Report
to
The LEGISLATIVE FINANCE COMMITTEE

Public Education Department
The Three-Tiered Licensure System and The Achievement Gap
June 4, 2009

Report #09-08

State of New Mexico
LEGISLATIVE FINANCE COMMITTEE

325 Don Gaspar, Suite 101 • Santa Fe, NM 87501
Phone: (505) 986-4550 • Fax: (505) 986-4545

David Abbey
Director

June 4, 2009

Dr. Veronica Garcia, Secretary
New Mexico Public Education Department
Jerry Apodaca Education Building
300 Don Gaspar
Santa Fe, NM 87501

Dear Dr. Garcia,

On behalf of the Legislative Finance Committee (Committee), I am pleased to transmit the *Program Evaluation of the Three-Tiered Licensure System and the Achievement Gap.*

The review team analyzed assessment data and evaluated the impact of the three-tiered licensure system on student performance. The evaluation also provides an update on the achievement gap. The report will be presented to the Committee on June 4, 2009. An exit conference was conducted on May 21, 2009 to discuss the contents of the report with your designated staff.

I believe this report addresses issues the Committee asked us to review and hope your department benefits from our efforts. We very much appreciate the cooperation and assistance we received from you and your staff.

Sincerely,

David Abbey, Director

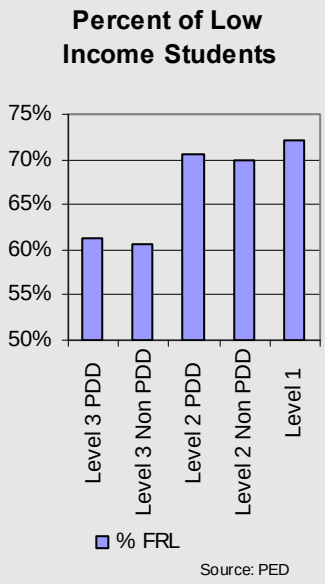
Table of Contents

Page No.

EXECUTIVE SUMMARY

Since 2004, the Legislature has invested over $278 million in teacher quality and compensation.

This evaluation analyzed data for 2,336 teachers and 40,199 students and is the first study connecting teachers to their students' growth on the NMSBA.



Beginning teachers are overrepresented in high poverty schools.

Teachers matter. Good teachers and quality instruction boost student performance and reduce the achievement gap between low income students and their peers. Education researchers demonstrated that effective instruction is 15-20 times more powerful than family background and income, race, gender, and other explanatory variables when using student growth models as opposed to a snapshot of student achievement. Recognizing the importance of teachers, the Legislature has invested over $278.4 million in appropriations from the general fund since 2004 to implement the three-tiered licensure system and boost pay.

The three-tiered system was designed to increase student achievement by recruiting and retaining high-quality teachers. In exchange for large increases in minimum salaries, teachers are expected to meet competencies and impact student achievement as demonstrated through annual evaluations and a professional development dossier (PDD). In 2007, the Legislative Finance Committee (LFC) conducted a joint evaluation of the three-tiered system with the Legislative Education Study Committee (LESC) and the Office of Education Accountability (OEA). That evaluation found the system had helped to recruit and retain teachers, but left whether the system was helping increase student achievement to future evaluations.

This LFC program evaluation sought to determine the effectiveness of the three-tiered licensure system on improving student performance as measured by student growth on the Standards Based Assessment (NMSBA) and to provide an update on the status of the achievement gap. In general, it is expected that teachers with advanced licensure, and thus higher pay, produce greater student growth on the NMSBA. The evaluation examined NMSBA student test results for reading and math of 4th, 5th and 6th grade teachers, by licensure level, for nearly all elementary schools in New Mexico and the amount of growth in student scores and proficiency levels that occurred between 2007 and 2008.

Overall, students with advanced licensed teachers (Level 3) who had gone through the PDD process generally outperformed their peers. Teachers at all licensure levels produced gains in student achievement, however, the differences in student achievement between licensure levels, in general, were not significant, especially in light of vastly different compensation, and achievement gains were insufficient to increase proficiency levels on a large scale. The results were similar when controlling for poverty. Beginning teachers were overrepresented in high poverty classrooms, and had more students entering their classes

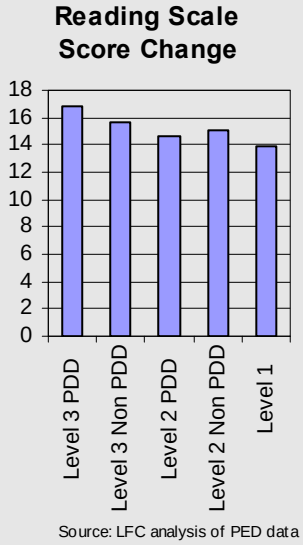
below proficient. Finally, the gap in achievement between low income students and their peers is larger than regularly reported and is persistently large regardless of race or ethnicity.

Key Findings.

The Effectiveness of the Three-Tiered Licensure System as Measured by Student Outcomes

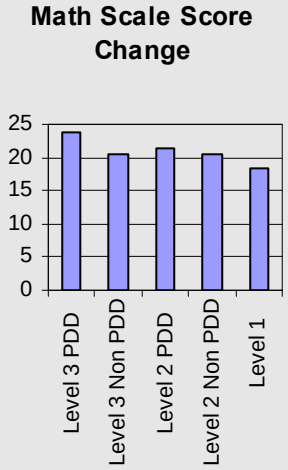
Growth in assessment scores can be used to identify effective teachers.

Growth In Assessment Scores Can Be Used To Identify Effective Teachers. The NMSBA is a vertically aligned assessment with scores scaled between 0 and 999 points. As a result, student test scores can be compared from year to year to determine "growth" from one grade level to next and whether the student's performance was sufficient to demonstrate proficiency or not. The evaluation uses growth in scores and proficiency level (beginning, nearing, proficient, advanced) as an appropriate proxy for effective instruction. Several elementary school principals, five from high performing schools and five from low performing schools, were interviewed and agreed that growth in assessment scores could serve as a suitable proxy for effective instruction.



Source: LFC analysis of PED data

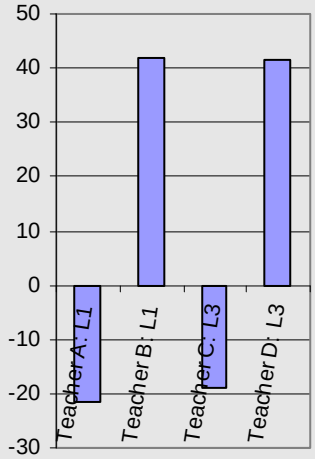
All groups of teachers produced average increases in student test scores, but gains were not large enough to see dramatic increases in the percentage of students that were proficient or above. In general, students in Level 3 PDD passer's classes experienced the largest gains in student achievement. Students in Level 1 beginning teachers' classes experienced smaller gains. For example, Level 3 PDD passers's student reading scores increased about 16.8 points versus Level 1 students growth of about 14 points. It is important to note, that in order for students to maintain proficiency level they ***must*** increase their test score, but to improve a proficiency level they must dramatically increase their score. For example, to maintain at "nearing proficient" a 4th grader must increase their reading score about 13 points higher on average than their 3rd grade score, but to move to "proficient" they must increase their score about 55 points on average. Level 3 PDD passers were the only group that had more students improve, rather than decline, in proficiency status for both math and reading. The difference in student achievement between these licensure levels (~2 points in reading) appears insufficient given the $20 thousand difference in minimum pay.



When classrooms are arranged into five groups of varying levels of classroom poverty, there are almost no statistically significant differences between licensure levels in terms of reading scale score changes and few statistically significant differences in terms of math score changes. Similar results occurred when analyzing teachers by grade level.

Research suggests that an effective way to improve the performance of the lowest achieving students is to improve the performance of the least effective teachers.

Differences in Effectiveness: Reading Score Change



Almost 1 in 10 respondents to an LFC survey expressed concerns about the occurrence of fraudulent PDD submissions.

Teachers at advanced licensure levels had more of the students who are likely to see larger score gains and improve their proficiency status. Students in Level 1 classrooms are more likely to start the year below proficiency. Students who are below proficient achieved much smaller gains in scores and were less likely to improve their proficiency status than students who entered a teachers' class at proficient or above. Also, low income students, as indicated by eligibility for the free/reduced lunch program, were more likely to be in beginning teachers' classrooms.

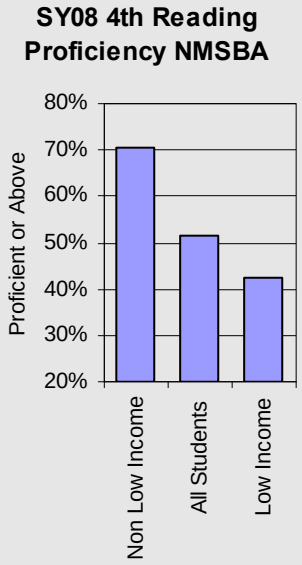
Variation Of Teacher Effectiveness

There Are Effective And Ineffective Teachers At Each Licensure Level Despite The Pay Being Similar At Each Licensure Level. The differences in teacher effectiveness between licensure levels were not substantial, but the differences in teacher effectiveness within each licensure level were immense. There is wide variation in terms of teacher effectiveness within a licensure level. Several teachers at all licensure levels produced negative average scale score changes and a few teachers at each licensure level achieved large scale score improvements. For example, of the 2,336 teachers in the analysis, thirty-two teachers increased, on average, their students scale scores in reading by over 35 points. Conversely, 177 teachers saw their students reading scores actually decrease on average.

Differences in teacher effectiveness have major consequences for students, both positive and negative. The few teachers that failed the PDD process had poor performance however many of the PDD passers demonstrated similar levels of poor student performance.

The LFC survey of teachers who went through the PDD process raises concerns about the integrity of the process and the occurrence of cheating. An online survey was conducted to receive input from teachers regarding the PDD process and received about 430 responses. Teachers had mixed opinions about the value of the PDD process with about half indicating that process does not identify highly effective teachers. One of the most concerning elements of the feedback received from teachers dealt with the integrity of PDD submissions and the general lack of internal controls to verify original work. Almost 10 percent of respondents provided unsolicited information citing concerns with the validity of some PDD submissions or the possibility of fraudulent submissions. PED has yet to provide educators and districts with guidance on what constitutes appropriate assistance on the PDD versus what kind of activity would result in an ethics violation.

PED guidelines do not require that student performance on assessments be central to the PDD submittal or the teacher evaluation process. Documenting student achievement is a minor factor in



The National Assessment of Educational Progress (NAEP) shows New Mexico at the bottom in performance comparisons to other states on overall achievement and closing the achievement gap, despite spending more per student.

advancement through the licensure system and the increasing salaries associated with advancement. The President and U.S. Secretary of Education Arne Duncan intend for the American Recovery and Reinvestment Act (ARRA) to be used as a means to promote new, reform-minded education plans, including performance pay. The keystone for many of these new reforms is quality instruction. Secretary Duncan has suggested that principal and teacher evaluations will have to take into account student performance in order for school districts to receive additional federal stimulus money.

The Academic Performance Of Students From Low Income Families

The federal No Child Left Behind Act of 2001 (NCLB) greatly increased the pressure on schools to close the "achievement gap" among various student groups. The achievement gap is the difference between the academic performance of students from low income families (economically disadvantaged) and wealthier students and between minority students and their non-minority peers. Closing the achievement gap requires accelerating the growth of minority and low income students at higher rates than the growth of their peers.

Until 2008, New Mexico did not disaggregate State student test data to compare low income versus non low income student scores, which masked the size of the true achievement gap. For example, in SY08 about 43 percent of low income 4th grade students were proficient or above on the NMSBA compared to about 71 percent of their non-economically disadvantaged peers scoring proficient and above – a 28 point gap. However, traditional achievement gap measures would have shown only an eight percent gap using the "all" students category results of 51 percent versus the economically disadvantaged student average of 43 percent proficient and above.

Generally, the achievement gap between low income students and their peers has not changed over the past four years, ranging from 22 to 28 percent in SY08. These differences are substantial, especially given the large numbers of low income students in the State. Results on the NMSBA show little growth in student achievement in fourth grade, but some positive improvement in eighth grade student achievement. Eighth grade reading is unique in improving performance and closing the achievement gap. However, the gap between non-economically disadvantaged and economically disadvantaged students remains high in most areas.

Socioeconomic status appears to have a more consistent impact on student achievement levels, regardless of race/ethnicity. The achievement gap between all low income students and their peers is larger than any socio-economic gaps within racial/ethnic groups. For

example, for all students the achievement gap is about 28 percent between low income students and their peers. The achievement gap between low income Hispanic students and their Hispanic peers was about 22 percentage points for 4th grade reading. Likewise, the achievement gap between Anglo (White) low income students and their Anglo peers was about 21 percentage points. Further analysis reveals that the overall gap in performance between all Anglo students and all Hispanic students is about 24 percentage points. However, after controlling for economic status the gap is much narrower among low income Anglos and low income Hispanic students at about 15 percentage points.

The low income student achievement gap starts high in 3rd grade and remains at a high level as students advance grade levels, using NMSBA data. This indicates that the achievement gap between non-economically and economically disadvantaged students is not only persistently large from year to year, and little progress has been made in closing the gap, but also remains high as students advance in grade levels. For example, the achievement gap was about 25 percent between low income students and their non low income peers in 2004 in reading, and by the time those students reached 6th grade the achievement was 26 percent.

"Performance pay is fairly new to education so there may not be a lot of studies showing that it boosts student achievement. But there's plenty of proof that it boosts worker productivity in other industries, so why not try it in schools?" - U.S. Sec. of Education Arne Duncan

Key Recommendations

PED should consider developing a bonus pay for performance pilot program and apply for competitive federal stimulus money for funding. PED should also work with LESC and LFC to design a pilot system that would provide bonuses or incentives to high performing teachers in general and even greater incentives to high performing teachers that relocate to high need schools.

