience	46.7%	48.4%	50.3%	48.1%	49.9%	43.9%	40.2%	39.3%
Percent 1–3 Years of Experienc	27.6%	28.3%	30.3%	25.8%	28.4%	32.6%	36.4%	36.0%
Percent More than 3 Years of Experience	25.8%	23.3%	19.4%	26.2%	21.7%	23.6%	23.4%	24.7%

Exhibit Reads: The proportion of African-American candidates has increased slightly across testing years. The proportion of those with no prior teaching experience decreases across testing years. Passing rates are stable. The relatively small number of candidates leads to more year to year fluctuation than for other tests.

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Because of the small number of individuals in particular cells when data are disaggregated by experience, gender, race and ethnicity, more detailed figures by gender, race and ethnicity are not presented.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Exhibit 36
Mean *Praxis* Scores (SD) of Those Who Pass by Year—Mathematics: Pedagogy

	1999	2000	2001	2002	2003	2004	2005	2006
All	147.33 (12.49)	149.35 (13.52)	146.44 (14.52)	151.06 (15.84)	151.47 (15.51)	153.74 (16.75)	152.48 (15.94)	155.05 (17.67)
Female	148.40 (12.21)	149.58 (13.46)	147.28 (14.69)	151.43 (16.12)	151.37 (15.88)	154.01 (16.92)	153.04 (15.61)	155.37 (17.43)
Male	145.92 (12.75)	148.94 (13.66)	145.17 (14.23)	150.48 (15.42)	151.62 (14.95)	153.38 (16.55)	151.28 (15.97)	154.73 (17.90)
African-American	142.38 (15.22)	140.65 (10.69)	147.00 (15.12)	145.00 (13.01)	145.23 (13.85)	152.02 (15.32)	147.67 (14.79)	153.28 (14.89)
White	148.13 (11.97)	149.65 (13.54)	146.62 (15.10)	152.00 (16.21)	151.82 (15.41)	154.14 (17.17)	153.25 (16.51)	155.67 (18.44)
No Experience	147.20 (12.63)	148.39 (13.27)	145.88 (13.59)	149.86 (15.26)	151.59 (15.17)	152.99 (15.75)	151.50 (15.50)	156.38 (17.87)
1-3 Years of Experience	144.58 (11.34)	148.71 (13.45)	147.87 (16.32)	152.74 (16.91)	151.96 (15.71)	155.59 (18.38)	154.01 (17.18)	155.21 (17.57)
More than 3 Years of Experience	150.00 (12.69)	151.62 (13.78)	144.93 (13.95)	151.29 (15.72)	149.57 (16.62)	152.75 (16.08)	151.76 (15.04)	153.08 (17.35)

Exhibit Reads: Scores appear to increase across testing years. There are consistent gender, race, and ethnicity differences.

Mean passing scores for all test-takers are disaggregated by experience. There are large differences in scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 152 (150) and 116 (115), respectively. Scores for those who pass are well above the median passing standards for states. Scores for those who fail are well below the median passing standards for states (see Appendix B and Exhibit 37).

There is more year-to-year variation in scores, likely attributable to the smaller sample size, and scores appear to increase for those who pass (see Exhibits 36 and 38). Plots are shown for those who pass, those who fail, and all candidates, regardless of passing status. Although failure rates fluctuate considerably, there appears to be a relatively consistent increase in scores for those who pass as well as for all candidates (see Exhibit 38).

Exhibit 37
Distribution of Scores in 2006 for Mathematics: Pedagogy

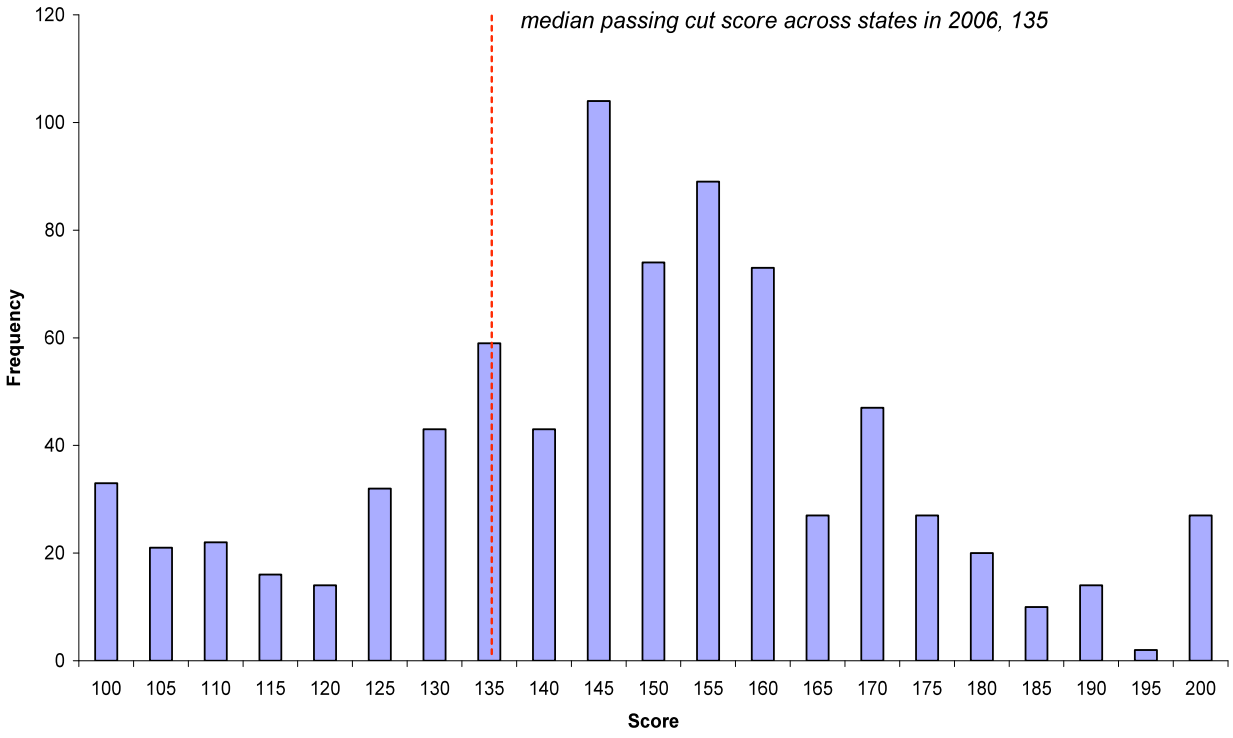


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Exhibit 38
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—Mathematics: Pedagogy

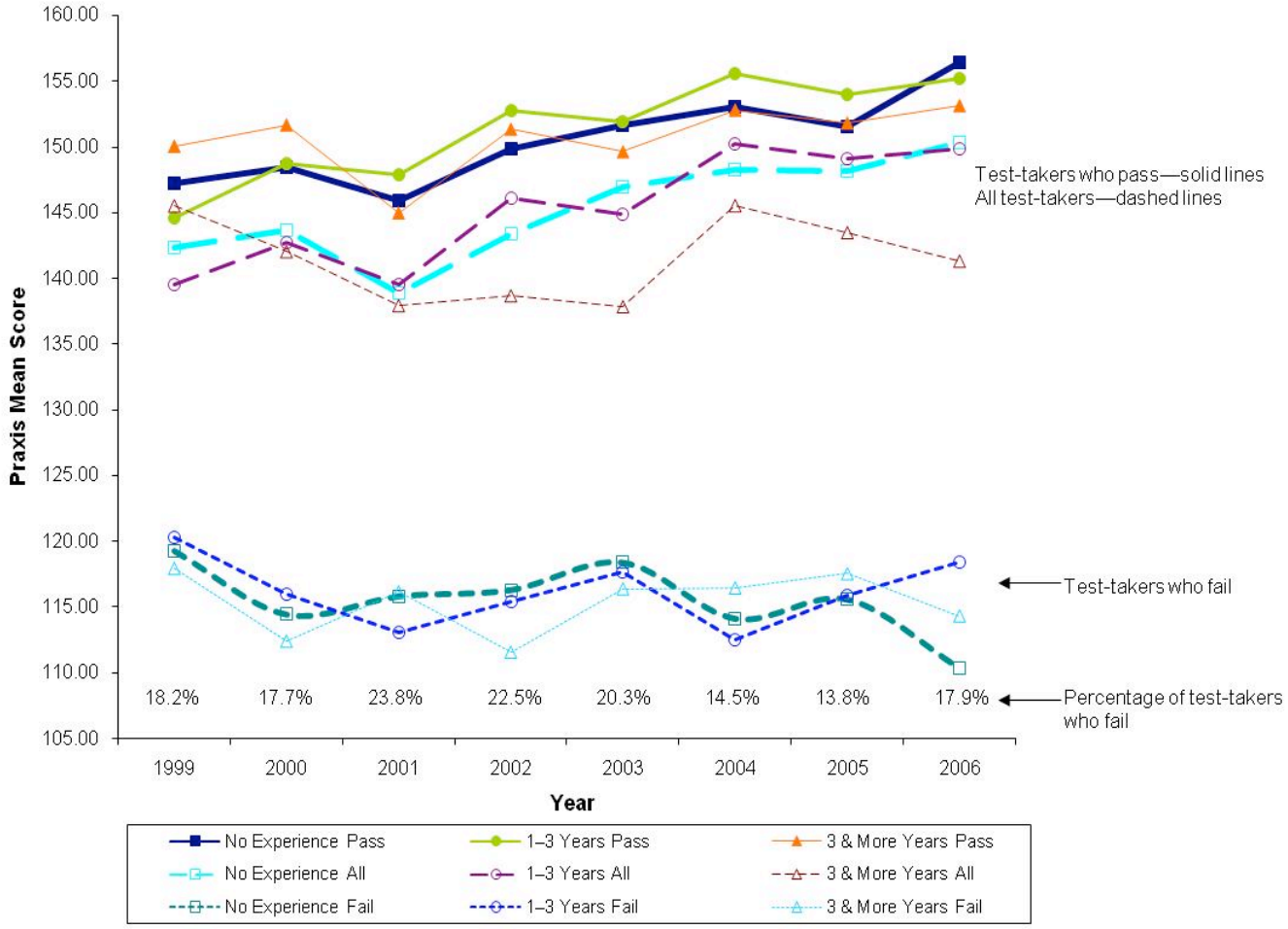


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Test scores increase substantially across testing years.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends? The regression analysis provides more detail about these trends. For all models there is a positive trend in scores over time (see Exhibit 39).

In the comprehensive model of passers there are relatively small race, ethnicity and experience effects, with those having more experience having slightly higher mean scores. For all candidates, female candidates score slightly higher, the race and ethnicity difference is greater and experience is no longer significantly related to test performance.

Exhibit 39
Regression Results—Mathematics: Pedagogy

Variable	Std Wts.— Simple Model Passers	Std Wts.— Simple Model All	Std Wts.— Comprehensive Model Passers	Std Wts.— Comprehensive Model All
Test Year	.151*	.119*	.156*	.140*
Gender			.029 ns	.037+
Race and Ethnicity			.097*	.240*
Experience			.052*	.021 ns
Adjusted R Squared	.023*	.014*	.033*	.072*

+ p=.016

Exhibit Reads: Scores increase substantially across testing years. Gender effects are non-significant for those who pass the test.

G. TEST 0069—MIDDLE SCHOOL MATHEMATICS: CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 40 multiple-choice and three constructed-response items. The multiple-choice questions focus on arithmetic and basic algebra, geometry and measurement, functions and their graphs, and data, probability, statistical concepts and discrete mathematics. The constructed-response items require candidates to complete problem solving tasks in mathematics.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Data from 23 states are included. However, because the adoption of this test was relatively recent, only three years of data are analyzed, as shown in Exhibit 40. Across the three years, the proportion of passers appears to decline as does the percent of candidates with three or more years of experience. Conversely, those with no prior teaching experience make up a larger portion of the pool over these three years.

Exhibit 40
Test-taker Characteristics—Middle School Mathematics: Content Knowledge

	2004	2005	2006
Sample Size	9,904	10,009	9,532
Percent Male	28.4%	27.3%	27.7%
Percent African-American	10.7%	11.8%	10.6%
Percent White	82.7%	80.2%	81.5%
Percent Passed	82.9%	80.2%	76.7%
Percent No Experience	44.0%	47.6%	48.0%
Percent 1–3 Years of Experience	22.8%	23.4%	23.6%
Percent More than 3 Years of Experience	33.2%	29.0%	28.4%

Exhibit Reads: Demographic characteristics are relatively stable across years while the passing rate has decreased, although it is difficult to detect trends with only three years of data.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Male candidates who pass this test appear to have somewhat higher passing scores than female candidates who pass. Those who pass without prior teaching experience have scores several points higher than those with experience. White candidates have higher passing scores than African-American candidates (see Exhibits 41, 42, and 43).

The analysis reinforces the large difference in scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of

approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 169 (167) and 140 (140), respectively (sd = 17.7). Scores for those who pass are well above the median passing standards for states. The standard error of measurement (SEM) for this test is 6.6. Three-fifths of candidates (60.2 percent) who do not pass the assessment score at least one SEM below (a score of 142) the median passing score. Thus, their scores are well below the state standards that have been established (see Exhibit 44**).** Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 41
Mean *Praxis* Scores (SD) of Those Who Pass by Year—Middle School Mathematics: Content Knowledge

	2004	2005	2006
All	169.30 (14.11)	168.54 (14.15)	168.99 (14.04)
Female	168.51 (13.84)	167.69 (13.80)	168.09 (13.70)
Male	171.07 (14.57)	170.36 (14.78)	170.90 (14.59)
African-American	159.61 (11.43)	157.83 (10.94)	158.58 (11.14)
White	170.10 (13.98)	169.53 (13.97)	169.86 (13.85)
No Experience	170.64 (14.18)	169.51 (14.22)	170.34 (13.83)
1–3 Years of Experience	168.50 (13.83)	167.77 (14.04)	167.41 (13.90)
More than 3 Years of Experience	167.56 (13.91)	167.18 (13.86)	167.71 (14.37)

Exhibit Reads: Scores appear to be stable across testing years. There are consistent gender, race, ethnicity, and experience differences.