PED should convene a workgroup to evaluate the following proposals:

- Require more evidence of student performance in PDD submissions and teacher evaluations.
- Establish goals for 'expected growth' on the NMSBA in grade levels and content areas.
- Give districts additional guidance about how to use student performance data to target professional development for teachers showing less than desirable performance.
- Continue to study the link between teacher effectiveness and student outcomes.

PED should ensure that principals review and authenticate that the PDD submittal is the teachers' actual work. PED should consider establishing penalties for false PDD submittals.

PED should establish clear guidelines for districts regarding the proper level of assistance teachers can receive when developing and submitting their PDDs.

PED and districts should ensure additional steps be taken for teachers who have poor student performance. Districts should direct principals to review and report on teacher performance and develop specific action plans to improve the results of the least effective teachers.

PED should update requirements for the Educational Plans for Student Success (EPSS) to incorporate the more accurate comparison of a school or district's achievement gap between low income students and their peers who are not from low income families.

OVERVIEW OF THE THREE-TIERED SYSTEM

In 2003, the Legislature passed legislation enacting comprehensive education reforms. The three-tiered licensure and evaluation system established new minimum salaries as a key component to these reforms. The three-tiered system has substantially increased teacher pay in New Mexico. According to a report issued by the National Education Association, New Mexico ranks third in the nation for the percentage increase in average teacher salaries since the 1997-98 school year (SY). Salaries for New Mexico public school teachers increased 49.6 percent from the 1997-98 SY to the 2007-08 SY. Nationally, salaries increased 32.9 percent in that same time frame. In the 1997-98 SY, New Mexico ranked 48th in the nation with an average salary of $30,152. In 2007-2008, New Mexico ranked 39th, with an average public school teacher salary of $45,112, according to the NEA report.

Fast Facts of the Three-tiered Licensure System and the PDD process

For the 2008-09 SY the State has:

- 703 level 0 teachers,
- 3,721 level 1 teachers,
- 10,504 level 2 teachers,
- 6,459 level 3 teachers.

The PDD process has three submittal periods per year: February, June, and November.

As of the November 2008 submission period there were 5,057 new submittals, and 717 resubmittals during twelve submission periods.

The PDD process takes about three or four months at cost to the teacher of $175.

The PDD process has a pass rate of 91.1% for teachers moving from level one to two and a rate of 90.2% for teachers moving from level two to three over the first twelve submission periods.

The three-tiered system requires teachers to submit a Professional Development Dossier (PDD) to advance in licensure level. The effective date for rules requiring teachers to submit PDD's for advancement was July 2004. Not all teachers at higher licensure levels submitted a PDD for advancement as over 2,700 teachers advanced to Level 3 between the effective date of HB 212 in April 2003 and the effective date of rules requiring teachers to submit PDD's for advancement in July 2004. The PDD is intended to provide sufficient evidence that a teacher is qualified to advance to a higher licensure level. The School Personnel Act of the Public School Code outlines the three-tiered system certification and compensation schedules which are contained in Sections 22-10A-7, 22-10A-10 and 22-10A-11.

The minimum salaries established in law were to be phased in over a five-year period as follows:
• Level 1, Provisional Teacher: $30,000 in SY 2003-2004;
• Level 2, Professional Teacher: $35,000 in SY 2004-2005 and $40,000 in SY 2005-2006; and
• Level 3-A, Master Teacher: $45,000 in SY 2006-2007 and $50,000 in SY 2007-2008.

The schedule of minimum salaries by level was fully implemented with SY 2007-2008 and the costs associated with the phased in salary schedule are shown in Table 1.

Table 1. Three-tiered Salary Increases and General Fund Appropriations

Year	Level One	Level Two	Level Three- A	General Fund Appropriation
2003-2004	$30,000	$30,000	$30,000	$5,700,000
2004-2005		$35,000	$35,000	$8,638,700
2005-2006		$40,000	$40,000	$51,800,000
2006-2007			$45,000	$6,841,300
2007-2008			$50,000	$9,118,600
Total Cost to Implement				**$82,098,600**

Source: LFC

Several other reports have been written about the Three-tiered System, including a Legislative Education Study Committee (LESC) memo in 2006 describing the extent to which the Three-tiered system requires documentation of student achievement. A 2007 joint evaluation by the Office of Educational Accountability (OEA), the LESC, and the LFC addressed recruitment and retention issues and suggested further study into the links between advanced licensure and student academic performance. The 2007 joint evaluation stated, "The most reasonable way to start such an examination is by looking at those teachers who have submitted their PDD for advancement to Level 2 or Level 3." This evaluation intends to serve as a preliminary examination of these relationships.

In determining a district's allocation from the State Equalization Guarantee (SEG), major impact comes from the training and experience (T&E) multiplier. This factor may increase the number of units generated by a district by as much as 30 percent based solely on teaching staff credentials and experience. High-poverty, rural districts, with the greatest needs, generally have the greatest difficulty hiring experienced teachers with advanced degrees that would increase T&E multipliers for these districts. For several years, the LFC has noted that, in its present form, the T&E index is not perfectly aligned to the three-tier system. The concern is whether the T&E index provides the appropriate amount of funding to districts given their teacher compensation expenses based on the three tier system. To the extent that the T&E index is not appropriately aligned with the three tier system, it has a considerable disequalizing effect on distributions.

THREE-TIERED SYSTEM AND IMPROVED STUDENT PERFORMANCE

In past discussions of student performance, the belief that achievement is determined mainly by factors beyond the control of a school has often been articulated. While some research suggests that family and community variables are noteworthy influences on student performance, those notions are less directly influenced by education policy makers when evaluating the costs and benefits of school district management decisions concerning programs, curriculum, faculty, and interventions. Research also clearly establishes that school level policies, particularly concerning teachers, can also have a significant impact on student performance. Several studies have shown that teachers matter tremendously and are perhaps the most important player in improving student performance. Researchers in Tennessee concluded that teacher effectiveness is "the single biggest factor influencing gains in achievement, an influence bigger than race, poverty, parents' education, or any of the other factors that are often thought to doom a child to failure."[1]

To assess the results of the Three-tiered Teacher Licensure System, it is important to understand the purposes of the system. One of the primary purposes of the system was to facilitate student

[1] W.L. Sanders and J.C. Rivers, Cumulative and Residual Effects of Teachers on Future Student Academic Achievement, University of Tennessee Value-Added Research and Assessment Center, 1996

success by recruiting and retaining qualified teachers. In the 2003 *Assessment and Accountability Act*, the legislative findings and purposes section begins and ends with an emphasis on student success. "The legislature finds that no education system can be sufficient for the education of all children unless it is founded on the sound principle that every child can learn and succeed, and the system must meet the needs of all children by recognizing that student success for every child is the fundamental goal. It is the purpose of this 2003 public school reform legislation to provide the framework to implement the legislative findings to ensure student success in New Mexico."[2] The legislation states "…the key to student success in New Mexico is to have a multicultural education system that attracts and retains quality and diverse teachers to teach New Mexico's multicultural student population." In addition, the LESC Ad Hoc Subcommittee for Education Reform stated in its final report "that the single, most important factor in improving student academic achievement is to ensure that a qualified, competent teacher is in every classroom…" As indicated in the 2007 joint OEA-LESC-LFC evaluation of the three-tiered licensure system, the public school reform legislation created the three-tiered system whose purpose, in addition to aligning with the federal NCLB requirements for "highly qualified teachers," was aimed at "increasing student achievement by recruiting and retaining high-quality teachers."

Given that one of the primary purposes of the three-tiered system is to ensure student success, it is appropriate to explore the connections between advanced licensure levels and increases in student performance. Arguably one of the most consistent and objective manners in which to evaluate this relationship is to study improvements in student achievement on standardized tests in reading and math. Therefore, this review focuses on outputs using the NMSBA test results as the measure of achievement. The method used is a simplified growth model in which the educational growth of individual students from entry to exit of a classroom is analyzed by teacher licensure level. The individual student serves as his or her own control since the baseline data is the students' performance level from the prior year. Using the students' performance upon classroom entry provides a reasonable starting point for assessing growth. The evaluation attempts to contribute to the literature on academic growth by providing a fair and constructive analysis of groups of teachers by licensure levels.

The three-tiered system is only one of many recent initiatives affecting teachers and students. Initiatives including New Mexico PreK, K-3 Plus, the School Improvement Framework, high school redesign, and increased graduation requirements have made, or soon will make, some impact on student achievement. Even so, it is possible to make certain associations between conditions in New Mexico and the three-tiered licensure system that shed light on the effectiveness of the system.

No Child Left Behind (NCLB) and the Achievement Gap

- The reporting requirements of NCLB require states to "produce results disaggregated by gender, major racial and ethnic groups, English proficiency, migrant status, disability, and status as economically advantaged. The assessment system must produce individual student interpretive, descriptive, and diagnostic reports. States must report itemized score analyses to districts and schools."
- The disaggregated results for these subgroups of the student population not only increase accountability for student performance, but they also allow comparisons between subgroups to be made and analyzed.

Source: NCLB Handbook

[2] 1978 NMSA Section 22-1-1.2

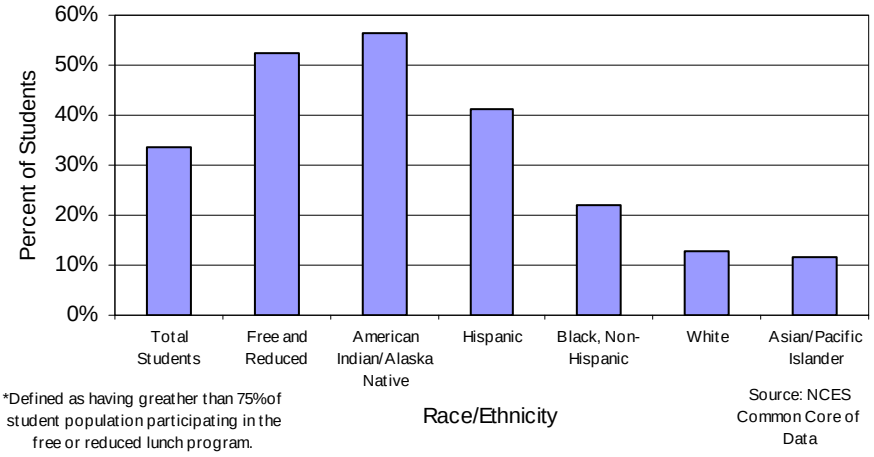
ACHIEVEMENT GAP

The achievement gap is the discrepancy in academic performance, as measured by assessment scores, between historically underperforming subgroups of students and their higher performing counterparts. In the past, historically underperforming groups have often been discussed in terms of ethnicity, and, to a lesser extent, socioeconomic status. The gap in student achievement between socioeconomic and ethnic subgroups has been the subject of past reports. The OEA and the LESC have published numerous reports that discuss the relationship between policy proposals and the achievement gap, but very little has been published on achievement gaps in recent years. The purpose of this update is to provide more timely information regarding the status of the achievement gaps, particularly since the federal No Child Left Behind Act (NCLB) reporting requirements allow for greater detail of data on socioeconomic status of students.

A high percentage of New Mexico children live in low income households, putting them at higher risk for academic difficulty. About 25 percent of New Mexico children live in poverty. Native Americans and Hispanics make up a disproportionate share of children under the age of 18 living in poverty.

Over 60 percent of New Mexico students are from low income families, using the broader socioeconomic indicator of student participation in the Free and Reduced School Lunch Program. Students from low income families make up the largest block of students in New Mexico schools, when compared to other NCLB racial/ethnic sub-groups.

Percent of New Mexico Students in High Poverty Schools* by Socioeconomic Status and Race/Ethnicity, 2006-07

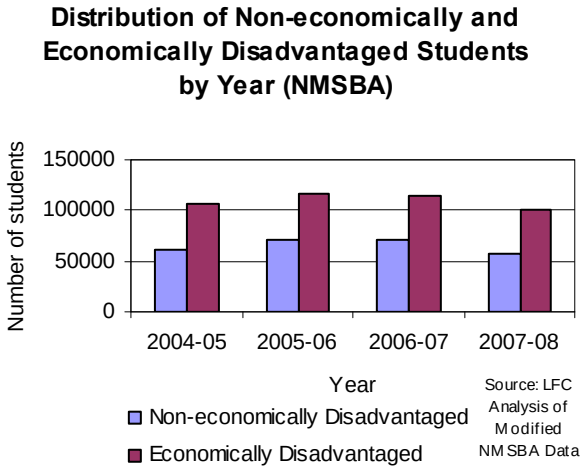


*Defined as having greather than 75%of student population participating in the free or reduced lunch program.

Source: NCES Common Core of Data

Over 33 percent of New Mexico students are concentrated in high poverty schools. High poverty schools are defined as having greater than 75 percent of students on free or reduced lunch. Not only are the subgroups of Native Americans and Hispanic students disproportionately represented among the population of children under 18 living in poverty compared to the total population, but also have a disproportionate share of students enrolled in schools where greater than 75 percent of the students are eligible for free and reduced lunch. The data show that a little over a third of the State's students attend a high-poverty school (as defined by this measure), and over half of the students eligible for free and reduced lunch attend a high poverty school. For example, although Native Americans constituted eleven percent of the overall student population, about 56 percent of Native American students attended a high poverty school.

In New Mexico, the persistent gap in achievement between economically disadvantaged students and non-economically disadvantaged students has grave implications because so many students are economically disadvantaged. Using the last four years of NMSBA data, the chart shows that the majority of students, nearly twice as many, are economically disadvantaged, and any efforts to raise the achievement of economically disadvantaged students will not only affect traditionally underperforming ethnic subgroups more, but these efforts will raise overall student performance as well.