Exhibit 42
Mean *Praxis* Scores of Test-takers by Gender, Passing Status and Teaching Experience — Middle School Mathematics: Content Knowledge

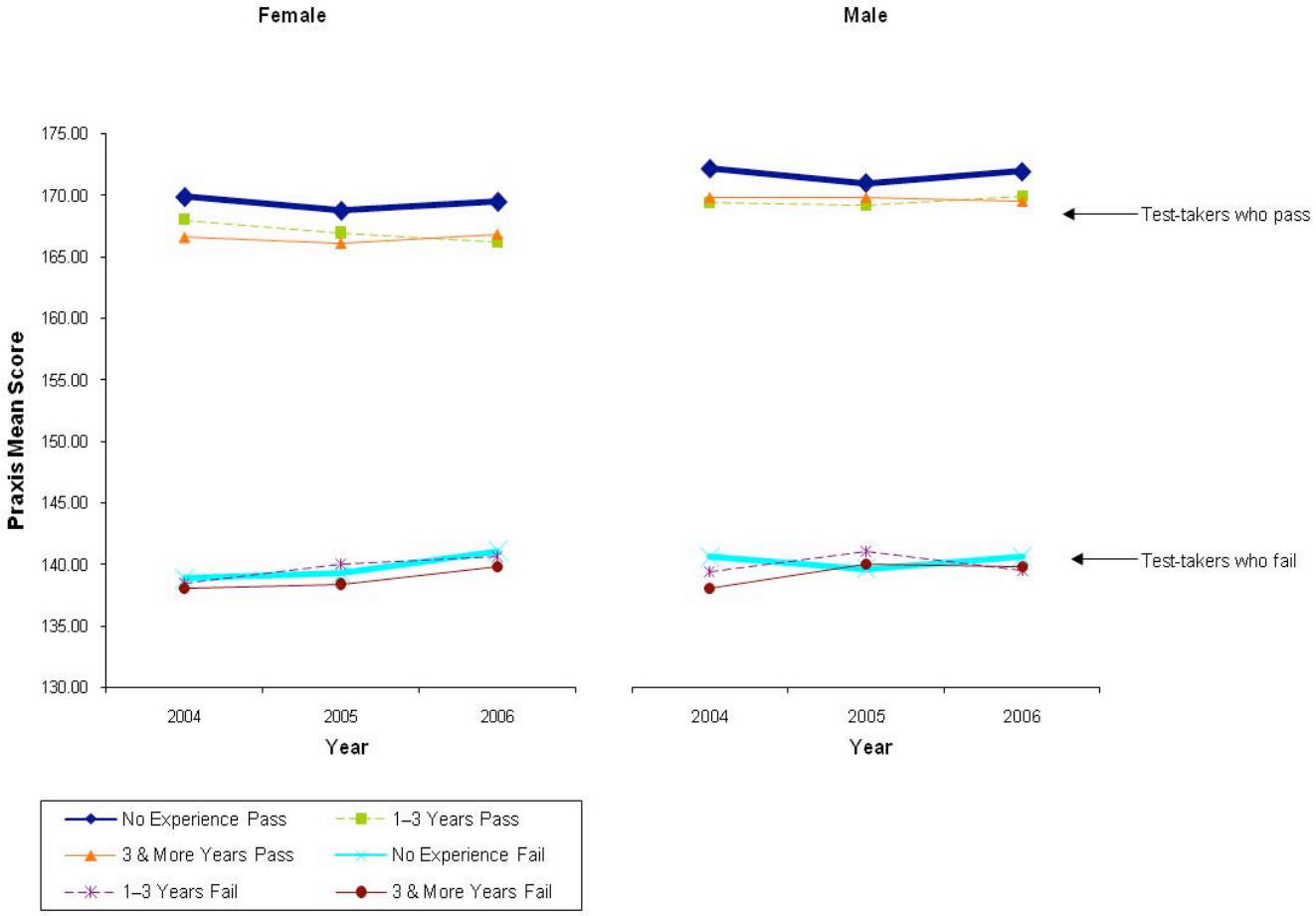


Exhibit Reads: For those who pass, scores for male candidate are higher than for female candidates.

Exhibit 43
Mean *Praxis* Scores of Test-takers by Race, Ethnicity, Passing Status and Teaching Experience—Middle School Mathematics: Content Knowledge

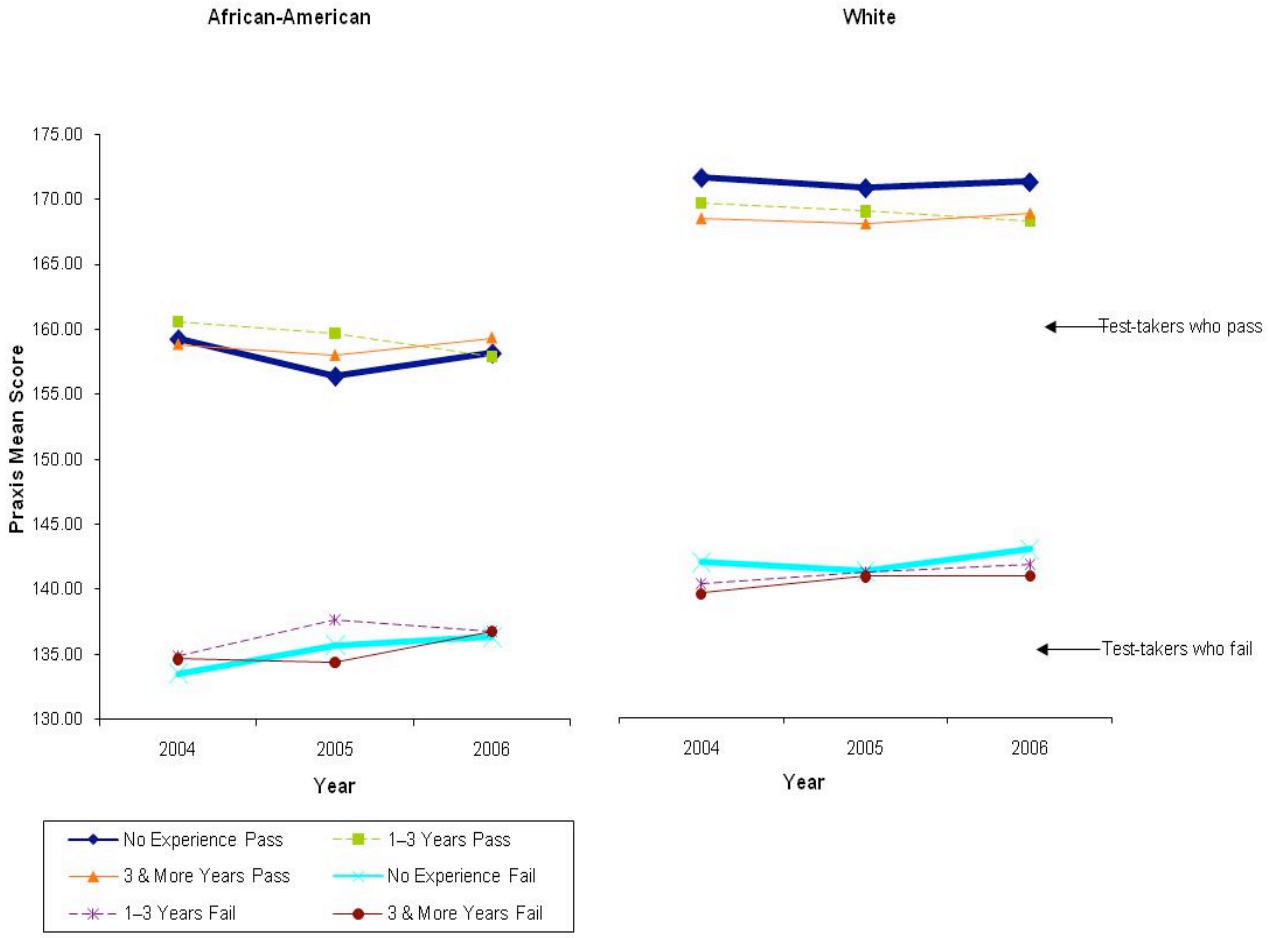


Exhibit Reads: Scores for white candidates are higher than for African-American candidates.

Exhibit 44
Distribution of Scores in 2006 for Middle School Mathematics: Content Knowledge

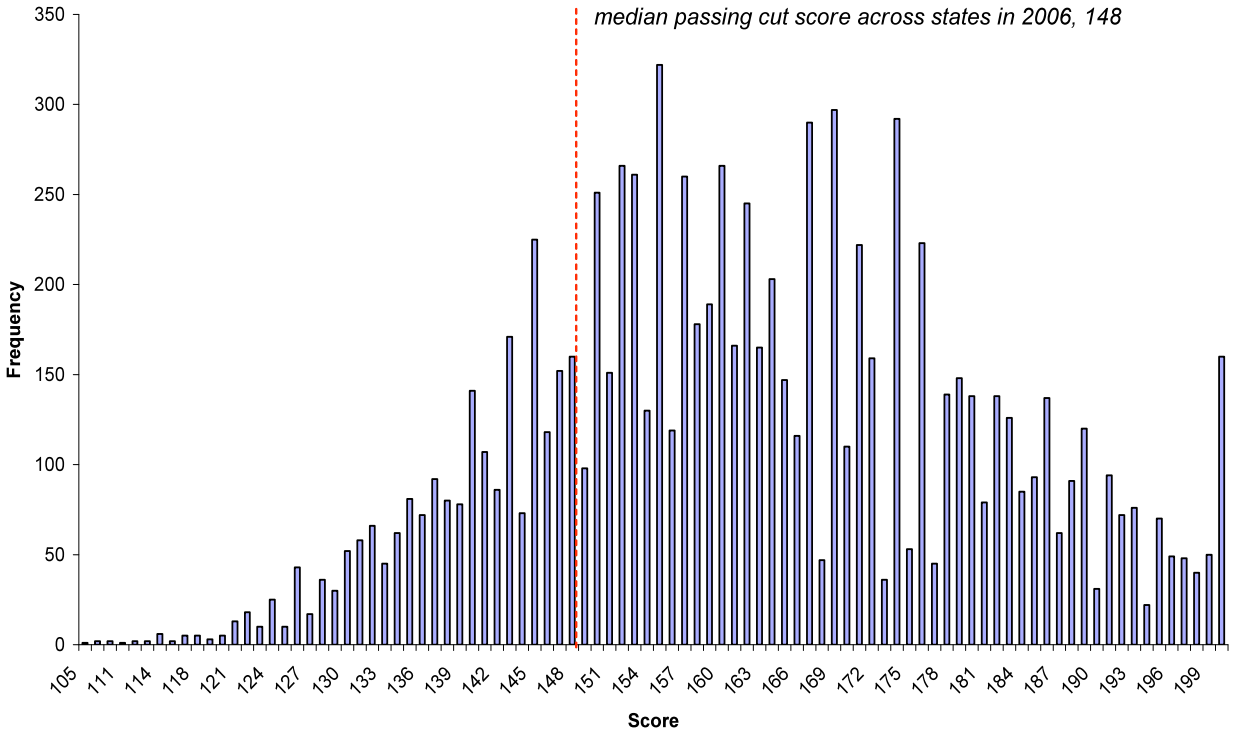


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Mean passing scores for all test-takers are disaggregated by experience. Plots are shown for those who pass, those who fail, and all candidates, regardless of passing status. The percent of those who fail is presented at the bottom of the exhibit. Interestingly, while scores for those who pass, regardless of experience, seem to be very stable, there does seem to be a slight downward trend when all candidates are included. For passers, those who have no prior teaching experience appear to have slightly higher mean scores (see Exhibit 45).

Exhibit 45
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—Middle School Mathematics: Content Knowledge

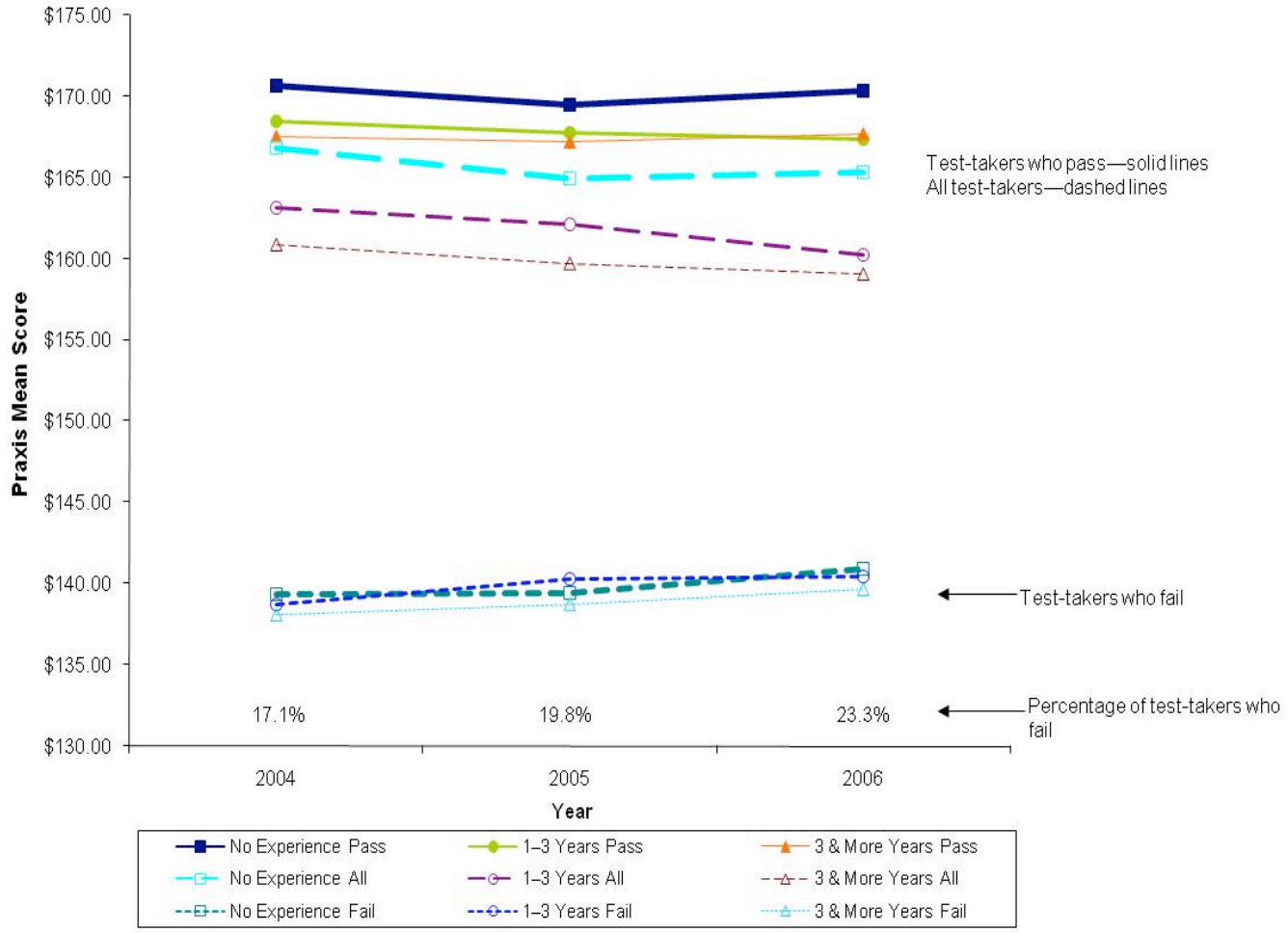


Exhibit Reads: Candidates who pass have much higher scores than those who fail. Candidates with no teaching experience have higher scores than experienced teachers.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

In the regression analyses, for those who pass, mean scores do not vary across years. However, there is a very slight decline when all candidates are considered. However, though the trend is discernible, it is extremely small and explains a very small portion of the score variation (see Exhibit 46).