Because traditionally underperforming ethnic groups make up disproportionately greater numbers of economically disadvantaged students, the focus for the achievement gap update is on the differences in student performance between non-economically disadvantaged and economically disadvantaged students as measured by eligibility in the National Free and Reduced Lunch program. Any policy efforts to close the achievement gap by raising the scores of economically disadvantaged students will inevitably impact Native Americans and Hispanics.

PROJECT INFORMATION

Program Evaluation Objective.

The purpose of the review was to evaluate the impact of three-tiered licensure system on student performance. The evaluation contained two primary objectives.

- Determine the effect of the three-tiered licensure system on improving student performance on the New Mexico Standards Based Assessment (NMSBA) scores. The objective sought to assess the extent to which teachers who have passed the Professional Development Dossier (PDD) process are successful at improving student performance. This objective also analyzed data regarding the distribution of teacher resources.
- Provide a detailed update on the status of the achievement gap in New Mexico. This objective focused on the performance of economically disadvantaged students compared to non-economically disadvantaged students.

Program Evaluation Activities (Scope and Methodology).

- Applicable laws and regulations, including appropriated funding.
- LFC file documents, including all available project documents.
- Available performance reviews from other states relative to this review.
- Information obtained from outside sources, including Internet searches.
- Extant literature studying the impacts of rigorous assessments on student achievement.
- Analysis of PED provided student test scores, broken down by teacher and student characteristics. See appendix for more detail on methodology.

Authority for Review. The LFC has the statutory authority under Section 2-5-3 NMSA 1978 to examine laws governing the finances and operations of departments, agencies and institutions of New Mexico and all of its political subdivisions, the effects of laws on the proper functioning of these governmental units and the policies and costs. The LFC is also authorized to make recommendations for change to the Legislature. In furtherance of its statutory responsibility, the LFC may conduct inquiries into specific transactions affecting the operating policies and cost of governmental units and their compliance with state law.

Program Evaluation Team.
Manu Patel, Deputy Director for Program Evaluation
Charles Sallee, Program Evaluation Manager
Craig Johnson, Program Evaluation, Project Lead
David Craig, Program Evaluator

Exit Conference. An exit conference with Deputy Secretary Don Moya, Assistant Secretary Sheila Hyde, and other Public Education Department staff was held on May 21, 2009.

Report Distribution. This report is intended for the information of the Office of the Governor, Department of Finance and Administration, Public Education Department, the Office of the State Auditor, and the Legislative Finance Committee. This restriction is not intended to limit distribution of this report which is a matter of public record.

Manu Patel
Deputy Director for Program Evaluation

THE EFFECTIVENESS OF THE THREE-TIERED LICENSURE SYSTEM AS MEASURED BY STUDENT OUTCOMES

Growth in assessment scores can be used to identify effective teachers. The three-tiered licensure system was not designed to be a true 'pay for performance' compensation system; however, improving student achievement was a key foundation for the development of the three-tiered system.

The New Mexico Standards Based Assessment (NMSBA) is the key accountability assessment used to determine whether or not a school has made adequate yearly progress (AYP). The NMSBA is administered to every student in grades 3 through 8 and 11 and establishes each student's level of performance using four levels of proficiency: Beginning Step, Nearing Proficiency, Proficient and Advanced. Each test taker receives a scale score in math and reading. The 'cut' score ranges shown in Table 11 in Appendix A establish the performance levels needed to achieve each level of proficiency. For example, the nearing proficient level in 3rd grade math includes scores ranging from 556 points to 610 points. Therefore, a 3rd grader earning a math score of 580 points would be classified as nearing proficiency. The NMSBA is a vertically aligned assessment which enables an analysis of growth over time. The NMSBA is designed so that a student's scale scores should increase each year as the student acquires greater knowledge and ability. For example, the nearing proficient level in 4th grade math includes scores ranging from 584 points to 635 points, an increase of about 25 points over 3rd grade scores. The cut score table (Table 11) indicates that the anticipated scale score change in math is larger than the anticipated scale score change in reading.

This evaluation measures growth in student achievement in two ways. First, the 2007 NMSBA scale score in math and reading serves as a starting point. This simple growth model attempts to control for non-school factors and focus on school factors by having an individual student serve as their own control by using their own NMSBA scores as a start and end point. The change in scale score from the 2007 NMSBA to the 2008 NMSBA in math and reading was calculated for each individual student. The change in scale score serves as a measure of academic growth that occurred during the year with a particular teacher. Individual scale score changes were averaged for each classroom and then averaged for each licensure level. Classroom average scale score changes for math and reading are shown for all licensure levels in the Classroom Performance Summary, Table 9, in Appendix A.

Second, student achievement is measured by analyzing the change, if any, in the students' proficiency status. For each individual student, their proficiency level in 2007 indicates the student's academic status as they enter a particular teacher's classroom. This level was compared to the proficiency level in 2008 which indicates the student's academic status as they left that teacher's classroom. For each individual teacher, percentages of students in their classrooms who improved in proficiency status, maintained proficiency status, and declined in proficiency status were calculated. These percentages were then averaged by licensure level. A performance score for each licensure level was calculated by subtracting the percent declining in proficiency status from the percent improving proficiency status. A positive performance score indicates more students improved their proficiency than declined. Detailed data showing change

in proficiency status for each licensure level is shown for math in Table 12 and for reading in Table 13 in Appendix A. Proficiency mobility and performance scores are also summarized in the Classroom Performance Summary in Table 9 in Appendix A.

While this three-tiered program evaluation is not designed primarily to evaluate individual teachers, several principals were asked if their best teachers are also the best at improving test scores. Essentially, they were asked if this narrowly defined method of using one year of growth in the NMSBA data would produce the same list of most effective teachers as a broader teacher evaluation method that might include other concepts such as creating good citizens, inspiring students to learn, knowledge in content areas, communicating with parents, principal observations, etc. Almost all of the principals (9 of 10) stated that the teachers with greatest success in NMSBA data would, in fact, be the best teachers overall. Many principals stated that if a teacher is effective at the broader measures of teacher effectiveness, this would translate into increases in NMSBA scores. Nearly all principals interviewed agreed that growth in assessment scores could serve as a suitable proxy to evaluate effective instruction.

All groups of teachers produced increases in average scale scores, but gains are not large enough to see dramatic increases in the percentage of students proficient. Teachers occasionally achieve scale score changes sufficient to maintain students at a given performance level but rarely enough to dramatically improve proficiency levels. An estimate of the average scale score change necessary to maintain proficiency levels from grade to grade was calculated using the cut scores which range from 0 to 999. The "change to maintain" estimate provides an informal benchmark that gives context when comparing growth in scale scores by licensure level.

Table 2: Change in Score Needed to Maintain Proficiency

Grade/Content	Estimated Change to Maintain	State Average Change
3rd to 4th Grade Reading	13.3	15.7
4th to 5th Grade Reading	12.7	17.9
5th to 6th Grade Reading	12.3	-1.4
3rd to 4th Grade Math	23.7	23.9
4th to 5th Grade Math	21.7	18.0
5th to 6th Grade Math	19.0	13.0

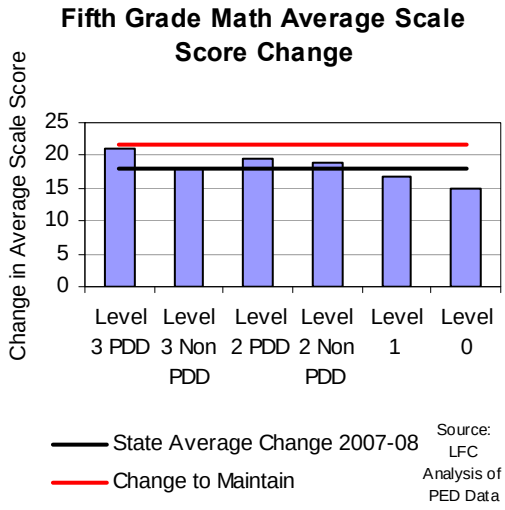
Source: LFC analysis of PED data

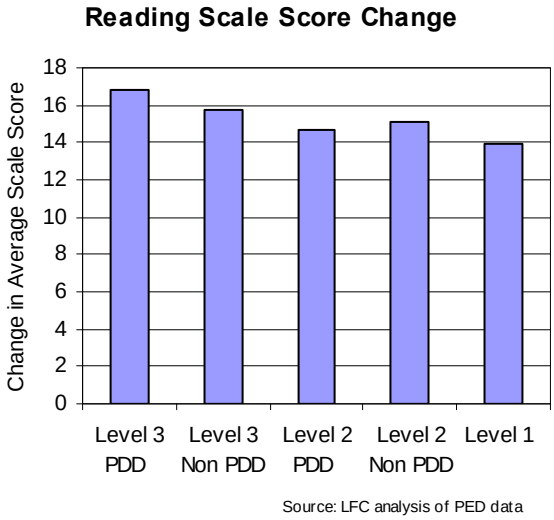
An elementary school teacher would need an average reading scale score change of about 13 points and an average math scale score change of about 22 points just to maintain students' proficiency levels. For example, to maintain proficiency in reading from 3rd grade to 4th grade, on average, an increase of 13.3 points would be needed; the state average change in scale score from 3rd grade in 2007 to 4th grade in 2008 was 15.7 points. The estimated changes needed to maintain proficiency levels are shown in Table 2.

Table 3: Score Change Needed to Move from Nearing to Proficient

MATH	
3rd to 4th	73.5
4th to 5th	67.0
5th to 6th	62.0
READING	
3rd to 4th	55.0
4th to 5th	56.0
5th to 6th	57.0

Source: NMSBA Technical Report





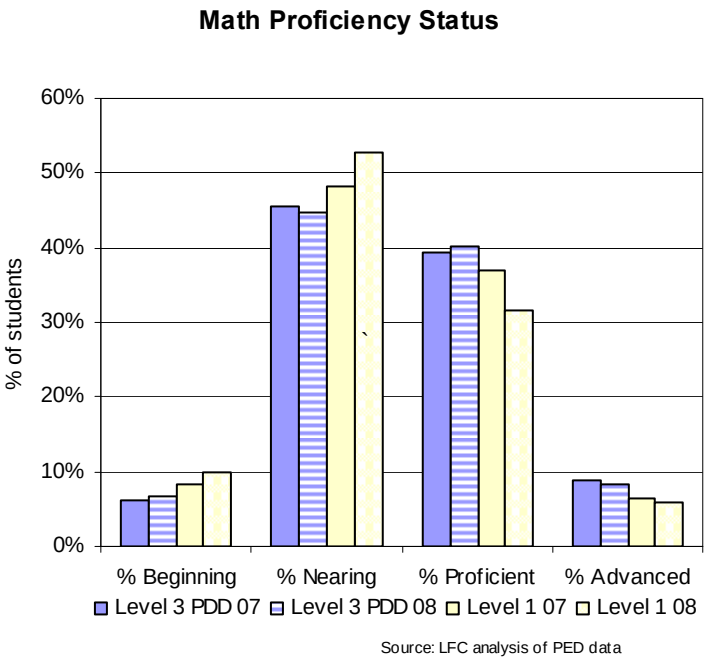
Much larger score changes are needed to improve the proficiency levels of the majority of students. The score changes needed to move most students forward is demonstrated by moving from the midrange of the 'Nearing Proficiency' level to the midrange of the 'Proficient' level in the following year. As indicated in Table 3, much larger scale score changes are needed to improve proficiency levels.

No group of teachers came close to reaching the very high levels of scale score changes necessary to see mass movement in students' proficiency levels, and many times groups of teachers were unable to reach the estimated change necessary just to maintain a student's proficiency level. As illustrated in the chart, for fifth grade math, no group of teachers were successful in reaching the average scale score necessary to maintain a students' proficiency level in this grade and subject area, although Level 3 PDD teachers came close.

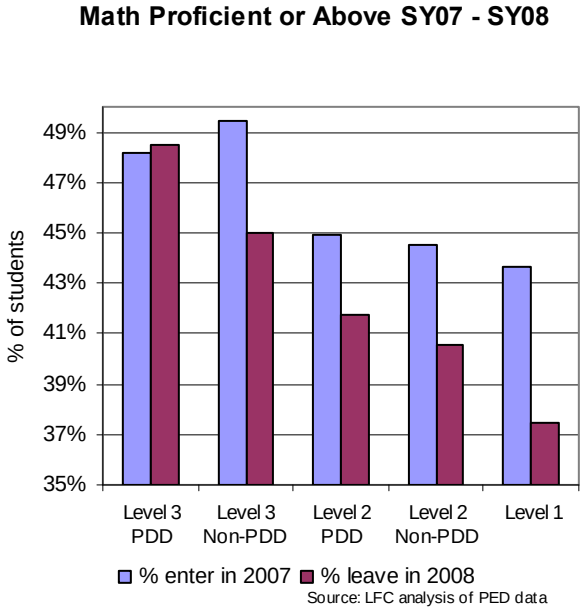
While Level 3 PDD teachers achieved the largest scale score increases for math and reading at all three grade levels, these average scale score changes are far short of the change needed to advance the majority of students in proficiency status.

Teachers with advanced licenses achieved the largest gains in student achievement, but these teachers also had more of the students who are likely to see larger gains in their classrooms. ***Students in Level 3 PDD passer classrooms experienced the largest average gains; Students in Level 1 classrooms experienced smaller gains.***

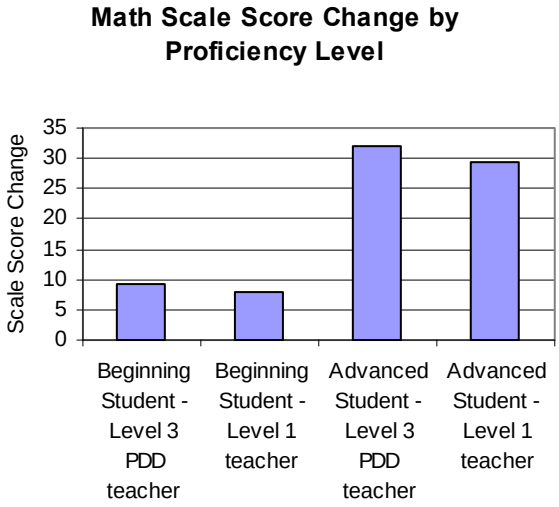
Level 3 PDD teachers had the largest gain in scale scores for math with an average gain of 23.7 points and reading with an average gain of 16.8 points. Of the licensure levels with more than 100 teachers, students of Level 1 teachers experienced the lowest growth in scale scores, in general, with an average math gain of 18.3 points and an average reading gain of 13.9 points. Level 3 PDD passers, Level 3 Non-PDD teachers, Level 2 PDD passers, and Level 2 Non-PDD teachers had more students improve their reading proficiency than decline. Level 0, Level 1, and Level 2 PDD failers had more students decline in reading proficiency than improve. Level 3 PDD passers were the only group to have more students improve their proficiency status in math than decline; therefore Level 3 PDD passers were the only group that was able to improve the proficiency status of more students in both math and reading.



Students in Level 1 classrooms are more likely to start the year below proficiency. Teachers with a Level 1 license have the highest percentage of students entering their class at the beginning or nearing proficient levels and the lowest percentage of students entering their class at the advanced levels. For example, the Math Proficiency Status graph illustrates that about 10 percent of Level 1 teachers' students enter the class at the beginning level, almost double the rate of Level 3 PDD passers. Level 3 teachers have the highest percentage of advanced students and lowest percentage of beginning students. The chart, Math Proficiency or Above SY07-SY08, shows the percentage of students who entered and left classrooms at proficient and above. Level 1 teachers struggled to keep their students who were proficient or above at those levels with about 44 percent of students entering at proficient or above and less than 38 percent leaving with the same designation. In the appendix, the Classroom Demographics Summary in Table 10 shows entering proficiency levels for all licensure levels.



Students who were below proficient achieved much smaller gains in scale scores and are less likely to improve their proficiency status than students who were proficient or above. There were substantial differences in the average scale score change depending on the student's proficiency status. The average reading scale score change for a student who was at the beginning step in 2008 was about zero, whereas the average reading scale score change for a student at the advanced level in 2008 was about 30 points. The chart shows the average math scale score change for beginning and advanced students in Level 1 and Level 3 PDD teacher classrooms. Similarly, for those students classified as nearing proficiency in math in 2008, more than 20 percent decreased in proficiency status from proficient in 2007 and only 7 percent improved in proficiency status from beginning in 2007. Of those students classified as proficient in math in 2008, about 22 percent improved in proficiency status from nearing proficient in 2007 and roughly 9 percent decreased in proficiency status from advanced in 2007. Detailed data on student performance by proficiency level is shown for math in Table 14 and for reading in Table 15 in Appendix A.



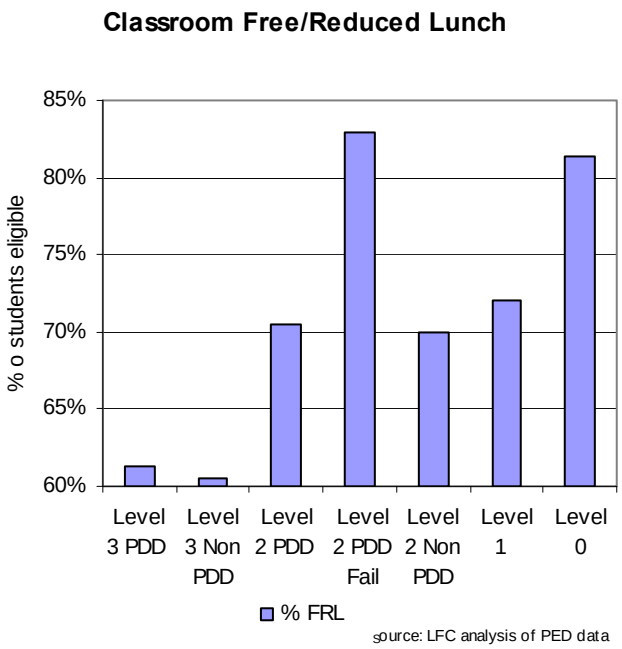
Students on free and reduced lunch are more likely to be in beginning teachers' classrooms. Eligibility for free and reduced lunch is a primary indicator of poverty and those schools with the largest proportion of these students have a greater percentage of Level 0-1 teachers and a lower percentage of Level 3 teachers as shown in Table 4. With over half of the schools statewide having 60 percent or more of their student population eligible for free and reduced (F/R) lunch, and another third of schools having 80 percent of the children eligible for free and reduced lunch, this distribution of teachers is a concern.

Table 4: 2009 Teacher Allocation by Free/Reduced Lunch

% F/R Students	% of Total Schools	Licensure Level 0 or 1	Licensure Level 2	Licensure Level 3	Totals
0%, <=40%	22.7%	19.4%	47.5%	33.1%	100%
>40%, <=60%	19.7%	17.0%	49.9%	33.1%	100%
>60%, <=80%	20.3%	21.3%	49.4%	29.3%	100%
>80%, <=100%	37.4%	23.8%	49.8%	26.4%	100%

Source: STARS data, 40th day 2009

There are more children on free reduced lunch in the typical Level 1 classroom as compared to the Level 3 classrooms. As shown in chart to the left, teachers with a Level 0 or Level 1 license had considerably higher percentages of poverty students in their classrooms; conversely, teachers with advanced licensure tend to have a lower percentage of their students in the free or reduced lunch program. This is consistent with national trends where teachers with less experience tend to work in schools with higher populations of low income students. The distribution of advanced licensed teachers is unfavorable to high poverty low performing schools in part because the three-tiered licensure system does not appropriately provide incentives for teachers to pursue more difficult teaching assignments. It is noteworthy that 20 of the 30 Level 0 teachers statewide were working in schools where over 80 percent of the students were eligible for free or reduced lunch.



When results are grouped by grade level, the differences in teacher performance between licensure levels are not substantial. Table 5 shows that Level 3 PDD teachers typically had the largest increases by grade level, however the differences are minimal.

Student performance shows a significant drop-off from 5th to 6th grade. The data suggest a significant decrease in students' performance as they move from fifth grade to sixth grade. Sixth grade performance shows the lowest average number of students scoring proficient or above in reading. The diminished performance of sixth grade students in reading and math is apparent and occurs independent of teacher licensure level. One of the most common explanations for the decline in sixth grade performance is that students are moving from a familiar elementary education setting into a middle school or entirely different school. Results from the analysis contradict this perspective and show that this decline occurs within elementary schools that include a sixth grade, and this decline in student performance of sixth graders may not be attributable to students changing schools.

There was more improvement in reading proficiency than math proficiency. Students were more likely to improve proficiency levels on average in reading than in math as indicated in Table 6. Although the number of students improving in reading is greater than math, fewer students score proficient or above as they progress through the elementary grades. Teacher level data supports this with the average percentage of students scoring proficient or above declining in all teacher levels in reading.

When results are grouped by socio-economic status, the differences in teacher performance between licensure levels are not substantial. Data was grouped by percent of classroom students eligible for free and reduced lunch in an attempt to compare licensure levels using demographically similar classrooms. Classrooms grouped by quintiles of students on free or reduced lunch were checked for statistically significant differences. A result is considered statistically significant if it is unlikely to have occurred by chance. A statistically significant difference simply means there is statistical evidence that there is a difference; it does not necessarily mean the difference is substantive or even large. The five main categories of teacher licensure levels, (Level 3 PDD, Level 3 Non-PDD, Level 2 PDD, Level 2 Non-PDD, and Level 1) allow for 10 direct comparisons within each quintile, for a total of 50 comparisons for math and 50 comparisons for reading. There are almost no statistically significant differences between licensure levels in terms of reading scale score changes. When classrooms are grouped by quintiles of free and reduced lunch, Level 3 PDD passers increases in math scale scores were occasionally statistically significant, particularly at the higher poverty schools. This further highlights the need to have the best teachers where they are needed most given that they were effective in the high poverty quintiles. See Tables 16 through 25 in Appendix A for all statistical significance tables.

Table 5: Scale Score change by Grade

	# teachers	# students	Change in Reading Scale Score	Change in Math Scale Score
4th Grade				
Level 3 PDD	45	769	18.5	28.2
Level 3 GF	222	3705	17.6	24.9
Level 2 PDD	141	2429	15.2	24.5
Level 2 GF	368	6180	17.0	23.9
Level 1	183	3139	15.3	21.3
Level 0	12	172	15.2	13.3
5th Grade				
Level 3 PDD	48	882	20.6	20.9
Level 3 GF	227	3981	17.9	17.8
Level 2 PDD	138	2422	17.8	19.4
Level 2 GF	374	6667	18.0	18.9
Level 1	178	2893	17.3	16.8
Level 0	13	231	17.1	14.8
6th Grade				
Level 3 PDD	20	359	4.0	20.2
Level 3 GF	70	1265	2.7	14.8
Level 2 PDD	42	729	2.7	17.1
Level 2 GF	155	2641	3.5	17.1
Level 1	70	1245	1.3	14.4
Level 0	5	98	5.9	19.4

Source: LFC analysis of PED data

Table 6. Improvement in Proficiency Rates

	Level 3 PDD	Level 3 Non-PDD	Level 2 PDD	Level 2 PDD Fail	Level 2 Non-PDD	Level 1	Level 0
N teachers	113	519	321	25	897	431	30
N students	2010	8951	5580	392	15488	7277	501
% Improved (Read)	18.6%	19.2%	19.0%	18.1%	19.3%	16.9%	16.1%
% Improved (Math)	18.0%	15.0%	15.6%	9.1%	14.7%	13.0%	10.8%

Source: LFC Analysis of PED Data

Recommendations

PED should study and design a bonus pay for performance pilot program and apply for competitive federal stimulus money for funding. PED should also work with OEA, LESC and LFC to design a pilot system that would provide bonuses or incentives to high performing teachers in general and even greater incentives to high performing teachers that relocate to high need schools. The system should target any bonus programs towards effective teachers, perhaps limited to Level 2 or Level 3, as measured by student performance. The return on the investment (increase in student achievement) in the three-tiered system could be greatly enhanced if policies were put in place to financially reward the most effective teachers for staying at or moving to low performing schools with high concentrations of low income students particularly in the early grades. See Appendix D for a discussion regarding the challenges associated with pay for performance systems.

PED should reevaluate licensure requirements based on licensure levels. For example, the requirement that principals qualify as a Level 3 teacher should be reevaluated to require demonstrated success in improving student achievement.

OEA may consider reviewing teachers that are clearly underperforming to see if proper evaluations have been made and peer interventions have occurred as required by statute. OEA is tasked annually by the General Appropriation Act (GAA) to verify that teachers are evaluated under the three-tiered licensure and evaluation system. OEA may need to conduct additional statewide reviews of teacher performance by licensure level to determine if the quality of evaluations is being maintained in accordance with expectations.

VARIATION OF TEACHER EFFECTIVENESS

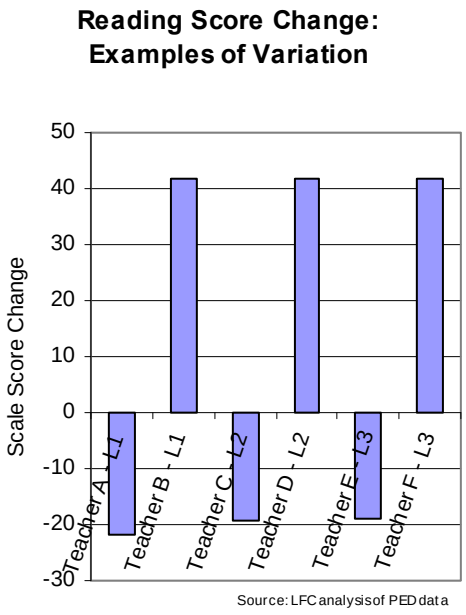
There are substantial differences in terms of student achievement gains amongst teachers and large gaps between the best and worst teachers at all licensure levels. There are effective and ineffective teachers at each licensure levels despite the pay being similar for teachers within each licensure level. Examples of teachers near the extremes in terms of performance are shown in Table 7 below.