The comprehensive analysis confirms that male candidates who pass have slightly higher scores than female candidates and white candidates who pass score higher than African-American

candidates. There is also a main effect of experience with those having no teaching experience also having the highest scores.

A slightly different picture emerges when all candidates are considered. There is a small, but significant, decline across selected years. Effects for race, ethnicity and experience are larger when all candidates are considered together.

Though scores are stable for those who pass and actually increase over the three years for those who fail, the observed decrease for all candidates is attributable to the larger proportion of candidates who fail the test over time.

Exhibit 46
Regression Results—Middle School Mathematics: Content Knowledge

Variable	Std Wts.— Simple Model Passers	Std Wts.— Simple Model All	Std Wts.— Comprehensive Model Passers	Std Wts.— Comprehensive Model All
Test Year	-.010 ns	-.041*	-.013 ns	-.042*
Gender			-.096*	-.098*
Race and Ethnicity			.216*	.305*
Experience			-.079*	-.123*
Adjusted R Squared	.000 ns	.002*	.063*	.123*

Exhibit Reads: There is no change in scores across testing years for those who pass. There is a declining trend when all candidates are included.

H. TEST 0431—GENERAL SCIENCE, PART 1: CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 60 multiple choice questions focused on basic principles of science, life science, earth and space science, technology and society, and methodology, measurement and safety.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Across eight years, data from six states are included. Sample sizes are relatively small, causing greater year to year fluctuation in demographic characteristics, though no strong trends are evident. The number of candidates taking this test does not increase in the same way as for other tests during this time period. There appears to be a general decline in passing rates (see Exhibit 47).

Exhibit 47
Test-taker Characteristics—General Science, Part 1: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
Sample Size	773	801	994	1,057	1,198	1,091	968	952
Percent Male	43.2%	45.8%	42.9%	50.0%	44.3%	43.6%	46.6%	42.3%
Percent African-American	6.4%	4.3%	5.3%	5.6%	6.4%	6.9%	5.7%	5.7%
Percent White	82.2%	81.8%	77.3%	77.0%	76.3%	74.2%	75.9%	76.3%
Percent Passed	92.1%	89.9%	90.6%	91.0%	85.3%	86.2%	86.9%	84.7%
Percent No Experience	59.7%	63.1%	62.3%	63.7%	59.7%	60.8%	58.9%	62.6%
Percent 1–3 Years of Experience	20.4%	18.3%	18.8%	19.4%	22.0%	20.3%	23.6%	21.0%
Percent More than 3 Years of Experience	19.9%	18.6%	18.9%	16.9%	18.3%	19.0%	17.5%	16.4%

Exhibit Reads: Demographic characteristics are relatively stable across years while the passing rate has decreased. The relatively small number of candidates leads to more year to year fluctuation than for other tests.

The mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Because of the small number of individuals in particular cells if data are disaggregated by experience, gender, race and ethnicity, more detailed figures by gender, race and ethnicity are not presented.

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

Male candidates appear to outperform female candidates and mean scores for white candidates are higher than for African-American candidates (see Exhibit 48). There are large differences in

scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 169 (168) and 139 (140), respectively (sd = 15.50). Scores for those who pass are well above the median passing standards for states. For those who fail, scores are well below the median passing standards for states (see Appendix B and Exhibit 49). The standard error of measurement (SEM) for this test is 7. One half of the candidates (51.2 percent) who do not pass the assessment score at least one SEM below (a score of 141) the median passing score. Thus, their scores are well below the state standards that have been established. Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 48
Mean *Praxis* Scores (SD) of Those Who Pass by Year—General Science, Part 1: Content Knowledge

	1999	2000	2001	2002	2003	2004	2005	2006
All	167.58 (12.09)	169.09 (12.58)	167.94 (12.61)	169.29 (13.08)	169.89 (12.31)	169.76 (11.67)	169.93 (11.90)	168.41 (11.77)
Female	165.78 (11.31)	167.08 (12.17)	164.89 (11.58)	166.09 (12.18)	167.41 (11.44)	166.97 (10.94)	166.71 (10.87)	165.96 (11.08)
Male	169.87 (12.66)	171.39 (12.68)	171.72 (12.84)	172.25 (13.20)	172.71 (12.65)	172.94 (11.77)	172.84 (12.11)	171.21 (11.98)
African-American	156.61 (8.26)	157.88 (9.17)	158.13 (10.03)	159.70 (12.07)	161.18 (9.74)	160.73 (8.16)	159.21 (9.34)	157.00 (6.41)
White	168.13 (11.96)	169.53 (12.24)	169.04 (12.61)	170.47 (12.84)	170.74 (12.35)	170.59 (11.58)	170.22 (11.50)	169.81 (11.87)
No Experience	166.84 (11.86)	169.12 (12.67)	168.36 (12.26)	169.60 (13.05)	169.81 (12.15)	169.63 (10.97)	169.86 (11.52)	168.17 (11.35)
1–3 Years of Experience	167.53 (11.45)	166.65 (11.59)	166.40 (13.17)	168.17 (13.08)	169.90 (12.27)	168.94 (12.56)	169.05 (11.53)	168.43 (12.41)
More than 3 Years of Experience	168.76 (12.86)	171.22 (12.95)	167.92 (13.00)	168.72 (13.42)	168.80 (13.24)	170.36 (12.96)	169.17 (13.62)	170.35 (12.85)

Exhibit Reads: Scores appear to be stable across testing years. There are consistent gender, race and ethnicity differences.

Mean passing scores for all test-takers are disaggregated by experience. Plots are shown for those who pass, those who fail, and all candidates, regardless of passing status. Although failure rates fluctuate considerably, there appears to be a relatively consistent increase in scores for those who pass as well as for all candidates. There is more year-to-year variation in scores, likely attributable to the smaller sample size (see Exhibit 50).

Exhibit 49
Distribution of Scores in 2006 for General Science, Part 1: Content Knowledge

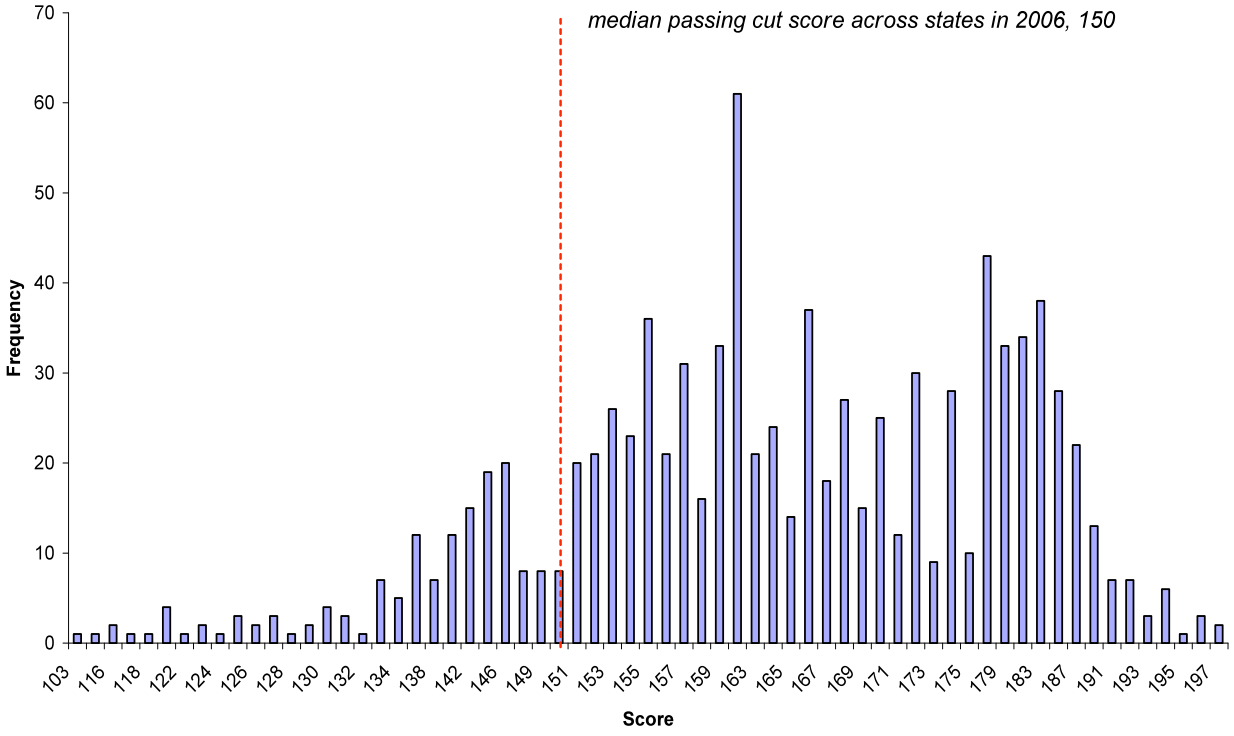


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Exhibit 50
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—General Science, Part 1: Content Knowledge

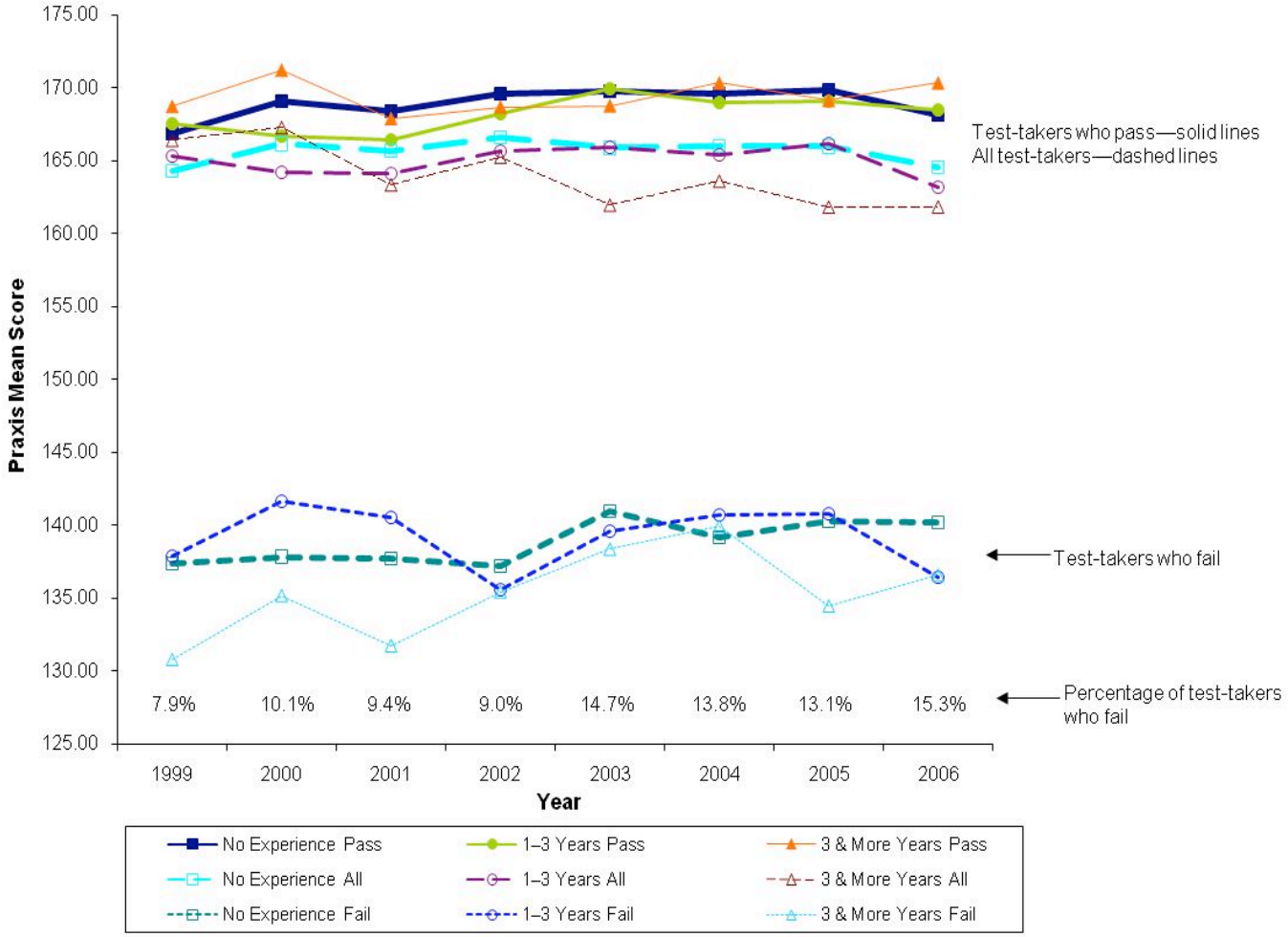


Exhibit Reads: Candidates who pass have much higher scores than those who fail.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

The regression analysis provides more detail about these trends. There is a small positive trend over time for those who pass the test in both the simple and comprehensive models. For all candidates, there is no trend effect (see Exhibit 51).

In the comprehensive models, findings for passers and all candidates are similar for variables other than the test year trend. Male candidate score higher than females, white candidates score higher than African-American candidates, and those with experience have slightly higher scores than those without experience.

Exhibit 51
Regression Results—General Science, Part 1: Content Knowledge

Variable	Std Wts.— Simple Model Passers	Std Wts.— Simple Model All	Std Wts.— Comprehensive Model Passers	Std Wts.— Comprehensive Model All
Test Year	.035*	-.016 ns	.041*	-.001 ns
Gender			-.218*	-.218*
Race and Ethnicity			.194*	.288*
Experience			.041*	.032*
Adjusted R Squared	.001*	.000 ns	.087*	.130*

Exhibit Reads: Scores increase across testing years for those who pass. There is no discernible trend when all candidates are included.

I. TEST 0571—EARTH AND SPACE SCIENCES: CONTENT KNOWLEDGE

Background Information on the Test:

This test consists of 100 multiple-choice questions, focused on basic scientific principles of earth and space sciences, tectonics and internal earth processes, earth materials and surface processes, history of the earth and its life-forms, earth's atmosphere and hydrosphere and astronomy.

1. Have the demographic or performance characteristics of individuals who took each test changed across the years included?

Across five years, data from 12 states are included. The sample for this test is relatively small and does not increase a great deal, and there are very few African-American test-takers. Thus, there are several points in this analysis in which data on race and ethnicity are not reported due to insufficient sample size in particular cells (see Exhibit 52).

Exhibit 52
Test-taker Characteristics—Earth and Space Sciences: Content Knowledge

	2002	2003	2004	2005	2006
Sample Size	631	874	809	796	829
Percent Male	50.9%	49.7%	47.7%	48.6%	48.7%
Percent African-American	3.4%	3.5%	2.7%	2.9%	2.1%
Percent White	86.8%	88.8%	92.2%	90.2%	93.1%
Percent Passed	86.8%	83.9%	82.8%	87.3%	80.2%
Percent No Experience	54.1%	56.1%	52.1%	56.4%	58.7%
Percent 1-3 Years of Experience	26.3%	26.2%	28.5%	26.9%	23.6%
Percent More than 3 Years of Experience	19.6%	17.7%	19.4%	16.7%	17.7%

Exhibit Reads: Demographic characteristics are relatively stable across years while the passing rate has decreased. The relatively small number of candidates leads to more year-to-year fluctuation than for other tests.

Mean *Praxis* scores for those who pass are disaggregated by gender, race, ethnicity and experience. Because of the small number of individuals in particular cells if data are disaggregated by experience, gender, race and ethnicity, more detailed figures by gender, race and ethnicity are not presented (see Exhibit 53).

2. Do mean *Praxis* scores differ by gender, race, and experience? Additionally, do the mean passing scores differ from those who do not pass as well or from all candidates considered together?

There are large difference in scores between those who pass and those who fail. Those who pass have substantially higher scores than those who do not, on the order of approximately two standard deviations difference. The mean (and median) scores for those who pass and those who fail are 169 (169) and 136 (138), respectively (sd = 17.5). Scores for those who pass are well

above the median passing standards for states. For those who fail, scores are well below the median passing standards for states (see Appendix B and Exhibit 54). The standard error of measurement (SEM) for this test is 5.3. More than three quarters of candidates (78.2 percent) who do not pass the assessment score at least one SEM below (a score of 145) the median passing score. Thus, their scores are well below the state standards that have been established (see Exhibit 56**).** Differences of this magnitude exceed those that might be attributed to measurement error.

Exhibit 53
Mean *Praxis* Scores (SD) of Those Who Pass by Year—Earth and Space Sciences: Content Knowledge

	2002	2003	2004	2005	2006
All	170.42 (14.12)	169.99 (12.68)	169.05 (12.59)	168.85 (12.51)	167.45 (12.74)
Female	167.61 (14.50)	167.45 (12.00)	166.53 (12.39)	166.01 (12.39)	165.52 (12.01)
Male	172.86 (13.34)	172.39 (12.84)	171.41 (12.47)	171.49 (11.98)	169.31 (13.20)
African-American	NR*	NR	NR	NR	NR
White	171.43 (13.75)	170.12 (12.62)	169.25 (12.51)	168.85 (12.29)	167.57 (12.67)
No Experience	170.54 (13.68)	169.85 (12.84)	170.19 (12.03)	169.30 (11.74)	167.74 (12.33)
1–3 Years of Experience	168.99 (14.33)	169.16 (11.84)	165.87 (12.78)	166.58 (13.50)	165.71 (13.00)
More than 3 Years of Experience	172.26 (14.78)	171.71 (13.33)	169.97 (13.25)	168.95 (13.20)	167.92 (13.50)

NR* - Data are not reported because of insufficient cell size in some cases (n<15).

Exhibit Reads: Scores appear to decline across testing years. There are consistent gender differences.

Exhibit 54
Distribution of Scores in 2006 for Earth and Space Sciences: Content Knowledge

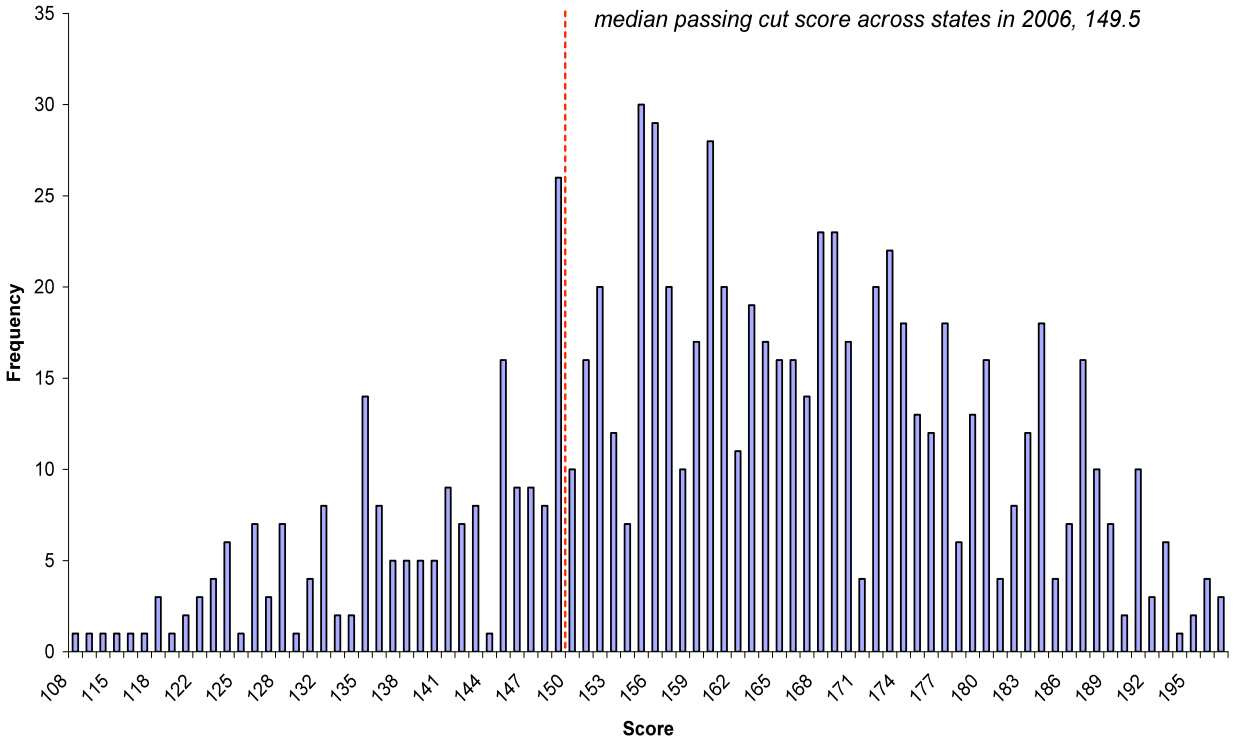


Exhibit Reads: Of the test-takers who do not pass, many have scores substantially below state passing standards.

Mean passing scores for all test-takers are disaggregated by experience. Plots are shown for those who pass, those who fail, and all candidates, regardless of passing status. Although failure rates fluctuate considerably, there appears to be a relatively consistent decrease in scores for those who pass as well as for all candidates (see Exhibit 55). There is more year-to-year variation in scores, likely attributable to the smaller sample size, and scores appear to decline over time for those who pass, as well as for all candidates (see Exhibit 55).

Exhibit 55
Mean *Praxis* Scores of Test-takers by Passing Status and Teaching Experience—
Earth and Space Sciences: Content Knowledge

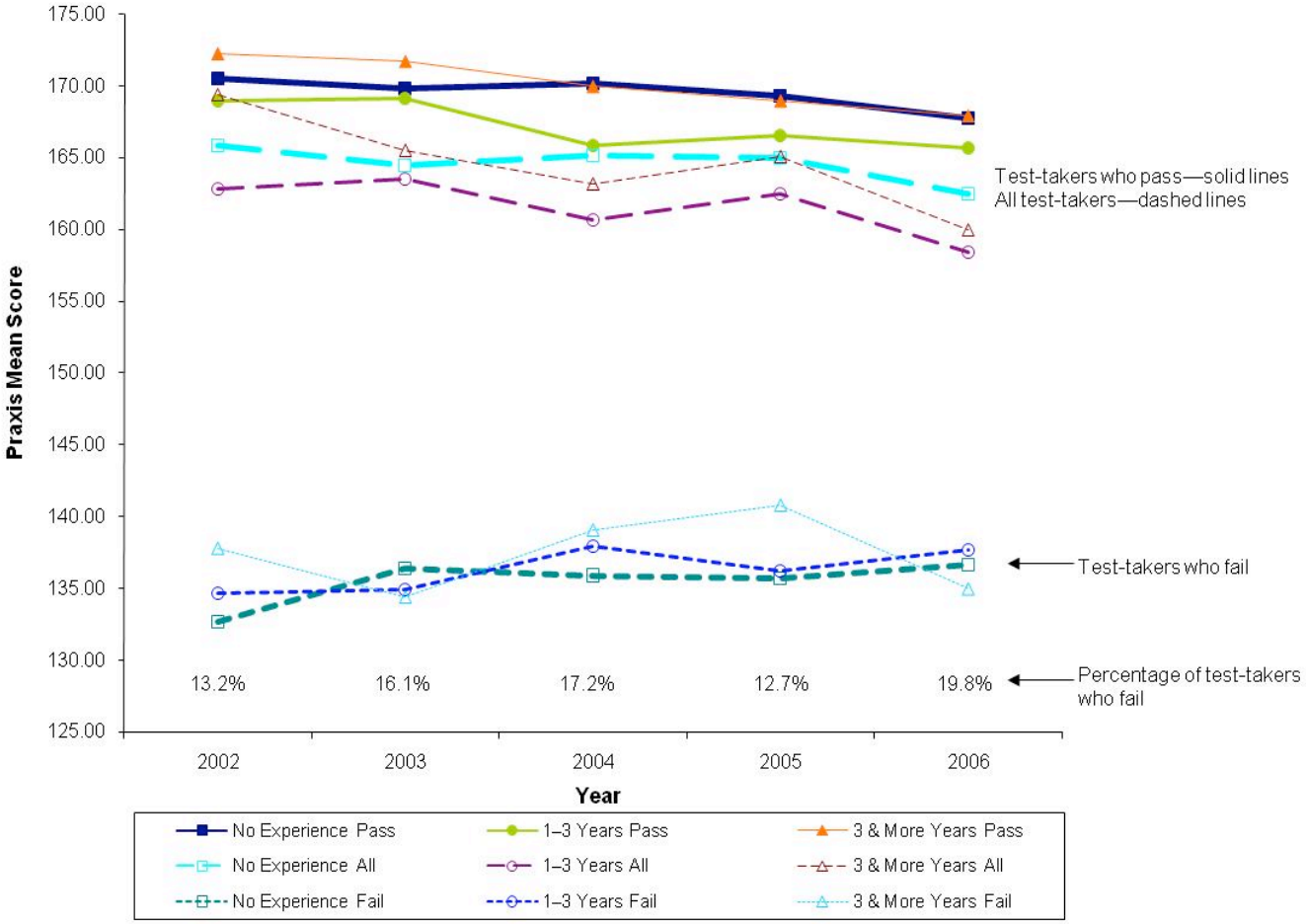


Exhibit Reads: Candidates who pass have much higher scores than those who fail. There is a decline in scores across testing years.

3. Across years, is there any evidence of a trend in *Praxis* score for those who pass? How do these trends compare with those who do not pass? Does including the factors of gender, race, and experience variables in the analytic model help to explain any observed trends?

The regression analysis provides more detail about these trends. For all models there is a declining trend in scores over time (see Exhibit 56).

In the comprehensive model for passers there are gender, race and ethnicity effects as male candidate score higher than female candidates and white candidates score higher than African-American candidates, with those having more experience having slightly higher mean scores. For all candidates, females score slightly higher, the race and ethnicity difference is greater and experience is no longer significantly related to test performance.

Exhibit 56
Regression Results—Earth and Space Sciences: Content Knowledge

Variable	Std Wts. –Simple Model Passers	Std Wts. –Simple Model All	Std Wts. – Comprehensive Model Passers	Std Wts. – Comprehensive Model All
Test Year	-.076*	-.065*	-.097*	-.088*
Gender			-.196*	-.196*
Race and Ethnicity			.113*	.261*
Experience			-.027 ns	-.026 ns
Adjusted R Square	.005*	.004*	.060*	.113*

+ p=.016

Exhibit Reads: Scores decline across testing years for those who pass as well as for all candidates

J. THE RELATIONSHIP OF STATE PASSING STANDARDS TO MEAN PASSING SCORES AND PASSING RATES

States set passing scores for particular tests in light of a number of considerations. Initially, states conduct a standard-setting process in which committees make judgments of expected skill levels for beginning teachers. These judgments lead to a recommended passing score, which relevant state authorities take into account along with other considerations, including supply and demand for particular certifications in the state, potential impact on teacher diversity, and political interests in setting passing standards at particular levels. After considering all of this information, the state then endorses a particular passing standard for candidates applying for licensure.