Table 7: Teacher Effectiveness and Pay

Teacher	Licensure	Content area	Grade	Students	% Improved	% declined	Avg Score change	% FRL	Salary	Yrs Exp
A	Level 1	MATH	4	18	5.6%	33.3%	3.8	22.2%	$35,518	1
B	Level 1	MATH	4	16	43.8%	0.0%	53.1	68.8%	$31,500	1
C	Level 1	READING	4	23	13.0%	52.2%	-10.8	100.0%	$33,499	2
D	Level 1	READING	4	12	58.3%	0.0%	41.9	100.0%	$33,280	3
E	Level 2 NonPDD	MATH	5	15	0.0%	53.3%	-5.6	60.0%	$48,296	20
F	Level 2 NonPDD	MATH	5	12	50.0%	0.0%	42.4	100.0%	$46,592	16
G	Level 2 NonPDD	READING	5	16	6.3%	31.3%	-2.3	62.5%	$51,995	19
H	Level 2 NonPDD	READING	5	13	46.2%	0.0%	35.6	100.0%	$43,900	20
I	Level 2 PDD	MATH	4	18	6.3%	50.0%	-2.8	100.0%	$44,101	3
J	Level 2 PDD	MATH	4	13	76.9%	0.0%	60.0	100.0%	$43,056	4
K	Level 2 PDD	READING	4	19	5.3%	47.4%	-1.0	63.2%	$44,603	3
L	Level 2 PDD	READING	4	13	61.5%	0.0%	35.2	100.0%	$43,056	4
M	Level 3 NonPDD	MATH	4	23	4.3%	47.8%	1.9	82.6%	$59,738	36
N	Level 3 NonPDD	MATH	4	15	66.7%	0.0%	51.8	100.0%	$52,079	25
O	Level 3 NonPDD	READING	4	14	0.0%	57.1%	-18.9	21.4%	$56,050	15
P	Level 3 NonPDD	READING	4	19	52.6%	5.3%	39.2	52.6%	$50,056	9
Q	Level 3 PDD	MATH	5	14	7.1%	28.6%	3.4	64.3%	$56,000	13
R	Level 3 PDD	MATH	5	18	44.4%	5.6%	43.0	66.7%	$50,060	9
S	Level 3 PDD	READING	5	14	7.1%	50.0%	-4.1	64.3%	$56,000	13
T	Level 3 PDD	READING	5	21	38.1%	0.0%	31.5	42.9%	$50,000	7

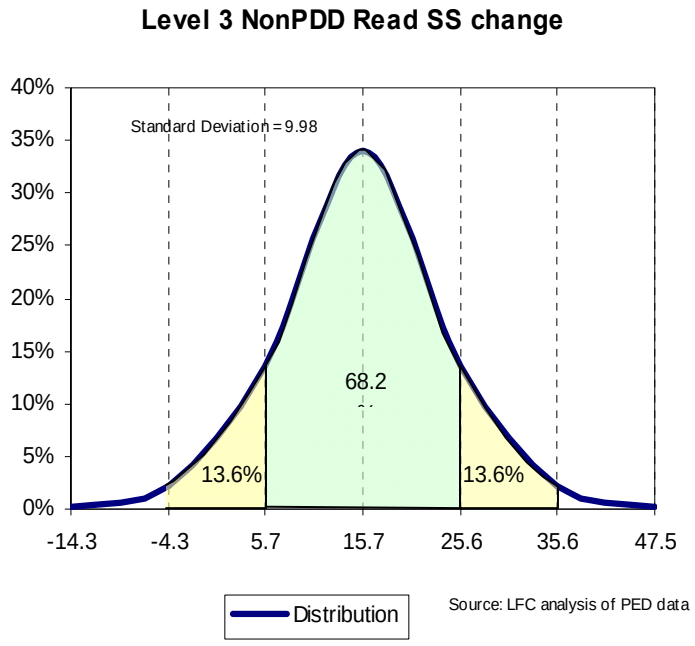
Source: LFC analysis of PED data

For example, of the 2,336 teachers in the analysis, thirty-two teachers (2-Level 1, 16-Level 2, and 14-Level 3) were able to increase, on average, their students scale scores in reading by over 35 points. Conversely, 177 teachers (40-Level 1, 98-Level 2, 39-Level 3) saw their students reading scale score decrease on average. Similarly, there were 100 teachers (11-Level 1, 58-Level 2, 31-Level 3) whose students increased their math scale scores by more than 40 points on average. Unfortunately, there were 48 teachers (14-Level 1, 20-Level 2, 14-Level 3) whose students average math scale score dropped.



Research suggests that an effective way to improve the performance of the lowest achieving students is to improve the performance of the least effective teachers.

Differences in teacher effectiveness have major consequences for students, both positive and negative. The chart on the left depicts average classroom reading score changes for individual teachers at Levels 1, 2, and 3. The differences between the lowest performing teachers and the best performing teachers impact changes in student test scores by a significant margin. Over several years, the cumulative difference in education quality between the lowest and highest performing teachers can significantly affect the extent to which the achievement gap is narrowed.



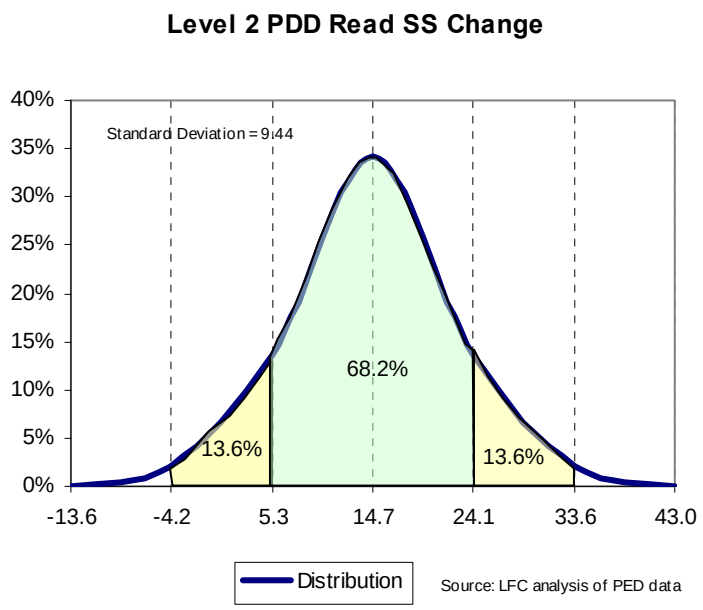
Some of the earliest and best analysis on teacher performance has been done in Tennessee, where researchers found that students with the most effective teachers for three years in a row performed 50 percentile points higher— on a 100-point scale—than comparable students assigned to the least effective teachers for three years in a row.[3] Other research documents the detrimental impacts of ineffective instruction. According to Eric Hanushek[4], "the estimated difference in annual achievement growth between having a good and having a bad teacher can be more than one grade-level equivalent in test performance."

All licensure levels had high standard deviations demonstrating similar variability in scale score changes. (See standard deviation data in Appendix A). The standard deviation is a measure of variability in classroom average scale score changes. The high standard deviations shown in Appendix A indicate that average classroom score changes were very spread out over a wide range of values. The scale score distribution charts depict the wide spread in scale score changes

[3] W.L. Sanders and J.C. Rivers, Cumulative and Residual Effects of Teachers on Future Student Academic Achievement, University of Tennessee Value-Added Research and Assessment Center, 1996.

[4] Hanushek, E. A. (1992). The Trade-Off between Child Quantity and Quality. *Journal of Political Economy, 100*(1): 84–117.

within a licensure level. For example, the average change in reading scale score for Level 2 PDD teachers was 14.7. The chart to the left shows that 13.6 percent of Level 2 PDD teachers achieved average scale score changes between 24.1 and 33.6 points. Likewise, 13.6 percent of Level 2 PDD teachers realized average scale score changes between -4.2 and 5.3 points. It is noteworthy that Level 3 Non-PDD teachers are the most expensive but are not the top performing group.



<u>The few teachers that failed the Professional Development Dossier (PDD) process had poor performance, however many of the PDD passers had similar levels of performance.</u>

The PDD process documents a teacher's own reasoning and practices in the classroom in line with the nine New Mexico teacher competencies. The PDD process appears to be more of a self-evaluation approach, rather than a student outcomes performance-based approach, to identifying quality teachers. The PDD process is less intensive than the National Board Certification (NBC) process, which takes over a year to complete at a cost to the teacher of $2,300 with a first time pass rate of 50 percent. The PDD process takes about three to four months to complete with a pass rate of over 91 percent. University of Washington and Urban Institute researchers found no evidence that even the more rigorous NBC process itself increases teacher effectiveness.

In general, those who fail the PDD were relatively ineffective at improving test scores, however passing the PDD does not guarantee a teacher's effectiveness. Level 2 teachers that failed the PDD had the least improvement in reading scale scores and the second worst performance in terms of improving math scale scores. This group of teachers was also the worst in terms of having more students improve in proficiency status instead of declining in proficiency status. Level 2 teachers that failed the PDD had average salaries higher than Level 1 teachers even though their performance was worse. To the extent that those that failed the PDD had dismal performance, the PDD is an accurate gatekeeper or filter on advancement to higher licensure level. However, to the extent that many of those that passed had similarly meager performance, it appears the PDD process should have filtered out more teachers. For example, as a group, Level 2 PDD failers improved the reading proficiency of 18 percent of their students while 22.2 percent of their students declined in reading proficiency for a difference (or performance score) of -4.1 percent. Of note, 44 of 113 Level 3 PDD passers had similar reading performance scores of -4.1 percent or less. Using this metric, about 39 percent of Level 3 PDD passers had performance similar to PDD failers.

In general, Level 3 PDD passers outperform the Level 3 teachers that did not go through the PDD process in both math and reading. Among Level 2 teachers, those that did not go through the PDD process outperformed the Level 2 teachers who had passed the PDD in terms of reading scale score changes. Overall, Level 3 PDD teachers did outperform Level 2 teachers; however, the performance of Level 3 Non-PDD teachers, the most expensive group of teachers, had performance similar to all Level 2 teachers.

The LFC survey of teachers who went through the PDD process indicates that cheating on the PDD may be occurring. Survey responses indicate there are validity issues with dossiers being submitted and raise serious concerns about the integrity of the process. An online survey was conducted to receive input from teachers regarding the PDD process. Teachers had mixed opinions about the value of the PDD process. Approximately half (218 out of 433) of the teachers responding said the process does not identify highly effective teachers. One of the most concerning elements of the feedback received from teachers dealt with the integrity of PDD submissions and the general lack of internal controls to verify original work. When asked to provide information concerning what teachers deemed to be appropriate evidence of effective instruction, 9.8 percent (32 of 326) of survey respondents cited concerns with the validity of some PDD submissions or the possibility of fraudulent submissions.

Responses regarding the validity of the PDD submissions range from questioning the level of appropriate assistance to hearsay to actual admission of misrepresenting information on the PDD. The following represent a sample of this range of the survey responses:

- "No [it is not a good measure], because a teacher could just do anything, copy off of someone else or make things up. It is not a good measure to see if teachers are effective."

- "…I don't feel as though it is effective because I have heard that too many teachers use one anothers [*sic*] dossier in parts or in whole as there [*sic*] own."

- "There have been rumors about that some teachers didn't even do their own dossier, that they paid for one and they passed to the next level. You cannot imagine how upset that made me knowing how hard and how long I worked on mine, only to find out that others didn't put in the time or effort and got passed."

- "No, I don't feel the documentation required for a PDD is appropriate evidence. I know I talked to several people who claimed to have 'made theirs up'."

- "…i wrote about student x and y, student x improved but student y just never improved despite all the help given. When i wrote up what actually happened, i was told that the dossier that failed were the ones where a student didn't improve. So i rewrote student y's out come, y became very successful, and my dossier passed with no problems. I choose y because y was an extremely poor student with a number of out side of school problems also."

Due to the fact that the survey instrument did not specifically solicit information concerning the authenticity of the dossiers being submitted, and that this information was freely given by nearly 1 out of 10 respondents, it raises serious concerns about how widespread the problem may be.

This problem affects the ability of the system to identify quality teachers, degrades the faith teachers place in the system, and may have detrimental effects to students receiving instruction from teachers who are misidentified as high-quality teachers.

Student outcomes are not central to the PDD submittal or the teacher evaluation process. PED's guidelines do not require documentation of student performance on standard based assessments for PDD submittal or teacher evaluations. While a modicum of student performance is part of teacher evaluations, student outcomes are peripheral to, rather than central to, teacher evaluations. Some small, likely inadequate, level of documentation of student work is required for the PDD for licensure advancement. Requiring documentation of student performance for purposes of teacher evaluations could assist those teachers in the preparation of future PDDs.

New Mexico's teacher evaluation and licensure system is based on the nine competencies and five strands of the PDD shown in Appendix A. Of the nine competencies, only 2 mention student learning. These are #3; The teacher communicates with and obtains feedback from students in a manner that enhances student learning and understanding; and #4; The teacher comprehends the principles of student growth, development and learning and applies them appropriately.

The guidelines for teacher evaluations provided to districts by PED do not comply with the spirit of PED's rules in the NMAC guiding these evaluations. As outlined in Title 6, Chapter 69, Part 4: Performance Evaluation System Requirements for Teachers, "every public school teacher must have an annual performance evaluation based on an annual professional development plan (PDP)." While the administrative code requires an annual evaluation for every teacher, Level 2 and Level 3 teachers are formally evaluated only every 3 years. For Level 2 and 3 evaluations, the PED guidelines describe classroom observation and reflection on the annual PDP as a 'formative' evaluation and the three year formal evaluation as a 'summative' evaluation. The case could be made that the annual classroom observation and reflection on the PDP alone do not constitute a sufficient evaluation as envisioned in administrative code.

Of the 5 strands of the PDD, Strands A and B, in part, relate to student outcomes. Strand A contains a section on Student achievement but requires demonstration of student learning using only three student examples. PED's instructions regarding this section are as follows: "Select three examples of student work that represent high, mid-range, and low levels of achievement on an assignment, performance, task or other activity completed in connection with the 3 to 5 hour segment of instruction. A work example in this section of Strand A should be a single assignment, performance, task or other activity rather than several collected pieces of student work." Strand B address competency four and intends to demonstrate how a student is learning over an extended period of time; however the required documentation or data is limited to reflect the learning of only two students. In March 2006, OEA presented to the LFC the findings of its analysis of the three-tiered system and found that "the competency that Level 1 teachers had the most difficulty passing was competency four, "The teacher comprehends the principles of student growth, development and learning, and applies them appropriately."