This process results in states setting different passing standards, which are displayed in Appendix B. In this section, we simply describe the relationship of state passing standards to the mean score of passers and to passing rates. While it might be intuitive to assume that higher passing standards would be associated with higher passing scores and lower passing rates, because these passing standards are defined in particular states, they may be drawing on differentially qualified pools of candidates. For example, if State X has a population of teacher candidates who routinely score higher on a particular test than candidates from State Y, then setting a higher passing standard in State X might not result in a lower passing rate than for State Y. Similarly, if candidates from State X score well above existing passing scores, then higher passing scores for that state might not have a substantial impact on the average score of passers—these individuals would have passed regardless of the passing score. However, if the performance of candidates from the two states is relatively similar, then we would expect that states with higher passing scores would have higher overall scores for those who pass and lower passing rates. In this section, we test the relationship between passing standards to average scores and passing rates.

For this set of analyses, we determine mean passing scores and passing rates for each passing standard set by a state over the course of the years included for each analysis. Thus, if a particular state changed its passing score during the included years, two data points from that state are included in the analysis.

We organize the analyses by general subject areas: elementary education, English, mathematics and science. For each subject, we first present charts describing the relationship of passing standards to mean scores for those who pass followed by charts describing the relationship of passing standards to passing rates. Specific state results are not identified.

An important caveat is that the regression lines presented are based on a relatively small number of data points. Therefore, the size of any particular correlation coefficient can be dramatically affected by a single point. Thus, the graphs, taken together, begin to provide a sense of these relationships, but the size of the coefficients and what they might mean should be interpreted with caution.

For the Elementary Education: Content Knowledge test, there is a nonsignificant relationship between state passing standard and mean score for passers (see Exhibit 57). The relationship is very strong for the few states included for the Elementary Education: Pedagogy test. For both tests, passing rates are lower for states with higher passing standards (see Exhibit 58).

Exhibit 57
Scatter Plots of Passing Standards vs. Mean Score by State for Passers—Elementary Education

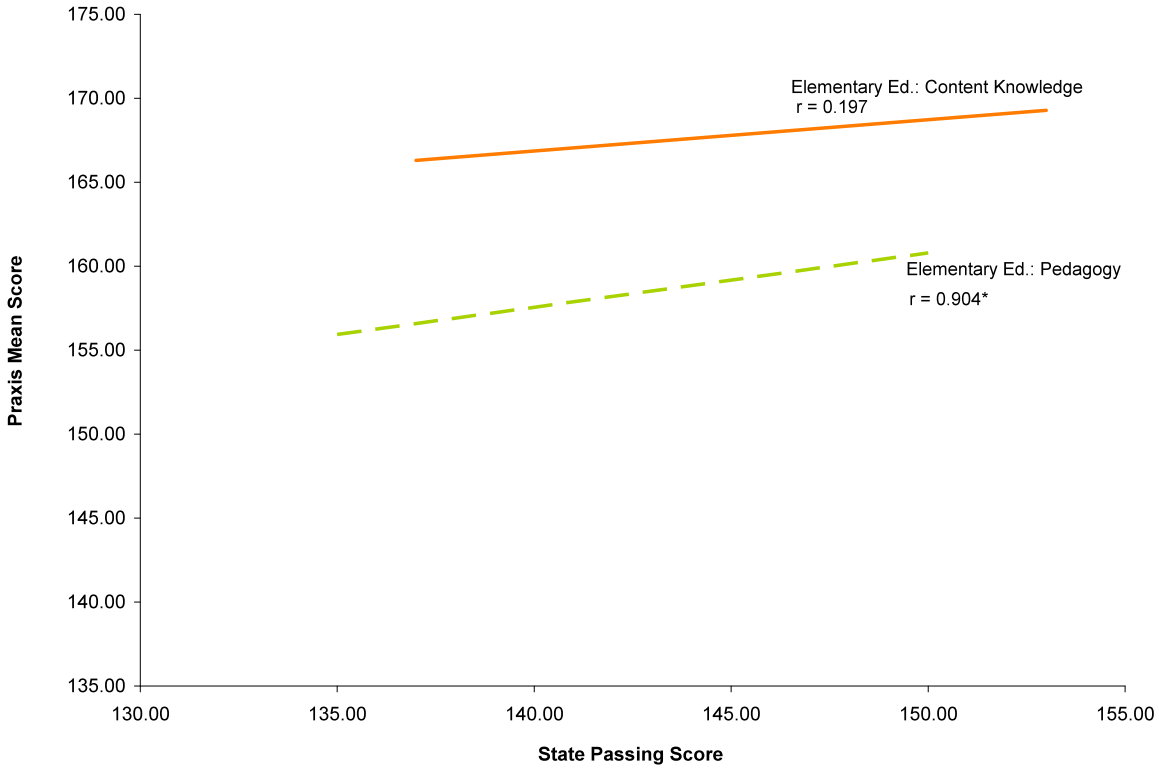


Exhibit Reads: The relationship between passing standards and mean scores for passers is not significant for the Elementary Education: Content Knowledge test.

Exhibit 58
Scatter Plots of Passing Standards vs. Passing Rate by State for Passers—Elementary Education

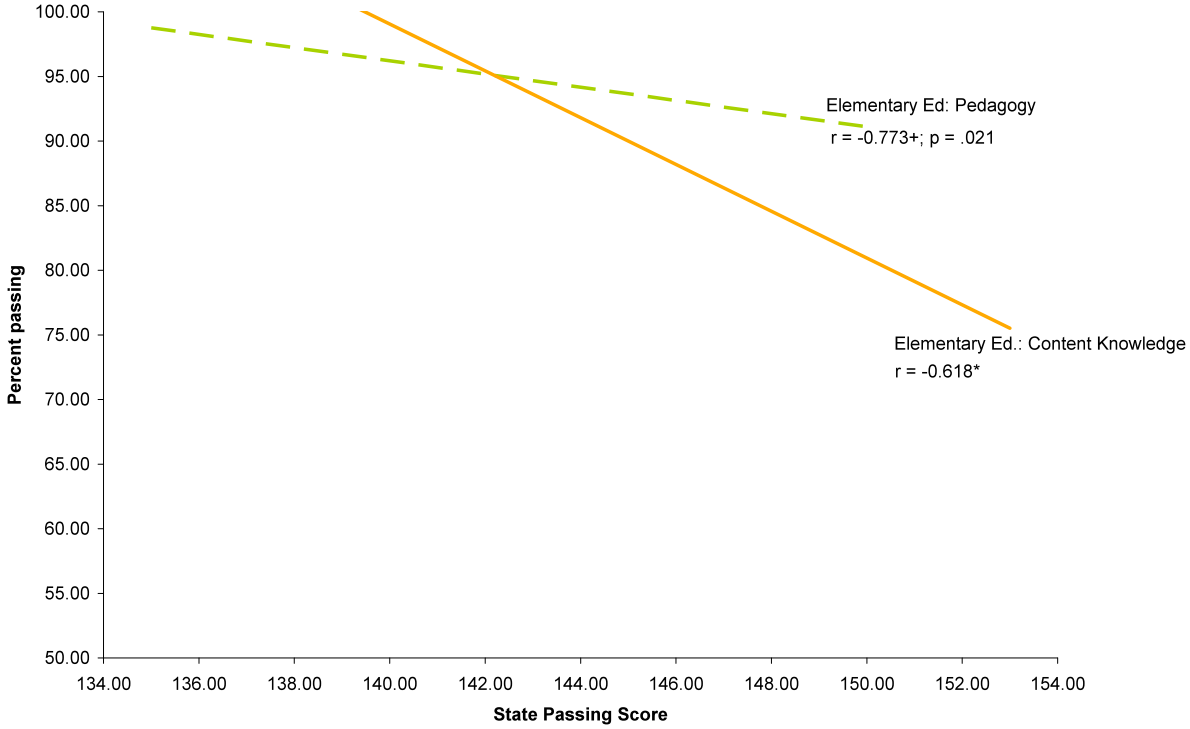


Exhibit Reads: The proportion of candidates who pass is lower in states with higher passing standards.

For the two English language tests (English Language, Literature and Composition: Content Knowledge and English Language, Literature and Composition: Pedagogy), there is a strong, positive relationship between state passing standards and mean score for passers (see Exhibit 59). For both tests, passing rates are lower for states with higher passing standards (see Exhibit 60).

Exhibit 59
Scatter Plots of Passing Standards vs. Mean Score by State for Passers—English Language, Literature, and Composition

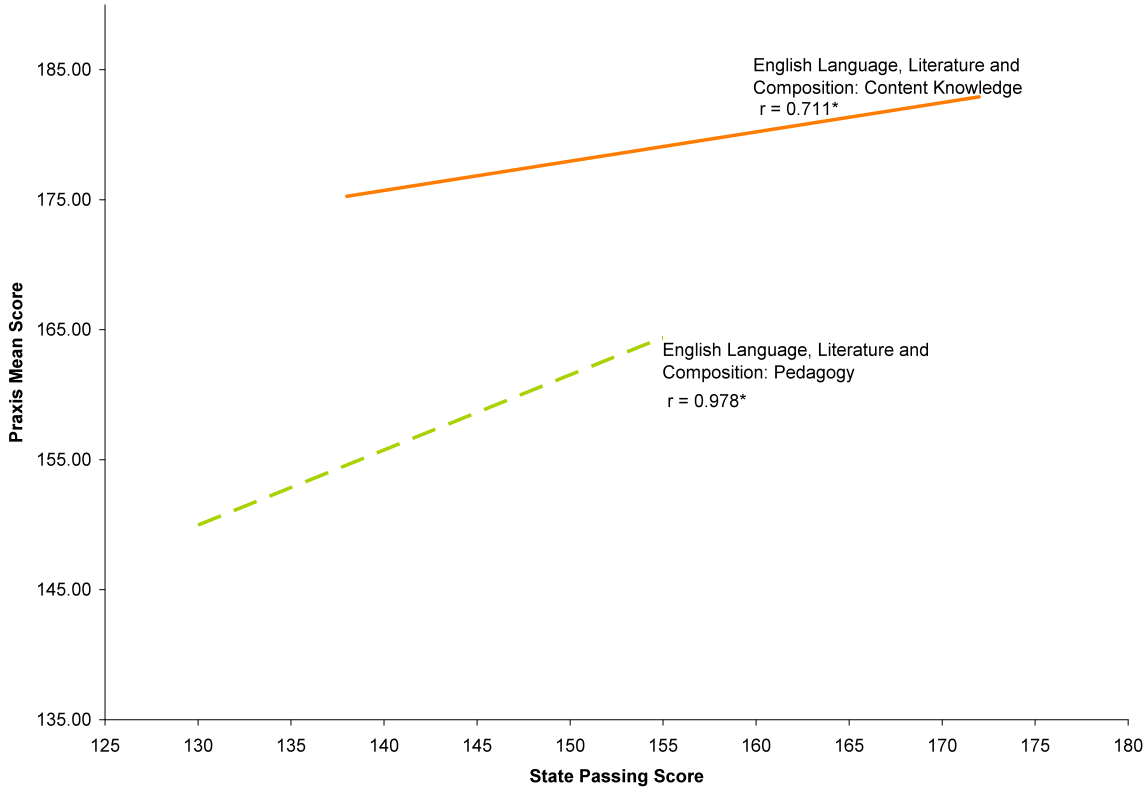


Exhibit Reads: Mean scores for passers are higher in states with higher passing standards.

Exhibit 60
Scatter Plots of Passing Standards vs. Passing Rate by State for Passers—English Language, Literature, and Composition

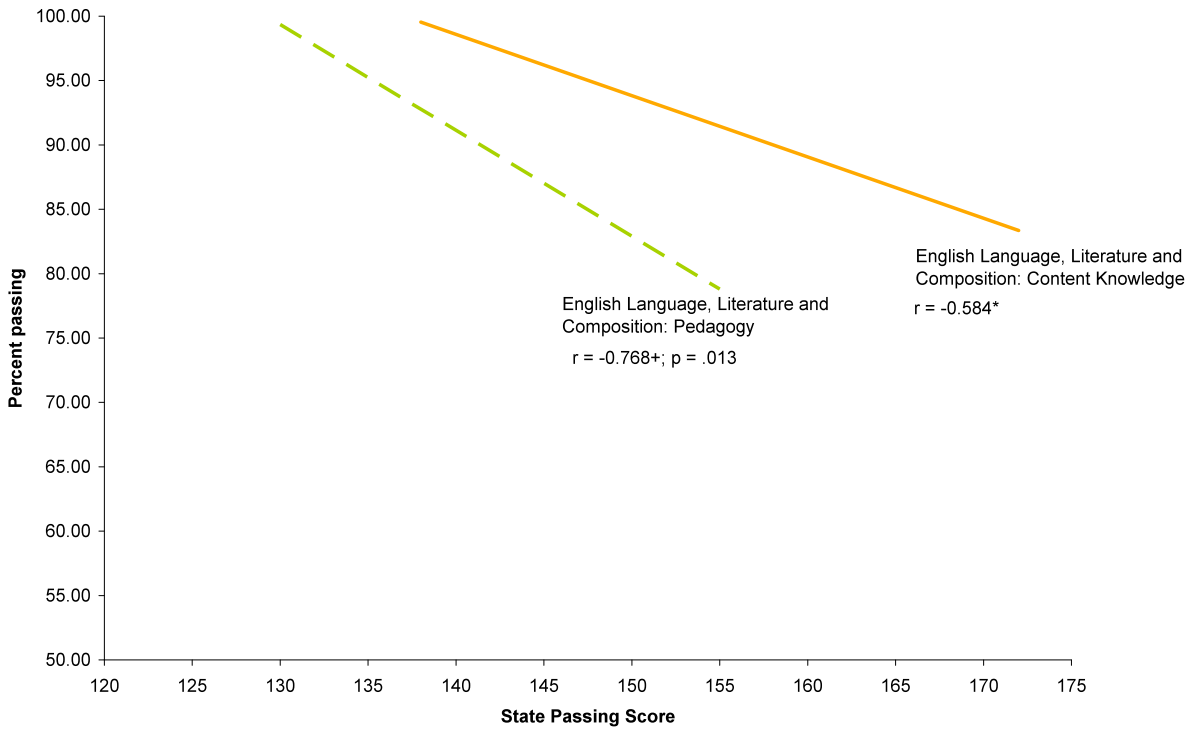


Exhibit Reads: The proportion of candidates who pass is lower in states with higher passing standards.

For the three mathematics tests (Math: Content Knowledge, Math: Pedagogy and Middle School Math: Content Knowledge), there is a strong positive relationship between state passing standards and mean score for passers (see Exhibit 61). Passing rates are lower with increased passing standards for the middle school test but no such trend is apparent for the Math: Content test (see Exhibit 62).

Exhibit 61
Scatter Plots of Passing Standards vs. Mean Score by State for Passers—Mathematics

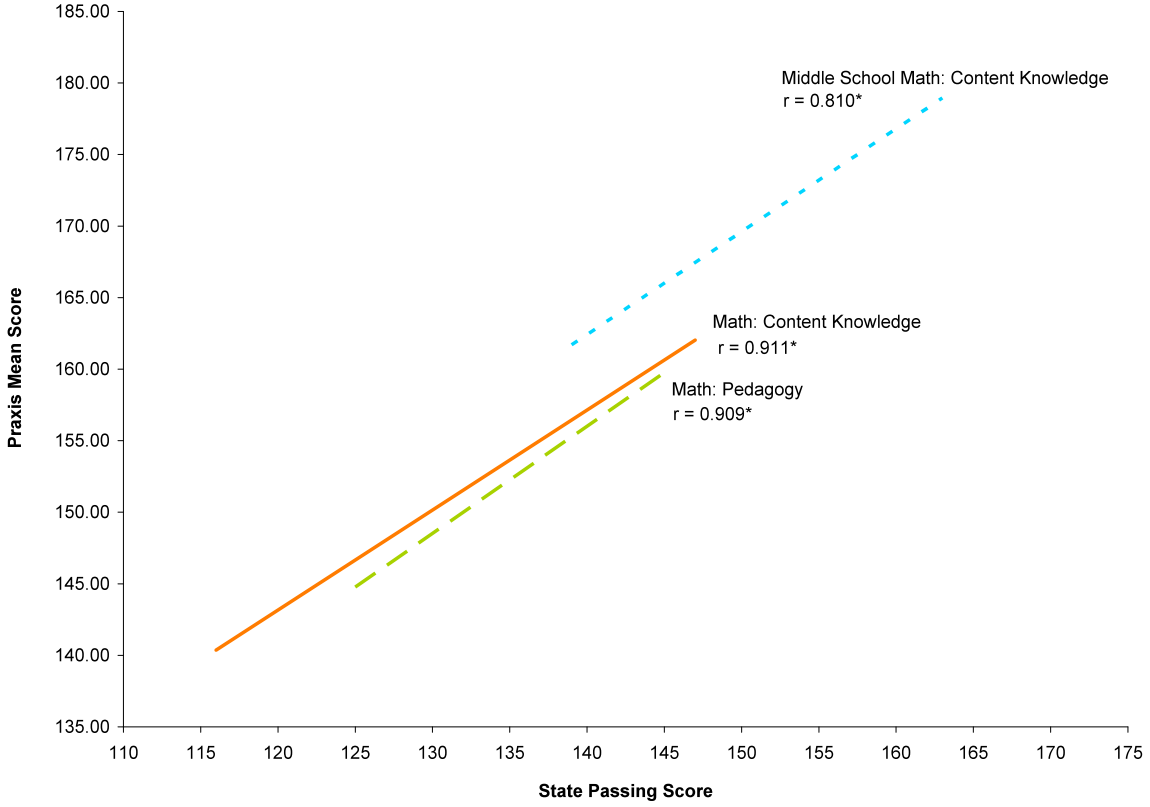


Exhibit Reads: Mean scores for passers are higher in states with higher passing standards.

Exhibit 62
Scatter Plots of Passing Standards vs. Passing Rate by State for Passers—Mathematics

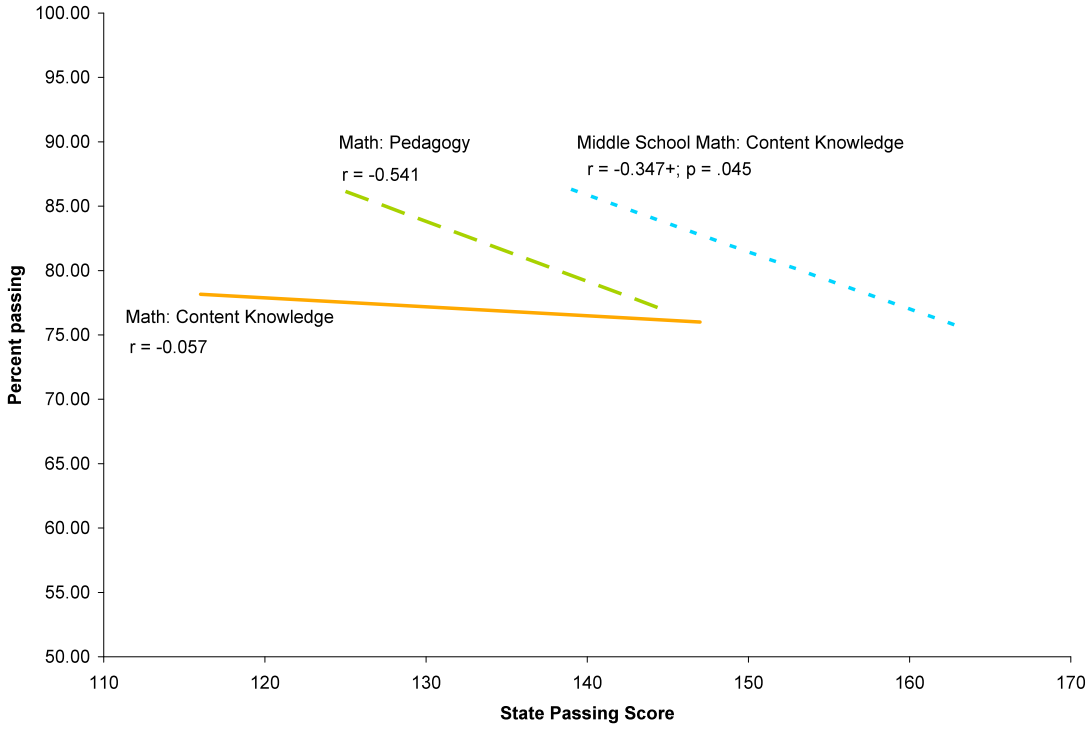


Exhibit Reads: Passing rates are lower for states with higher standards for the Middle School Mathematics test but no such relationship is observed for the Math: Content Knowledge test.

For the two science tests (Earth and Space Sciences: Content Knowledge and General Science, Part 1: Content Knowledge), there is strong positive relationship only for Earth and Space Science (see Exhibit 63). Passing rates are lower for states with higher passing standards for General Science only (see Exhibit 64).

Exhibit 63
Scatter Plots of Passing Standards vs. Mean Score by State for Passers—Sciences

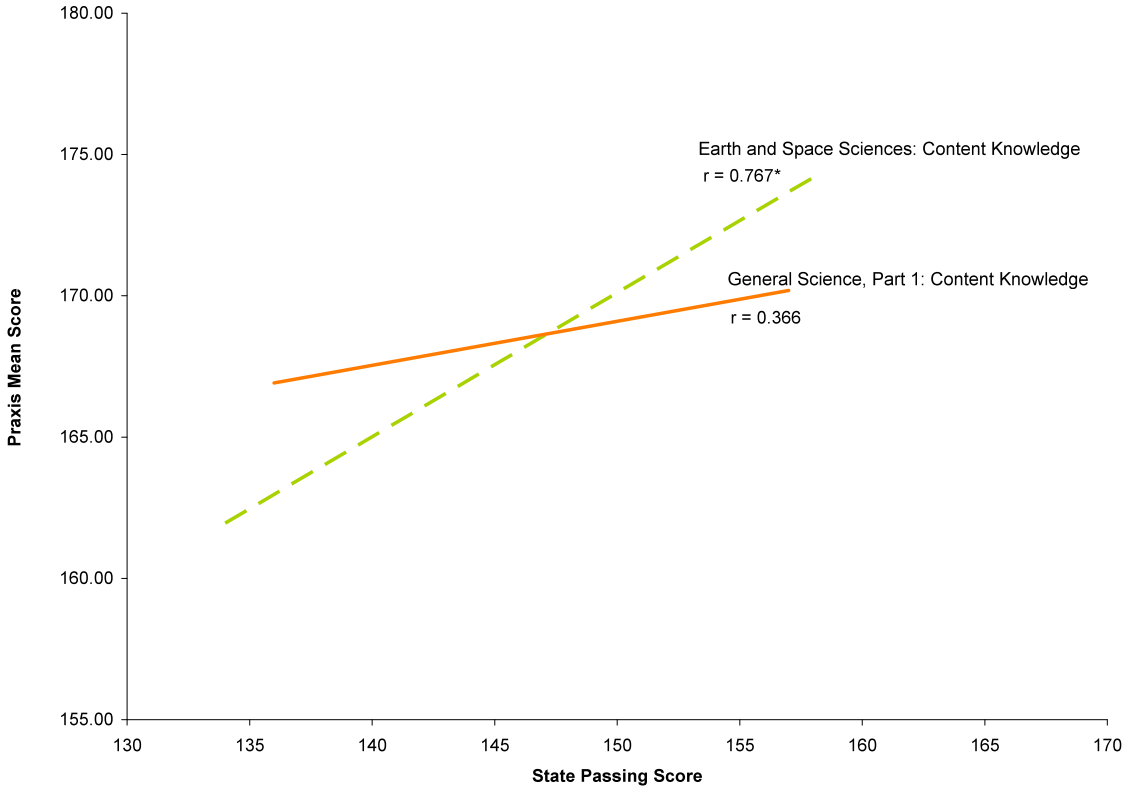


Exhibit Reads: Mean scores for passers are higher in states with higher passing standards for Earth and Space Sciences.

Exhibit 64
Scatter Plots of Passing Standards vs. Passing Rate by State for Passers—Sciences

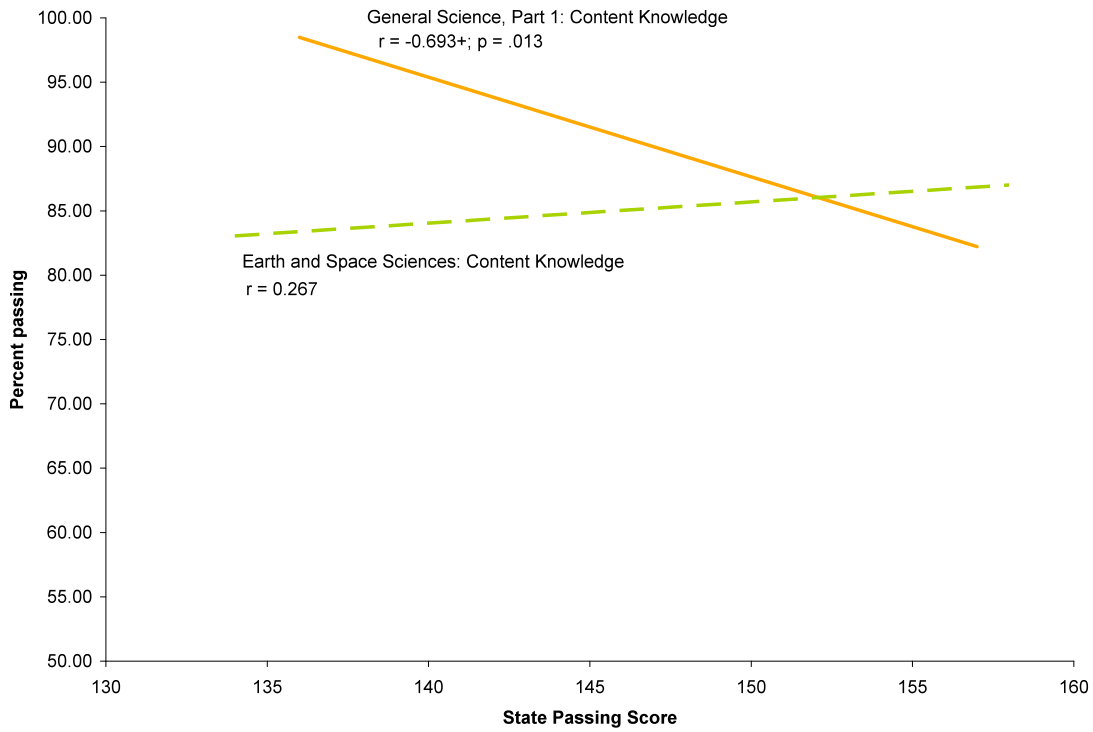


Exhibit Reads: Passing rates are lower in states with higher passing standards for General Science only.

Summary. For seven of the nine tests reviewed, individuals who pass the tests in states with higher passing scores have, on average, higher scores than for individuals in states with lower passing scores. For six of the nine tests included in the analysis, passing rates are lower when passing standards are higher. The tests that do not show such a pattern are in mathematics and science, tests that tend to have the lowest passing rates in general.

IV. SUMMARY

This study analyzed changes in teacher licensure scores during recent years. Nine *Praxis* tests were analyzed. Six of the tests focused on content knowledge and the others focused on pedagogy. The primary question was whether or not there were discernible trends in scores during the recent past, a period in which there has been a policy emphasis on improving the quality of the teaching force. *Praxis* tests selected for study included:

1. Elementary Education: Pedagogy (0012)
2. Elementary Education: Content Knowledge (0014)
3. English Language, Literature, and Composition: Content Knowledge (0041)
4. English Language, Literature, and Composition: Pedagogy (0043)
5. Mathematics: Content Knowledge (0061)
6. Mathematics: Pedagogy (0065)
7. Middle School Mathematics: Content Knowledge (0069)
8. General Science, Part 1: Content Knowledge (0431)
9. Earth and Space Sciences: Content Knowledge (0571)

The study appropriately focuses on individuals who pass the *Praxis* tests, for it is these individuals who will be allowed to teach in the public school system. However, we also examined the performance of all candidates, regardless of passing status, in order to provide a broader view of trends for all individuals who took *Praxis*.

FINDINGS

The adoption of tests by states varies widely. States adopt different tests at different times and apply quite different passing standards. This study attempted to make fair trend comparisons across years by including a stable set of states within each analysis and applying appropriate state- and year-specific passing standards for each test. Thus, different combinations of states and years are used in each of the analyses. Because states have been transitioning to new testing requirements, some analyses include only three years of data, while others include up to eight years.