Recommendations

Professional Development Dossier and teacher evaluations

PED should convene a workgroup consisting of PED staff, policy makers, and educators to improve the Three-Tiered Licensure and Evaluation System. The workgroup should submit a report to the LESC and LFC which evaluates the following proposals:

- Require more evidence of student performance in PDD submissions and teacher evaluations. The PDD process should increase the focus on a teacher's impact on all students by requiring documentation of outcomes for more than two or three students. Evidence of instructional effectiveness should play a larger role in the PDD process and teacher evaluations. Demonstration of effective instruction should include objective data, if available, such as standardized test scores or results from short cycle assessments. Objective test data may take less time to assemble and evaluate than the 'portfolio' approach currently employed by the PDD process.

- Establish goals for 'expected growth' in grade levels and content areas.

- Give districts additional guidance about how to use student performance data to target professional development for teachers showing less than desirable performance.

- Continue to study the link between teacher effectiveness and student outcomes to determine whether a consistently high performing Level 2 teacher could use multi-year NMSBA or short cycle assessment results to qualify for Level 3 pay.

PED should ensure that principals review and authenticate that the PDD submittal is the teachers' actual work. PED should consider establishing penalties for false PDD submittals. Require that the principal, not 'administrator' sign off on strands D and E. Principals are required to attend bi-annual training on evaluation of teachers and they are most knowledgeable about teacher performance through their interaction in developing and reviewing the Professional Development Plans (PDP). This would provide an important internal control establishing the validity of the PDD submission.

PED should establish clear guidelines for districts regarding the proper level of assistance teachers can receive when developing and submitting their PDDs. In order to communicate statewide expectations for PDD submittals, PED needs to formalize the expectation that PDD submittals should reflect the original work of the teacher and clearly articulate the levels of outside assistance that are appropriate. Assistance from supervisors, other teachers, and/or organizations, whether paid or unpaid, that include editing for content and guidance on how to complete strands detract from how much of the PDD submittal is the original work of the teacher. This in turn limits the precision of the PDD process in reviewing the actual performance of the teacher. Formalizing what PED considers appropriate levels of assistance in guidelines will also help to clarify what is considered a valid PDD submittal and what is considered a fraudulent submittal and subject to an ethics violation.

PED and districts should ensure additional steps be taken for teachers who have poor student performance. PED and districts should require teachers failing the PDD to take additional mentoring or professional development because these teachers, in general, have relatively poor student outcomes. Districts should direct principals to review and report on teacher performance in terms of student academic growth. Districts should develop specific action plans to improve the results of the least effective teachers. PED should consider requiring teachers with poor performance to submit or resubmit a PDD.

The "Race to the Top" is a competition for $4.35 billion dollars in discretionary grants provided through the American Recovery and Reinvestment Act (ARRA) that will give support to top states that have demonstrated commitment to improving education in their states. States are expected to demonstrate progress in several areas including improving the use of longitudinal data, achieving equity in the distribution of effective teachers, improving low performing schools, and linking teacher evaluations to student outcomes. New Mexico would be in a stronger position to receive any federal competitive grant monies from the Race to the Top fund contingent on stringent teacher evaluations if:

- Summative evaluations' were required every year for Level 2 and Level 3 teachers
- Additional documentation of student performance was required
- Teacher evaluations included additional performance levels beyond the simple "Meets" or "Does Not Meet" levels.

LFC REQUESTS FOR DATA

LFC and PED staff encountered numerous challenges in developing a statewide data set connecting teachers to their students' test scores. LFC and PED staff worked to eliminate New Mexico Alternate Performance Assessment (NMAPA), Braille and Spanish language versions of the NMSBA. Removing these groups resulted in numerous iterations of data requests, and the result was multiple data request addendums. Because of the nature of the request, information on student assessment from the Student Teacher Accountability Reporting System (STARS) database needed to be paired with information for the PDD from a licensure database. Although it was unnecessary, PED staff elected to scramble teacher IDs which made linking datasets and ensuring data accuracy overly cumbersome.

Further complicating analysis of the request, LFC and PED staff discovered errors in both data sets including:

- the omission of a grade in one of the years of achievement gap data (which was received via an addendum),
- and the omission of data for four districts from the teacher licensure data set, including Gallup, Corona, Floyd, and Cimarron. Data for these districts was not included in the analysis because the error was discovered late in the evaluation process.

LFC staff took appropriate steps to provide reasonable assurance that the data provided by PED was accurate and reliable. LFC staff took steps to perform quality checks to maintain data analysis accuracy, and determined that the omitted districts would not change the high-level findings as they represent a very small percentage of the student population analyzed. For these reasons the analyses based on these data sets are valuable. PED staff indicated that working with LFC staff to identify issues with the accuracy of the data requests has improved PED's intra-agency use of the STARS application.

Recommendations

LFC and LESC staff should be provided access to STARS and provided training on how to replicate these data extractions to minimize the impact of future data requests on PED staff.

PED should formalize a process to produce timely, accurate and complete data linking statewide teacher licensure data to student assessment data and ensuring the data integrity of information presented to interested parties. In order to facilitate future studies, LFC recommends that PED take steps to make sure that statewide data produced on student performance and associated links to teacher licensure levels or regional areas is accurate and complete.

THE ACADEMIC PERFORMANCE STUDENTS FROM LOW INCOME FAMILIES

The federal No Child Left Behind Act of 2001 (NCLB) greatly increased the pressure on schools to close the "achievement gap" among various student groups. NCLB requires 100 percent proficiency of all students, regardless of race/ethnicity or socioeconomic status to be proficient on state standardized tests by 2014. While this is an aggressive goal, the requirement for schools to meet interim performance targets for each student sub-group (economically disadvantaged, African American, Caucasian, Hispanic, Native American, etc) and publish results has heightened the focus on reducing disparities in student achievement.

The achievement gap is the difference between the academic performance of students from low income families (economically disadvantaged) and wealthier students and between minority students and their non-minority peers. According to the Education Commission of the States, "the gap in achievement separating poor and minority students from less disadvantaged students has been the focus of discussion, research and controversy for nearly 40 years." A large body of research has identified a variety of factors that appear related to the achievement gap, including:

- students' economic background
- peer influences
- their parents' education level
- teachers' expectations
- their access to high-quality preschool instruction, and
- curricular and instructional quality.

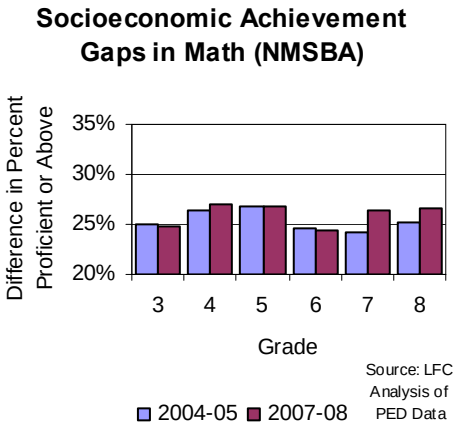
Closing the achievement gap requires accelerating the growth of minority and low income students at higher rates than the growth of their peers. Since the achievement gap likely emerges prior to these students entering school, students enrolled in high-poverty schools must have greater growth from the day they begin their formal education. In a report by the Economic Policy Institute (EPI), achievement gaps, in large measure, appear when children first enter kindergarten as evidenced by substantial variations in children's cognitive ability. According to research, socioeconomic status accounts for more of these variations than any other factor, including race/ethnicity, family educational expectations, and access to quality child care. A larger proportion of racial and ethnic minority children are from low income families, which contribute to the disparities in achievement between these students and their Caucasian peers.

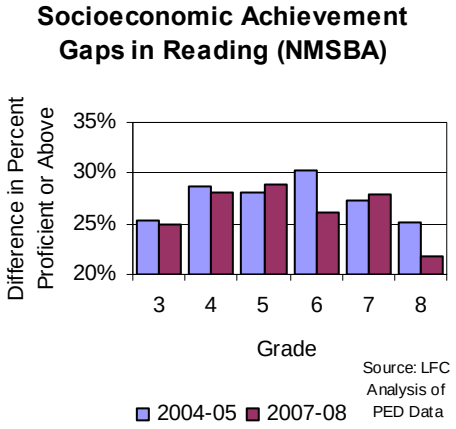
Until 2008, New Mexico did not disaggregate student test data to compare low income versus non low income student scores, which masked the size of the true achievement gap. For example, in SY08 about 43 percent of low income 4th grade students were proficient or above on the NMSBA compared to about 71 percent of their non-economically disadvantaged peers scoring proficient and above – a 28 point gap. However, traditional achievement gap measures would have shown only an eight percent gap using the "all" students category results of 51 percent versus the economically disadvantaged student average of 43 percent proficient and above.

This evaluation compares economically disadvantaged students to non-economically disadvantaged students using the New Mexico Standards Based Assessment (NMSBA); a new development. Past evaluations have examined the differences between economically disadvantaged students and the state average for all students' performance when discussing socioeconomic achievement gaps. The old methodology provided an inaccurate picture of the gap in student achievement as the traditionally lower achievement of economically disadvantaged students is included in the state average for all students and shows a misleadingly smaller gap than what may actually exist. The large population of economically disadvantaged students in New Mexico skews the state average. This evaluation provides the most accurate picture available of the socioeconomic achievement gap based on NMSBA data.

Generally, the achievement gap between low income students and their peers has not changed over the past four years, ranging from 22 to 28 percent in SY08. These differences are substantial, especially given the large numbers of low income students in the State.

When the achievement gap is viewed on its own, it becomes clear that the gap has closed in sixth and eighth grade reading over the last four years. When compared to gaps in other grades, most other areas have not shown large losses or gains in closing the gap. Also, the gap has grown in the difference between non-economically disadvantaged students and economically disadvantaged students in some grades. The achievement gap also remains fairly high, with all groups showing a difference of twenty percentage points or more.



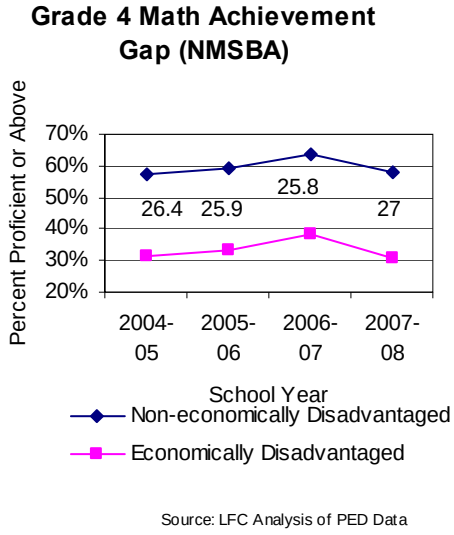


Results on the NMSBA show little growth in student achievement in fourth grade, but some positive improvement in eighth grade student achievement. Eighth grade reading is unique in improving performance and closing the achievement gap. However, the gap between non-economically disadvantaged and economically disadvantaged students remains high in most areas.

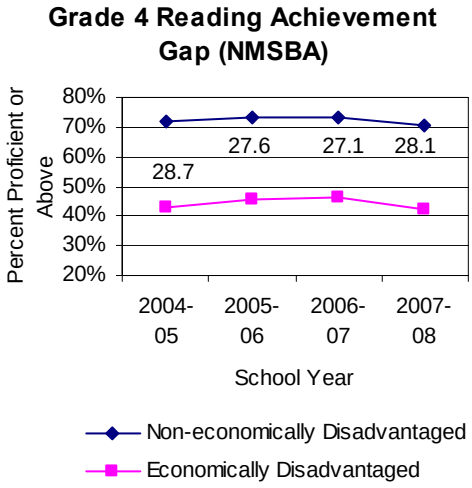
NMSBA scores for fourth and eighth grade Math and English show:

- stagnant achievement levels in fourth grade math and reading,
- improving achievement levels in eighth grade math and reading,
- and an achievement gap that is relatively stable in all four areas.

In fourth grade reading and math there is little change in the achievement gap over the last four years. In fourth grade reading, students scoring proficient or above increases over a three year period until the 2007-08 SY when scores decline and the gap grows to the 2004-05 level. Test scores in math also follow this trend generally, with the gap and scores also returning to approximately the same level they were in 2004-05.



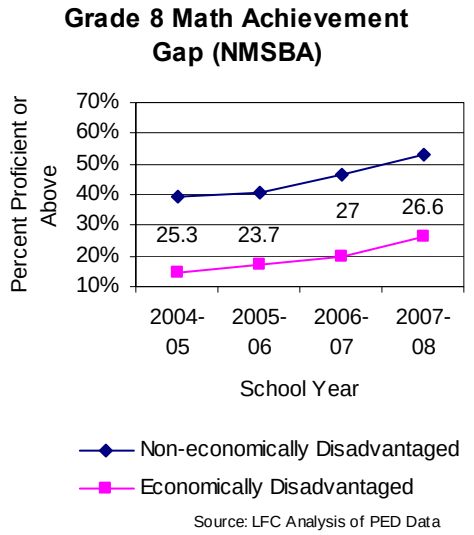
Source: LFC Analysis of PED Data

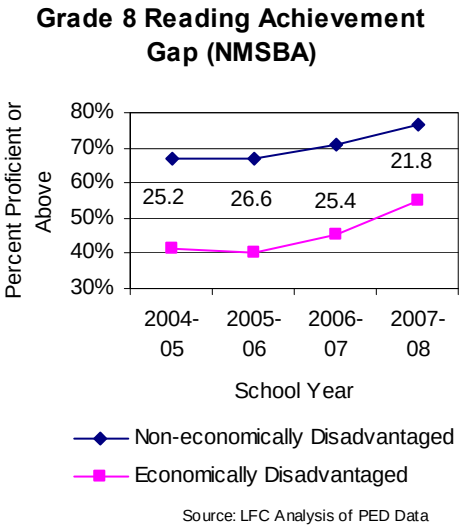


Source: LFC Analysis of PED Data

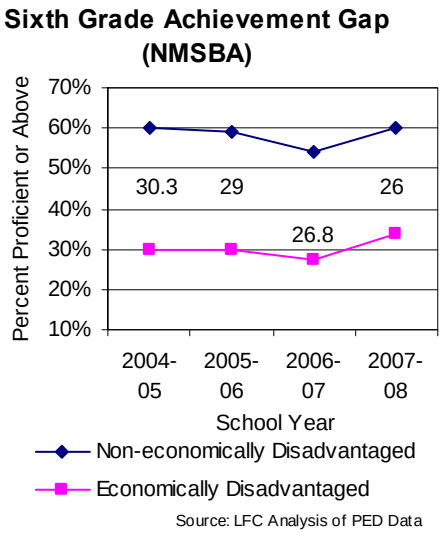
The only area to show marginal improvement in both achievement and in closing the gap is in eighth-grade reading. This is the condition that the State wants to see, with economically disadvantaged students' performance improving at a faster rate than non-economically disadvantaged students. In theory, this type of improvement while closing the gap should be present across all grades in both subjects if positive gains in closing the achievement gap were present statewide. In eighth-grade math, scores improve though the achievement gap stays relatively constant showing marginal growth from the 2004-05 SY. There is little change in the size of the achievement gap in any of the other areas, and any changes in eighth grade-reading achievement gap are small when compared to the size of the achievement gap as illustrated below.