Changes in mean test scores over years are extremely small. While the study identifies many significant trend effects, the magnitude of these effects is very low. Of course, because the samples include thousands of teachers, even small differences can be statistically significant. However, the practical import of any observed trends on student outcomes is likely to be minimal.

Those who pass have *Praxis* scores substantially higher than those who do not. This has two practical implications. First, licensure tests are filtering out individuals who attain very low scores on tests of content knowledge. Second, it is unlikely that many of these low-scoring individuals will achieve a passing score simply through taking the test multiple times without learning more of the content that is measured on the test. The difference between the mean passing score and state standards far exceeds any score variations that might be attributed to measurement error.

Interpretations based on those who pass differ from interpretations based on all candidates. When looking at just those who pass, there is either no trend or an increase for the majority of tests. However, when all candidates are included, the majority of tests show a decline (see Exhibit 65). This is due to the fact that, for many tests, the proportion of individuals who fail has increased over time. Because there is no prerequisite for taking a *Praxis* test and also because increasing standards in many states (see Appendix B) are increasing failure rates, the critical policy question should focus on the characteristics of those who pass the tests.

Trend patterns vary across tests. Trends vary across the nine tests included in this study. For those who pass, three show positive trends, two show declines, and the remaining four do not trend in any particular direction (see Exhibit 65). By subject area findings are:

> Elementary Education—No trend for content, Negative trend for pedagogy.
> English Language, Literature and Composition—No trend for content, No trend for pedagogy.
> Mathematics—Positive trend for content, Positive for pedagogy, No trend for middle school mathematics.
> Science—Positive for general science content, Negative for earth and space sciences content.

Exhibit 65
Study Summary of Trends in Scores Across Testing Years Analyzed

Praxis Test	Number Years	Simple Passers	Simple All	Comp Passers	Comp All
0012 Elementary Education: Pedagogy	6	Negative	Negative	Negative	Negative
0014 Elementary Education: Content Knowledge	3	No Trend	Negative	No Trend	No Trend
0041 English Language, Literature, and Composition: Content Knowledge	8	No Trend	Negative	No Trend	Negative
0043 English Language, Literature, and Composition: Pedagogy	8	No Trend	No Trend	No Trend	Positive
0061 Mathematics: Content Knowledge	8	Positive	Negative	Positive	Negative
0065 Mathematics: Pedagogy	8	Positive	Positive	Positive	Positive
0069 Middle School Mathematics: Content Knowledge	3	No Trend	Negative	No Trend	Negative
0431 General Science, Part 1: Content Knowledge	8	Positive	No Trend	Positive	Positive
0571 Earth and Space Sciences: Content Knowledge	5	Negative	Negative	Negative	Negative

NOTE**:** The simple model examines mean score changes over year. The comprehensive model includes gender, race, and experience variables.

Exhibit Reads: **Tests vary in the number of years included in the analysis and in trends over years. Conclusions based on passers are substantially different than if all candidates are included.**

Exhibit 66
Study Summary of Trends by Gender and Experience (Passers)

Test	Gender	Experience
0012 Elementary Education: Pedagogy	F	I
0014 Elementary Education: Content Knowledge	M	E
0041 English Language, Literature, and Composition: Content Knowledge	M	E
0043 English Language, Literature, and Composition: Pedagogy	F	E
0061 Mathematics: Content Knowledge	M	I
0065 Mathematics: Pedagogy	0	E
0069 Middle School Mathematics: Content Knowledge	M	I
0431 General Science, Part 1: Content Knowledge	M	E
0571 Earth and Space Sciences: Content Knowledge	M	0

(F=females score higher; M=males score higher; E=experienced score higher, I=inexperienced score higher)

Exhibit Reads: **Gender and experience differences vary across the different tests studied. Differences tend to be relatively small.**

There has been an increase in the number of individuals taking the *Praxis II* tests. For many of the tests included in the analyses, the size of the sample of test-takers nearly doubled over the different range of years included. At the same time that a substantial increase in test-takers was occurring, there was very little movement in scores. There is no clear reason for why this pattern of data exists. It appears that the increase in candidates has not materially affected the overall preparation of individuals to succeed on the respective tests.

Race and ethnicity differences are consistent across tests. Across all tests, white candidates who pass have higher mean scores than African-American candidates.

Gender differences vary by test. Female candidates who pass have higher mean scores than male candidates on two of the three pedagogy tests included in the analysis. Male candidates who pass have higher mean scores on the six content knowledge tests analyzed (see Exhibit 66). Gender differences tend to be relatively small.

Experience effects vary across tests. Test-takers who pass and have prior teaching experience have slightly higher scores than those without experience for five tests. Those without experience have higher scores for both mathematics content tests (see Exhibit 66). Experience differences tend to be relatively small.

Passing standards are related to mean passing scores and passing rates in predictable ways. Given relatively sparse data, there does appear to be a general pattern that states with higher passing standards have candidates who have higher mean scores. Passing rates are typically lower as well, though there are notable exceptions in mathematics and science.

V. CONCLUSIONS

This study posed the question of whether teacher licensure scores have changed across recent years. The fact that *Praxis* tests are administered across multiple states allowed us to explore this issue in a relatively thorough fashion. The overarching conclusion is that scores during the periods analyzed have not moved substantially in either direction. There are a number of findings and study limitations worth further discussion.

Perhaps the most striking finding in this study is the very large disparity in scores between those who pass and those who do not. This is not the case of having a large number of candidates hovering near a passing standard such that if people simply took a test one or two more times, their status could change to pass. Instead the *Praxis* tests appear to serve the function of filtering out individuals with extremely poor test performance.

Aside from this score disparity, it is also true that passing rates for many of the tests have decreased over time. This might be due to increases in passing scores established by some states during the course of the study. However, that may not be the only reason and should be explored further. There is a need to understand who is taking the tests, including the nature of their preparation and academic qualifications. Further work needs to explore the extent to which failure rate changes can be explained by the increase in passing scores.

Certainly, there has been a systematic increase in the number of individuals taking these tests over the time periods studied, for experienced and inexperienced individuals alike. If this were simply a matter of more unqualified individuals taking the tests, we would have expected to see consistent decreases in scores for test-takers. This is no evidence to support any systematic decreases for candidates regardless of experience level. Thus, it might be useful to begin to gain a better understanding, over time, of who is taking teacher licensure tests, why they are taking them, and the extent to which they enter the teaching force.

The study makes clear why it is so important to clarify whether the sample is comprised only of passers or of all candidates. Different inferences will be made about trends depending on which sample is used. We contend that looking at passers is the most policy-relevant focus, because these are the individuals who are actually meeting the teaching qualifications established by states. The legitimate policy concern has to do with the qualifications of prospective teachers, not prospective test-takers.

The different inferences made as a function of the sample included in the analyses appear to be explained by Simpson's paradox. For a number of tests, we observe more positive trends for passers than for the whole sample. Yet, for most of these tests, we also see more positive trends for those who fail. This apparent contradiction is a classic example of Simpson's paradox. The increasing proportion of candidates who fail the test can mean that the overall means can decrease at the same time that the means for both those who pass and those who fail either increase or remain stable.

Experience factors are relatively small except for the Middle School Mathematics test. For this test, those new to the field of teaching score substantially higher than more experienced test-takers. This finding is consistent with that of Gitomer (2007), which showed that the middle school test was being taken by a large number of elementary trained teachers who were teaching in middle school and needed to be subject matter certified under the Highly Qualified Teacher

provision of *ESEA*. It is very possible that this group of teachers was not as well prepared in mathematics as newly prepared math teachers who are targeting middle school for a career path.

Gitomer (2007), primarily on the basis of SAT scores, concluded that the academic quality of teachers had improved a substantial amount over recent years. This begs the question of whether there are reasons to explain these more modest findings compared with the prior study. We believe there are a number of critical differences between the studies that might account for the somewhat different findings.

Perhaps the most important distinction concerns the cohorts being sampled. Gitomer (2007) compared *Praxis* candidates from 1994–97 to those from 2002–05, which actually contains data that vary by 11 years (1994 to 2005). In the present study, no data are included prior to 1999. It was during the late 1990s that concerted efforts were made to include testing as part of policies to address teacher quality issues. As examples, the *Higher Education Act* was reauthorized and included requirements for teacher education institutions to report their students' licensure test performance. The National Council for the Accreditation of Teacher Education (NCATE) developed new standards for accreditation that focused much more on student performance in teacher education programs. Other state and institutional efforts also were being put in place at this time.

Thus, if these policies had an effect, it is possible that they were more visible using a lens that begins in 1994. It is plausible that any changes brought about by this confluence of policies were well in place by 1999. Indeed, if we look at just the five tests in this study for which there are eight years of data, we obtain a more positive view than one which looks at all nine tests, including those using a briefer lens. Of the five tests with data going back to 1999, three show positive trends and two show no trend.

In preliminary analyses, we also observe that, when positive trends are observed, they are largely due to score increases in states that had increased their passing scores at some point during the time period studied. Higher passing scores are unambiguous indicators of more rigorous requirements being applied within a state. Less certain is whether moving to the new *Praxis* tests also reflects more rigorous requirements, although that was the clear intention when the new tests were designed. The improvement in SAT scores reported by Gitomer may be accounted for by the increased standards brought about by the adoption of *Praxis*. However, there is no empirical way to validate this speculation.

It is possible that more substantial changes would be evident if we could include the years 1994–98. Unfortunately, this is just speculative and there is no way of examining this hypothesis. However, given when the policy activity occurred, changing the window of examination may explain the variation in findings.

Of course, the measures themselves may be more or less sensitive to policy changes and instructional interventions. On the one hand, if greater emphasis is placed on entrance into teacher education, then we might expect an increase in college admissions scores, particularly if those scores are part of the admissions equation. If *Praxis* scores reflect an undergraduate curriculum that has stayed fairly consistent across years, we might expect less change in those scores. Again, this is speculative but creates hypotheses that might be worth further study.

Finally, we do not understand how the increased presence of alternate route candidates influences trends in test scores. It is possible that many of these students bring in relatively strong academic skills as measured by SAT scores. However, they may have a wide range course work as background which may or may not prepare them well for the *Praxis* tests. The current system of data collection does not capture the nature of teacher preparation routes, so this issue would need to be studied separately.

Taken together, for a limited set of *Praxis* tests across a number of recent years, we see little or no change in scores. Policies that gave significant attention to teacher licensure test performance, if they did have an impact as suggested by Gitomer (2007), may have already had their effect prior to the years studied in this analysis.

REFERENCES

Amendments to the Higher Education Act of 1965, Pub. L. No. 105-244, §. 112 Stat 1581 (1998).

Blyth, C. R. (1972). On Simpson's Paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338), pp. 364–366.

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st century.* New York: Carnegie Corporation.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity* (NCES. No. OE-38001). Washington, D.C.: U. S. Department of Health, Education, and Welfare, Office of Education.

Elementary and Secondary Education Act, Pub. L. No. 107-110, §. 115 Stat 1425 (2002).

Feistritzer, C. E. (2007). *Alternative teacher certification: A state-by-state analysis 2007.* Washington, D.C.: National Center for Education Information.

Gitomer, D. H. (2007). *Teacher quality in a changing policy landscape: Improvements in the teacher pool.* (ETS Policy Information Report). Princeton, N.J.: Educational Testing Service.

Gitomer, D. H., Latham, A. S., and Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing* (ETS Teaching and Learning Report Series, ETS RR-03-25). Princeton, N.J.: Educational Testing Service.

Greenwald, R., Hedges, L. V., and Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361–396.

Kerr, D. H. (1983). Teaching competence and teacher education in the United States. In L. S. Shulman and G. Sykes (eds.), *Handbook of teaching and policy* (pp. 126–149). New York: Longman.

Koerner, J. D. (1963). *The miseducation of American teacher*s. Boston: Houghton Mifflin.

Lagemann, E. C. (2000). *An elusive science: The troubling history of education research* (1st ed.). Chicago: University of Chicago Press.

National Council for Accreditation of Teacher Education. (2006). *Professional standards for the accreditation of schools, colleges, and departments of education*. Retrieved Sept. 29, 2008, from http://www.ncate.org/documents/standards/NCATE%20Standards%202008.pdf.

National Educational Association (1895). *Report of the Committee of Fifteen on elementary education: With the reports of the sub-committees; on the training of teachers; on the correlation of studies in elementary education; on the organization of city school systems.* New York: American Book Company.

State of Pennsylvania. (2007). 22 Pa. Code § 354.

Teacher Education Accreditation Council. (n.d.). *TEAC accreditation process overview*. Retrieved Sept. 29, 2008, from http://teac.org/accreditation/index.asp.