While achievement gains in 8th-grade math have improved, the achievement gap has not. This example illustrates that while economically disadvantaged 8th graders experienced a double digit increase on proficiency levels, it was not enough to keep up with their non-economically disadvantaged peers.





The State reduced the achievement gap four percentage points for sixth grade reading between SY05 and SY08, but only because non-economically disadvantaged students did not make any achievement gains over the period. As a result, eighth grade reading is the only grade and subject area to close the gap while improving student performance for both non-economically disadvantaged and economically disadvantaged students.

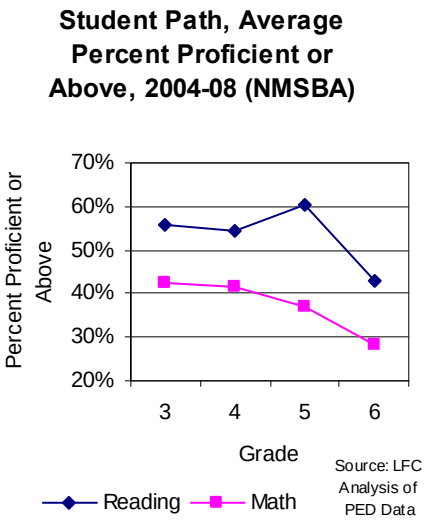


<u>The achievement gap between all non-economically disadvantaged and economically disadvantaged students is larger than any socio-economic gaps within racial/ethnic groups.</u> The NMSBA provided by PED allowed LFC staff to combine ethnicity and socio-economic status to determine if the socio-economic achievement gap was more pronounced among students from certain ethnic backgrounds. For example, does the achievement gap between low income students and their peers persist across racial/ethnic subgroups? Achievement gaps in fourth and eighth grade between all non-economically disadvantaged students and their peers is higher than for any sub-group comparison and has seen very little change from this high level over the last four years in either grade, where other gaps fluctuate.

Socioeconomic status appears to have a more consistent impact on student achievement levels, regardless of race/ethnicity. For example, for all students the achievement gap is about 28 percent between low income students and their peers. The achievement gap between low income Hispanic students and their Hispanic peers was about 22 percentage points for 4th grade reading. Likewise, the achievement gap between Caucasian, low income students and their Caucasian peers was about 20 percentage points.

Further analysis of the overall gap in performance between all Anglo students and all Hispanic students shows about a 25 percentage point difference. However, after controlling for economic status the gap is much narrower among low income Anglos and low income Hispanic students at about 15 percentage points, as shown in Table 26 in the appendix. The differences between other racial and ethnic groups were all lower than the socioeconomic gap with the exception of the difference between Native American students and white students.

As students progress through elementary school the rate of student achievement declines, but the achievement gap stays high throughout a student's elementary school career. For example, about 55 percent of 3rd graders in 2004 were proficient and above in reading, but by the time those students reached 6th grade only 41 percent were still proficient and above. Data on the average number of students scoring proficient or above along the student path shows much higher numbers of students scoring proficient or above on reading and at earlier grades, with an increase in fifth grade followed by a sharp decline in sixth. The data show decreasing numbers of students scoring proficient or above in math across the student path. The state would like to have students become increasingly proficient in all subject areas as they advance grade levels, but this is not occurring.

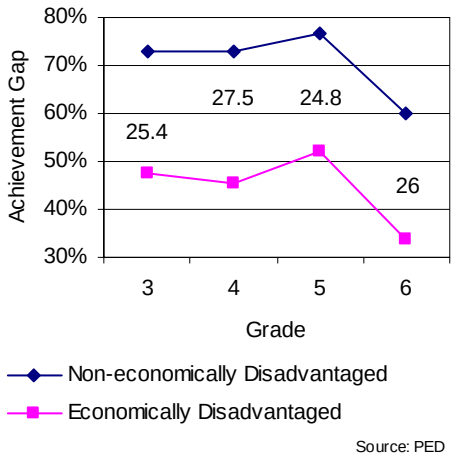


The evaluation team constructed a high-level student cohort path by taking information on third-grade students in the 2004-05 SY and comparing it to their fourth-grade scores in the 2005-06 SY, fifth-grade scores in the 2006-07 SY, and sixth-grade scores in the 2007-08 SY. By following this high level student path of performance by grade level, LFC staff were able to show a high level trend of a cohort of almost the same students over a four year period. Although the cohort chosen to represent students advancing through grades by year does not control for factors such as student mobility or population size, it is felt that enough students from grade to grade would be similar enough to provide a picture of how student performance changes in the State as children advance through the grades.

The low income student achievement gap starts high in 3rd grade and remains at a high level as students advance grade levels, using NMSBA data. This indicates that the achievement gap between non-economically and economically disadvantaged students is not only persistently large from year to year, and little progress has been made in closing the gap, but also remains high as students advance in grade levels. Research shows this gap starts well before 3rd grade.

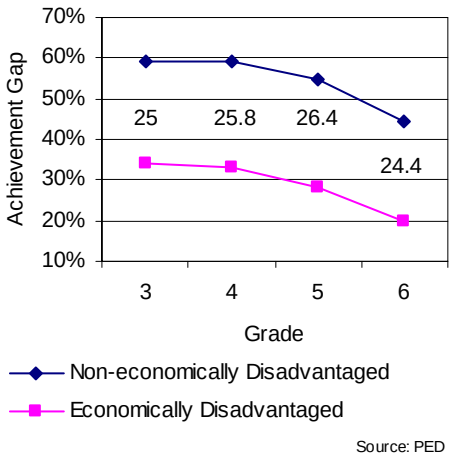
For example, the graphs below show the achievement gap remaining relatively constant as students progress to higher grades. Both low income students and their peers appear to have been experiencing declining rates of achievement as they progress through elementary school. This specific issue could be a topic for future study to assess whether other cohorts of 3rd graders have similar achievement rates as they progress in elementary school and using a tighter cohort of students to account for mobility.

Student Path Reading Average Percent Proficient or Above, 2004-2008 (NMSBA)



Source: PED

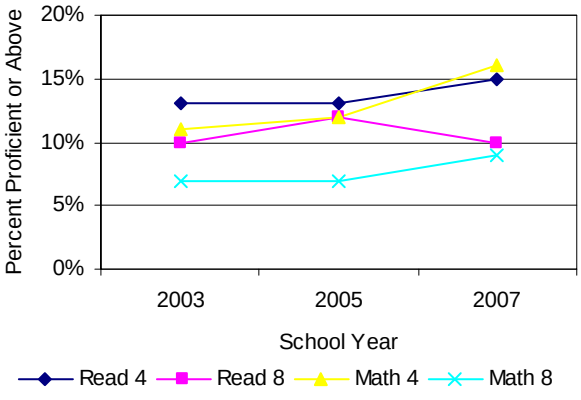
Student Path Math Average Percent Proficient or Above, 2004-2008 (NMSBA)



Source: PED

<u>National test score rankings show New Mexico near the bottom in performance comparisons to other states on overall achievement and closing the achievement gap, despite spending more per student.</u> The National Association of Educational Progress (NAEP) is the test the U.S. Department of Education, National Center for Education Statistics uses to assess fourth, eighth and twelfth grade students and measure their achievement. The only federally administered nation-wide student assessment test, NAEP is the only assessment from which comparisons between states can be made. A comparison of NAEP scores of other states shows New Mexico ranking in the bottom of student achievement nationwide, and below the national average in every reportable category. New Mexico's level of student achievement is extremely low in comparison to other states, and these low levels occur in both fourth and eighth grades, in math and reading, and as measured by both average scale score and students scoring proficient or above.

Free and Reduced Lunch Eligible Students (NAEP)

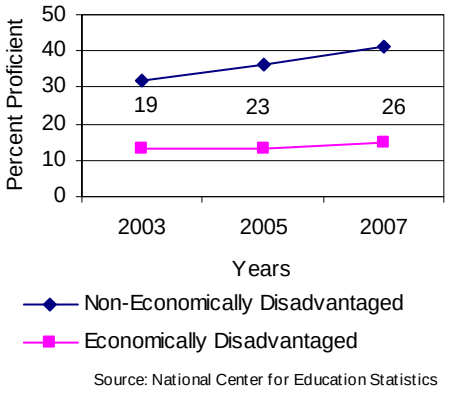


Source: National Center for Education Statistics

In 2007, between nine and 16 percent of economically disadvantaged students demonstrated proficiency or above on the NAEP, depending on the grade level. Student performance, as measured by the number of students scoring proficient or above on the NAEP, has increased in most subject areas for fourth and eighth grade. The only exception is eighth grade reading, which has seen a decline in performance over the last three years.

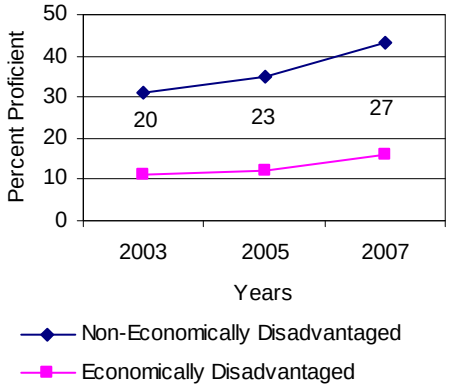
Despite gains in achievement, the achievement gap has increased on the NAEP to as high as 27 point difference between economically disadvantaged students and their peers. In all four of the areas presented below, the income achievement gap as measured by students scoring proficient or above on the NAEP has either remained at the same level it was in years previous or has grown over the last three years. This shows that student performance is not increasing at the rate necessary to close the achievement gap. To close the achievement gap economically disadvantaged students must improve at a faster rate than non-economically disadvantaged students.

4th Grade Reading Scores, NAEP Percent Proficient or Above



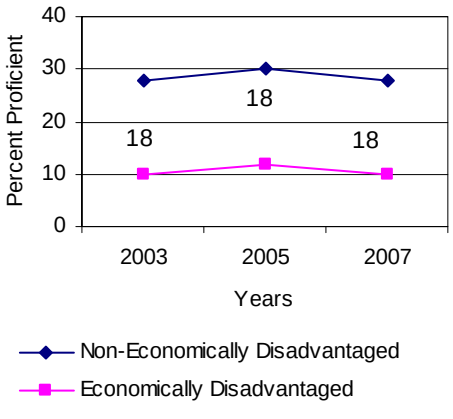
Source: National Center for Education Statistics

4th Grade Math Scores, NAEP Percent Proficient or Above



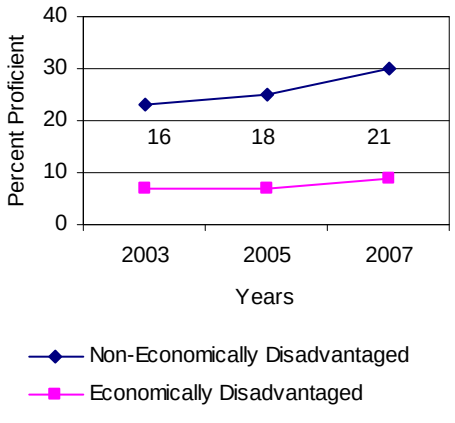
Source: National Center for Education Statistics

8th Grade Reading Scores, NAEP Percent Proficient or Above



Source: National Center for Education Statistics

8th Grade Math Scores, NAEP Percent Proficient or Above



Source: National Center for Education Statistics

New Mexico's growth in the achievement gap is the highest among western states as measured by the NAEP. This trend of increased student performance and an increase in achievement gaps between socio-economic groups occurs across western states on the NAEP test for both fourth and eighth grade math scores. Even as more students become proficient or above in math, more students who are not economically disadvantaged are becoming proficient at a faster rate in nearly every western state.

Western States,
Average Expenditure per Pupil, FY07

National Rank	States	Total from Federal/State/Local Sources (in dollars)
7	Wyoming	13,266
22	United States†,‡,*	9,683
29	Oregon	8,958
30	California ‡	8,952
32	New Mexico	8,849
38	Washington*	8,524
41	Colorado	8,286
46	Nevada	7,806
49	Arizona	7,338
51	Idaho	6,648
52	Utah	5,706
†50 States and District of Columbia		
		Source: NCES, Common Core of Data

New Mexico has also seen the largest magnitude of change in the size of the achievement gap between non-economically disadvantaged and economically disadvantaged students of all western states. For example, the NAEP achievement gap was 7 percent higher in 2007 than 2003 in New Mexico, while others states' gap growth was between -3 percent and 6 percent for 4th grade math, as shown in table 27 in the appendix. New Mexico ranks in the bottom of student achievement for both fourth and eighth grade math and any gains made in student achievement in math over the last four years have not been enough to change this status.