APPENDIX A
INITIAL YEAR OF *PRAXIS* TEST USE BY STATE FOR EACH ASSESSMENT

Exhibit A.1
Initial Year of Test-taking Sample Stability With Established Cut Score—Elementary Education: Content Area Exercises, Pedagogy (0012)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CT				X								
DC						X						
GA				X								
HI		X										
MD							X					
NV	X											
RI											X	
SC						X						
Total (Cumulative)	1	2	2	4	4	6	7	7	7	7	8	8

Exhibit A.2
Initial Year of Test-taking Sample Stability With Established Cut Score—
Elementary Education Content Knowledge (0014)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											X		
AL											X		
CO										X			
DC												X	
DE													X
ID											X		
KY											X		
LA								X					
MD							X						
ME												X	
MN									X				
MS								X					
NH												X	
NJ							X						
OH											X		
RI											X		
SD												X	
TN										X			
TX							X						
UT												X	
VA									X				
VT								X					
WA										X			
WY											X		
Total (Cumulative)							3	6	8	11	18	23	24

Exhibit A.3
Initial Year of Test-taking Sample Stability With Established Cut Score—
English Language, Literature, and Composition: Content Knowledge (0041)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							X						
AL											X		
AR						X							
CO										X			
CT	X												
DC	X												
DE											X		
GA				X									
HI		X											
ID											X		
IN						X							
KS												X	
KY			X										
LA						X							
MD						X							
ME												X	
MN									X				
MO					X								
MS							X						
ND													X
NH						X							
NJ					X								
NV									X				
OH						X							
OR		X											
PA	X												
SC						X							
SD												X	
TN		X											
UT												X	
VA						X							
VT								X					
WA										X			
WI											X		
WV			X										
Total (Cumulative)	3	6	8	9	11	19	21	22	24	26	30	34	35

Exhibit A.4
Initial Year of Test-taking Sample Stability With Established Cut Score—
English Language, Literature, and Composition: Pedagogy (0043)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AR						X							
DC						X							
HI		X											
LA						X							
MD						X							
NV	X												
TN						X							
Total (Cumulative)	1	2	2	2	2	7	7	7	7	7	7	7	7

Exhibit A.5
Initial Year of Test-taking Sample Stability With Established Cut Score—
Mathematics: Content Knowledge (0061)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							X						
AL											X		
AR						X							
CO										X			
CT				X									
DC	X												
DE											X		
GA				X									
HI				X									
ID											X		
IN								X					
KS												X	
KY		X											
LA										X			
MD						X							
ME												X	
MN									X				
MO					X								
MS						X							
ND													X
NH						X							
NJ					X								
OH						X							
OR		X											
PA				X									
SC								X					
SD												X	
TN		X											
UT												X	
VA						X							
VT								X					
WA										X			
WI											X		
WV			X										
Total (Cumulative)	1	4	5	9	11	17	18	21	22	25	29	33	34

Exhibit A.6
Initial Year of Test-taking Sample Stability With Established Cut Score—
Mathematics: Pedagogy (0065)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AR						X							
DC						X							
HI				X									
MD						X							
NV					X								
TN						X							
Total (Cumulative)				1	2	6	6	6	6	6	6	6	6

Exhibit A.7
Initial Year of Test-taking Sample Stability With Established Cut Score—
Middle School Mathematics: Content Knowledge (0069)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											X		
AL											X		
CT					X								
DE											X		
GA								X					
HI													X
ID													X
IN											X		
KS												X	
KY							X						
LA											X		
MD										X			
ME												X	
MN											X		
MO							X						
MS											X		
NC						X							
ND													X
NH							X						
NJ											X		
NV													X
OH						X							
OR						X							
PA										X			
RI													X
SC											X		
SD												X	
TN										X			
VA									X				
WA										X			
WV						X							
Total (Cumulative)					1	5	8	9	10	14	23	26	31

Exhibit A.8
Initial Year of Test-taking Sample Stability With Established Cut Score—
General Science, Part 1: Content Knowledge (0431)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							X						
DC	X												
HI		X											
NJ					X								
NV		X											
OR		X											
TN					X								
Total (Cumulative)	1	4	4	4	6	6	7	7	7	7	7	7	7

Exhibit A.9
Initial Year of Test-taking Sample Stability With Established Cut Score—
Earth and Space Sciences: Content Knowledge (0571)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											X		
AL											X		
AR									X				
CT				X									
DE											X		
ID											X		
IN								X					
KS												X	
KY										X			
MD						X							
MN											X		
MO					X								
NC										X			
ND													X
NH							X						
NJ					X								
OH						X							
PA						X							
SD												X	
TN					X								
UT													X
VA						X							
VT								X					
WA											X		
Total (Cumulative)				1	4	8	9	11	12	14	20	22	24

APPENDIX B
PASSING SCORE ESTABLISHED BY EACH STATE FOR EACH ASSESSMENT BY YEAR

Exhibit B.1
Praxis Passing Score by Test by Year—Elementary Education: Content Area Exercises, Pedagogy (0012)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CT				148	148	148	148	148	148	148	148	148
DC						148	148	148	148	148	148	148
GA				137	137	137	137	137	137	137	137	137
HI		135	135	135	135	135	135	135	135	135	135	135
LA						137	137	137	137	137		
MD							150	150	150	150	150	150
NV	135	135	135	135	135	135	135	135	135	135	135	135
RI											148	148
SC					145	145	145	145	145	145	145	145
TN								138				
UT												150

NOTE: For this exhibit **Bolded* cells** indicate cases in which a state raised the passing score of a test and *italicized** cells indicate cases when a state lowered the passing score of a test.

Exhibit B.2
Praxis Passing Score by Test by Year—Elementary Education Content Knowledge (0014)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											143	143	143
AL											137	137	137
CO										147	147	147	147
DC												145	145
DE													151
ID											143	143	143
KY											148	148	148
LA								147	147	147	**150***	150	150
MD							136	**142***	142	142	142	142	142
ME												145	145
MN									140	140	140	**145***	145
MS								153	153	153	153	153	153
NH												148	148
NJ							133	133	133	133	**141***	141	141
OH											143	143	143
PA								142					
RI											145	145	145
SD												137	137
TN										140	140	140	140
TX							151	151	151	151	151	151	151
UT												150	150
VA									143	143	143	143	143
VT								148	148	148	148	148	148
WA										141	141	141	141
WI											147	147	147

Exhibit B.3
Praxis Passing Score by Test by Year—English Language, Literature, and Composition: Content Knowledge (0041)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							158	158	158	158	158	158	158
AL											151	151	151
AR						159	159	159	159	159	159	159	159
CO										162	162	162	162
CT	172	172	172	172	172	172	172	172	172	172	172	172	172
DC	142	142	142	142	142	142	142	142	142	142	142	142	142
DE											159	159	**163***
FL			165	165	165	165	165	165					
GA				163	163	163	**168***	168	168	168	168	168	168
HI		164	164	164	164	164	164	164	164	164	164	164	164
ID											158	158	158
IN						153	153	153	153	153	153	153	153
KS												165	165
KY			138	138	138	138	**160***	160	160	160	160	160	160
LA						160	160	160	160	160	160	160	160
MD						164	164	164	164	164	164	164	164
ME												160	160
MN									148	148	148	**157***	157
MO					158	158	158	158	158	158	158	158	158
MS							157	157	157	157	157	157	157
ND													151
NH						164	164	164	164	164	164	164	164
NJ					155	155	155	155	155	155	**162***	162	162
NV									150	150	150	150	150
OH						167	167	167	167	167	167	167	167
OR		164	164	164	164	164	164	164	164	*159**	159	159	159
PA	153	153	153	153	153	**160***	160	160	160	160	160	160	160
SC						162	162	162	162	162	162	162	162
SD												154	154
TN		157	157	157	157	157	157	157	157	157	157	157	157
UT												168	168
VA						172	172	172	172	172	172	172	172
VT								172	172	172	172	172	172
WA										158	158	158	158
WI											160	160	160
WV			151	151	**155***	155	155	155	155	155	155	155	155

Exhibit B.4
***Praxis* Passing Score by Test by Year—English Language, Literature, and Composition: Pedagogy (0043)**

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AR						145	145	145	145	145	145	145	145
DC						150	150	150	150	150	150	150	150
HI		150	150	150	150	150	150	150	150	150	150	150	150
LA						130	130	130	130	130	130	130	130
MD						155	155	155	155	155	155	155	155
NV	155	155	155	155	155	*150**	150	150	*140**	140	140	140	140
TN						145	145	145	145	145	145	145	145

Exhibit B.5
Praxis Passing Score by Test by Year—Mathematics: Content Knowledge (0061)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							146	146	146	146	146	146	146
AL											118	118	118
AR						116	116	116	116	116	116	116	116
CO										156	156	156	156
CT				137	137	137	137	137	137	137	137	137	137
DC	141	141	141	141	141	141	141	141	141	141	141	141	141
DE											121	121	**141***
GA				124	124	124	**136***	136	136	136	136	136	136
HI				136	136	136	136	136	136	136	136	136	136
ID											119	119	119
IN								136	136	136	136	136	136
KS												137	137
KY		125	125	125	125	125	125	125	125	125	125	125	125
LA										125	125	125	125
MD						141	141	141	141	141	141	141	141
ME												126	126
MN									124	124	124	**125***	125
MO					137	137	137	137	137	137	137	137	137
MS						123	123	123	123	123	123	123	123
ND													139
NH						127	127	127	127	127	127	127	127
NJ					130	130	130	130	130	130	**137***	137	137
OH						139	139	139	139	139	139	139	139
OR		147	147	147	147	147	147	147	147	*138**	138	138	138
PA				127	127	**136***	136	136	136	136	136	136	136
SC								131	131	131	131	131	131
SD												124	124
TN		136	136	136	136	136	136	136	136	136	136	136	136
UT												138	138
VA						147	147	147	147	147	147	147	147
VT								141	141	141	141	141	141
WA										134	134	134	134
WI											135	135	135
WV			133	133	133	133	133	133	133	133	133	133	133

Exhibit B.6
Praxis Passing Score by Test by Year—Mathematics: Pedagogy (0065)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AR						135	135	135	135	135	135	135	135
DC						135	135	135	135	135	135	135	135
HI				135	135	135	135	135	135	135	135	135	135
MD						145	145	145	145	145	145	145	145
NV					135	135	135	135	135	135	135	135	135
OR		140	140	140	140	140	140	140	140				
TN						125	125	125	125	125	125	125	125

Exhibit B.7
Praxis Passing Score by Test by Year—Middle School Mathematics: Content Knowledge (0069)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											145	145	145
AL											139	139	**142***
CT					158	158	158	158	158	158	158	158	158
DE											148	148	148
GA								139	139	139	**145***	145	145
HI													143
ID													145
IN											156	156	156
KS												158	158
KY							143	143	143	143	143	**148***	148
LA											148	148	148
MD										152	152	152	152
ME												148	148
MN											152	152	152
MO							158	158	158	158	158	158	158
MS											140	140	140
NC						141	141	141	141	141	141	141	141
ND													148
NH							151	151	151	151	151	151	151
NJ											152	152	152
NV													139
OH									143	143	143	143	143
OR						163	163	163	163	*156**	156	156	156
PA										151	151	151	151
RI													158
SC											149	149	149
SD												139	139
TN										143	143	143	143
VA									163	163	163	163	163
WA										152	152	152	152
WV						148	148	148	148	148	148	148	148

Exhibit B.8
Praxis Passing Score by Test by Year—General Science, Part 1: Content Knowledge (0431)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK							155	155	155	155	155	155	155
DC	136	136	136	136	136	136	136	136	136	136	136	136	136
HI		157	157	157	157	157	157	157	157	157	157	157	*150**
NJ					148	148	148	148	148	148	**152***	152	152
NV		150	150	150	150	150	150	150	150	150	150	150	150
OH						155	155	155	155	155			
OR		152	152	152	152	152	152	152	152	*145**	145	145	145
TN					138	138	138	138	138	138	138	**145***	145

Exhibit B.9
Praxis Passing Score by Test by Year—Earth and Space Sciences: Content Knowledge (0571)

State	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
AK											144	144	144
AL											148	148	148
AR									145	145	145	145	145
CT				157	157	157	157	157	157	157	157	157	157
DE											150	150	150
ID											144	144	144
IN								150	150	150	150	150	150
KS												150	150
KY										145	145	145	145
MD						152	152	152	152	152	152	152	152
MN											149	149	149
MO					147	147	147	147	147	147	147	147	147
NC										136	136	136	136
ND													149
NH							148	148	148	148	148	148	148
NJ					134	134	134	134	134	134	**145***	145	145
OH						151	151	151	151	151	151	151	151
PA						157	157	157	157	157	157	157	157
SD												150	150
TN					144	144	144	144	**146***	146	146	146	146
UT													153
VA						156	156	156	156	156	156	156	156
VT								158	158	158	158	158	158
WA											150	150	150

APPENDIX C
PRAXIS REGISTRATION FORM—BACKGROUND INFORMATION

Exhibit C.1
Praxis Registration Form—Background Information
(Select one answer for each question below)

a. How do you describe yourself?
1. _ African-American or Black
2. _ Asian American/Asian (Ex.: Japanese, Chinese, Korean)
3. _ Southeast Asian American/Southeast Asian (Ex.: Cambodian, Hmong, Khmer, Laotian, Vietnamese)
4. _ Pacific Island American/Pacific Islander
5. _ Mexican, Mexican American, or Chicano
6. _ Puerto Rican
7. _ Other Hispanic, Latino, or Latin American
8. _ Native American, American Indian, or Alaskan Native
9. _ White
10. _ Other

b. What is your best language of communication?
1. _ English
2. _ Another language

c. What language(s) did you first learn as a child?
1. _ English only
2. _ English and another language
3. _ Another language only

d. What is the highest education level you have attained?
1. _ Freshman (first year)
2. _ Sophomore (second year)
3. _ Junior (third year)
4. _ Senior (fourth or final year)
5. _ Earned bachelor's degree
6. _ Earned bachelor's degree plus additional credits
7. _ Earned master's degree
8. _ Earned master's degree plus additional credits
9. _ Earned doctoral degree

e. How many years has it been since you attended college or graduate school?
1. _ Currently attending college or graduate school
2. _ Less than 1 year
3. _ 1–3 years
4. _ 4–6 years
5. _ 7–10 years

6. _ More than 10 years

f. What is your cumulative undergraduate grade point average to date (based on a system where 4.0 = A)?
1. _ 3.5–4.0
2. _ 3.0–3.49
3. _ 2.5–2.99
4. _ 2.0–2.49
5. _ 1.5–1.99
6. _ Below 1.5

g. Indicate the highest level of education completed by your father or male guardian.
1. _ Some high school or less
2. _ High school diploma
3. _ Some postsecondary education
4. _ Associate degree
5. _ Bachelor's degree
6. _ Some graduate or professional school
7. _ Graduate or professional degree
8. _ Unknown

h. Indicate the highest level of education completed by your mother or female guardian.
1. _ Some high school or less
2. _ High school diploma
3. _ Some postsecondary education
4. _ Associate degree
5. _ Bachelor's degree
6. _ Some graduate or professional school
7. _ Graduate or professional degree
8. _ Unknown

i. Is either of your parents in the education profession?
1. _ Yes
2. _ No

j. What was your most recent full-time occupation?
1. _ Student
2. _ Food service
3. _ Maintenance
4. _ Truck driver
5. _ Technician
6. _ Clerical/administrative support
7. _ Sales/retail
8. _ Managerial
9. _ Self-employed
10. _ School aide
11. _ Teacher
12. _ Professional/executive
13. _ Other

14. _ None

k. Are you or have you ever been enrolled in a teacher education program?
1. _ Currently
2. _ Formerly
3. _ Never

l. Your teaching status is:
1. _ Planning to enroll or currently enrolled in a teacher education program
2. _ Recently graduated and expect to begin teaching in the near future
3. _ 1 to 3 years teaching experience
4. _ More than 3 years teaching experience
5. _ Not planning to teach at this time

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access

www.ed.gov