Although New Mexico ranks in the bottom in performance comparisons on the NAEP, it is in the middle of most western states with regard to expenditures per pupil. Although New Mexico spends below the national average with regard to expenditures per pupil, it is allocating more resources per student than many western states that are shown to perform better on the NAEP. New Mexico ranks 32nd in the nation when counting the District of Columbia and the national average. This suggests that for the resources allocated, the State should be seeing increasing returns in the form of improved student performance.

Currently, the Educational Plan for Student Success (EPSS) does not view the achievement gap of economically disadvantaged students versus non-economically disadvantaged students. Districts and schools develop their EPSS using PED's template. The template includes a data table that allows comparisons of NMSBA scores for the NCLB subgroups (i.e., Hispanic, Native American, and economically disadvantaged) and the data for the current year's annual measurable objective (AMO) and the previous year's average for all students. Comparing economically disadvantaged students to those in the average for all students can unintentionally decrease the size of the gap.

Recommendations

The Committee and PED should regularly monitor the status of New Mexico's achievement gap between low income students and their peers who are not from low income families.

PED should implement the following.

- update requirements for Educational Plans for Student Success to incorporate the more accurate comparison of a school or district's achievement gap between low income students and their peers who are not from low income families; and
- incorporate information on how to use STARS data to further disaggregate student performance information by low income status of different ethnic/racial subgroups in its STARS training and in technical assistance to underperforming districts and schools.

STATE OF NEW MEXICO
PUBLIC EDUCATION DEPARTMENT
300 DON GASPAR
SANTA FE, NEW MEXICO 87501-2786
Telephone (505) 827-5800
www.ped.state.nm.us

DR. VERONICA C. GARCÍA
SECRETARY OF EDUCATION

BILL RICHARDSON
Governor

June 1, 2009

MEMORANDUM

TO: Manu Patel, Deputy Director, Legislative Finance Committee

FROM: Veronica C. García, Ed.D.
Secretary of Education

RE: **PED RESPONSE TO REVIEW OF ACHIEVEMENT GAP AND THREE-TIER LICENSURE**

I want to begin my response by praising the LFC staff for its thorough review of the achievement gap and the three-tier licensure system and for providing the Public Education Department (PED) with an early copy of the report memorandum. I believe that the work on this evaluation has assisted the PED in expanding its data and reporting capabilities to the Legislature.

The PED with its legislative partners now has the capability to measure student learning, follow a cohort over several years and link learning proficiency to specific instructors. This type of system will ultimately enable us to better inform instruction.

History of Three-Tier Licensure

In 2003, the Legislature passed HB 212, which enacted major education reforms for New Mexico and noted in statute that the teacher shortage had affected the ability of New Mexico to compete for the best teachers and that, unless the state and school districts found ways to mentor beginning teachers, intervene with teachers while they still show promise, improve the job satisfaction of quality teachers and elevate the teaching profession by shifting to a professional educator licensing and salary system, public schools would be unable to recruit and retain the highest quality teachers in the teaching profession in New Mexico.

It further noted that teachers who do not meet performance standards must improve their skills or they will not continue to be employed as teachers.

Ultimately, this resulted in the three-tier licensure system, which has been in place since July 1, 2004. Since then, New Mexico has achieved the following milestones:

- For 2008, 94% of core classes in New Mexico were taught by highly qualified teachers, an increase over 2003 and 2004 when just 67% of core classes were taught by highly qualified teachers.

- New Mexico ranks third in the nation in percent change in average teachers' salaries from 1997-98 to 2007-08.

- New Mexico ranks 17th in the nation for its efforts to improve teaching.

- Regarding teacher recruitment, there was an increase from 2004 to 2008 in reciprocity Level I of 45%, Level II of 44% and Level III of 49%. This indicates that New Mexico is now able to attract more highly qualified individuals, spanning all levels, from out of state.

- New Mexico's Student Teacher Accountability Reporting System (STARS) was recognized as a national model for longitundinal data systems.

Clearly, the three-tier licensure system has met the Legislature's intent to elevate the teaching profession by shifting to a professional educator licensing and salary system.

Over the past five years, New Mexico has consistently invested in high-quality teachers, closing the achievement gap and increasing accountability to transform New Mexico schools. Tying student performance to teachers is an important dimension to elevating the teaching profession and closing the achievement gap. However, the PED believes that utilizing standardized assessments and scale scores is not the most effective method for holding teachers accountable.

Concerns Related to Professional Development Dossiers (PDDs)

Within the PDD system there are different levels of responsibility for teachers, principals, superintendents, the PED and dossier reviewers. Based on some of the responses to the LFC's online survey, it appears that in some cases principals and superintendents are not providing proper oversight, and some teachers are not meeting the requirements that pertain to them. What these individuals reported in the survey constitutes an ethics violation and needs to be handled as such. This will be addressed by the PED in future professional development conferences and trainings across all levels of responsibility, with the goal of ensuring integrity and a system that is beyond reproach.

Most particularly, principals must be strong in their evaluations of teacher performance. Although New Mexico is not a pay-for-performance state, student learning has to be a consideration in the teacher evaluation process.

Regarding the placement of less experienced teachers in high-needs schools, NCLB requires that districts place highly qualified teachers in schools in need of improvement. A teacher can be highly qualified at Levels I, II or III. We believe that through the oversight of our Priority Schools Bureau this dynamic is improving. However, for schools not in the school improvement cycle, school districts have autonomy regarding teacher placement. New Mexico has been able to increase the percent of highly qualified teachers from 67% to 98% over the last five years.

Given that five years have passed since the three-tier licensure system took effect, the PED agrees with the recommendation that a task force should be convened this interim to consider the strengths and weaknesses in the system and make recommendations for changes to the state's PDD requirements.

Analysis of State Achievement Gap

The PED does not dispute the achievement gap findings in the LFC's evaluation report and commends the comparisons made between economically disadvantaged and non-economically disadvantaged students. While we have long recognized that there has been some improvement in the achievement gap, we must do more.

The PED, with the support of the Legislature and Governor Richardson, has instituted reforms under its School Improvement Framework; brought expanded school breakfast and physical education into elementary schools; launched PreK statewide; expanded K-3+ learning opportunities; ensured improved data reporting from districts and charter schools on student proficiency; worked with districts on utilizing short-cycle assessments to measure student progress; instituted high school redesign; and revamped the status of math and science education in the state. It should be noted that reforms take time to work. Specific to economically disadvantaged students, funding for education reform under Governor Richardson has been focused on areas serving low income students.

As a result, there have been demonstrated improvements in recent years in scores on the New Mexico Standards-Based Assessment (SBA), the National Assessment of Educational Progress (NAEP) and the ACT. For example, the percentage of all students scoring proficient or better in reading on the SBA improved from 50% in 2005 to 53% in 2008. In math, the percent of students proficient or better has improved from 30 to 36% and in science student proficiency has increased from 40 to 43% in the same period.

In 2007, according to NAEP, New Mexico students showed gains in reading and math achievement with some significant gains being made by Hispanics and free-and-reduced-price lunch students. Fourth grade students showed progress in both math and reading, 8th grade students showed gains in math and scores in 8th grade reading remained stable.

The decrease reported by the LFC for 8th grade reading (12 to 10% proficient and above) is not statistically significant. Because NAEP is an assessment survey, changes need to be statistically significant. We recommend that the LFC consider the statistical significance in the achievement gap conclusions that it has drawn.

Compared with results from the 2005 NAEP, New Mexico was one of only 14 states in 2007 that improved in both 4th and 8th grades in math. It was among only four states to show significant increases for Hispanic students in 4th grade reading. The other three states were Massachusetts, New Jersey and Illinois. There was also an increase in the average scale score for Hispanic students in both 4th and 8th grade math.

Students who were eligible for free-and-reduced-price lunch in New Mexico showed significant increases in 4th and 8th grade math as well as an increase in the percentage performing at or above the basic level in grade 4 reading.

The Education Trust in March 2009 credited New Mexico for its gains on the NAEP in 4th and 8th grades in 2007 while noting that the achievement gap stayed constant or did not change significantly. While we are improving, we must continue our efforts to see increased performance for all students.

New Mexico does not rank at the bottom on NAEP. It does, unfortunately, share in rankings near the bottom with up to nine other states, depending on grade level and subject.

New Mexico Standards-Based Assessment (SBA)

We must be mindful that the student proficiency data reviewed for this evaluation represent only a snapshot. First, the evaluation only considered the student test results in reading and math for 4th, 5th and 6th grade teachers. Second, a more accurate analysis would have been to measure only the students who stayed in the system for three consecutive years, removing from the study students who came and went during that time period. The study did not do this.

Also of concern is the study's use of scaled scores and wide reliance on them to draw conclusions. While scaled scores are consistent throughout the state, the purpose and psychometric characteristics of the SBA supports their interpretation only in large groups. The smaller the group size of analysis the more inaccurate scaled scores become. That is why New Mexico, as required by the *No Child Left Behind Act of 2001* (NCLB), relies on Annual Measurable Objectives (AMOs) and limits accountability to a school-wide population combining all grades. A more precise assessment for interpreting growth in a classroom of students would be a statewide short-cycle assessment, which is not available.

The LFC notes in its report that both low income students and their peers appear to be experiencing declining rates of achievement as they progress through elementary school. Please be mindful that the state's standards become more stringent as students progress from grade to grade. Moreover, local and national evidence exists that in general, all students experience a decrease in achievement scores in the 6th grade. The PED is aware of the dip in 6th grade scores and is investigating the alignment of our standards to the state's curriculum and assessments.

Growth in student proficiency has long been a concern of the PED. In 2008, the PED received notice that its proposal to move to a growth-based accountability model was not approved by the U.S. Department of Education (USDE). Under the current system, Adequate Yearly Progress (AYP) assesses individual school performance based only on current-year information. The growth-based model would have allowed the state to measure a school or district's growth or improvement over three years.

New Mexico's proposal was recognized by the USDE for several strengths including "focus on the lower-performing schools and on reducing achievement gaps for non-proficient students." Unfortunately, under proposal guidelines, New Mexico had a difficult time making the growth-based model fair for our small schools. I am committed to exploring every opportunity that will help us move to an accountability system that truly captures the progress that so many of our schools are making.

I am concerned with the statement in the LFC report on page 2, as follows: "The difference in student achievement between these licensure levels (~2 points in reading) appears insufficient given the $20 thousand dollar difference in minimum pay." These ~2 points in reading for certain students can represent tremendous improvement. It is noteworthy that a scaled score improvement in the group of students that is lowest in proficiency impacts twice as many students as changes for students in the highest category of proficiency. For example, in math statewide, 22,632 students were in the Beginning Step, whereas only 11,374 were in the Advanced category. The interpretation of meaningful change should balance the magnitude of growth with the size of the group experiencing the growth.

We vehemently disagree with any suggestion that the PED has sought to "mask" the true size of the achievement gap (p.p. 30-31), or that the methodology used is "old." In 2008, the USDE congratulated the PED for having a student assessment system that meets all statutory and regulatory requirements under NCLB, making New Mexico the 33rd state in the nation to receive its approval. We have consistently and openly reported on the achievement gap and student proficiency as required under NCLB and have expanded public reporting to include non-mandatory disaggregation whenever needed.

Regarding Educational Plans for Student Success (EPSS), the PED agrees with the report's recommendation to incorporate into the EPSS for schools/districts comparisons between low income students and those who are not from low income families. However, PED will not be able to make that change until the September 2009 submissions since districts/schools are currently working on their June 2009 submissions. Additionally, the PED will incorporate that data point as it develops the web-based EPSS tool.

STARS and Other Data Requests

PED's reporting system, in concert with a stable assessment system, is now capable of providing reliable student-linked data to inform important growth research. STARS was designed to provide student, teacher and course data from the 89 school districts and charter schools, which have disparate data systems. This has been accomplished, thus enabling student linkage to teachers. This is an integral design of STARS.

Specific to the issue of scrambling the data, the PED takes seriously its responsibility to protect the confidentiality of individual students. The federal Family Educational Rights and Privacy Act (FERPA) of 1996 requires state educational agencies to protect the identity of individual students.

Contrary to the LFC report, the PED did not have difficulty providing assessment test data. At issue was that the LFC's initial data request grew from 40 elementary schools to a request for all elementary schools (more than 400). Adding all elementary schools resulted in a radically different data format than what had been requested initially, which required a complete rebuilding of the data format.

The PED notified the LFC that four districts provided incorrect grade-level categories during one of the years examined by the LFC. Therefore, we were unable to provide consistent linkage/statistics across two years' worth of student and teacher data for the districts in question.

PED has strived to ensure that teacher licensure data are very closely linked in STARS via nightly uploads to the Data Warehouse. However, the PDD system is a standalone system that employs a "snapshot" view of progression in submissions provided by teachers and is completely unrelated to STARS. For the LFC's requests, which required data from the PDD system, PED's IT staff developed a set of data extractions and merged other spreadsheets with the STARS data.

Funding for STARS has permitted PED to develop the first components of an Education User Interface (EUI). We plan to expand and welcome the opportunity to collaborate and understand information needs with and for the Legislature.

Critical to the success of future evaluations is for initial data parameters to be established and adhered to by the requesting legislative agency. This is especially needed when conducting evaluations during legislative sessions and standard reporting periods for districts and charter schools, when the PED's workload is heightened.

Looking to the Future

The PED is in agreement that other systems to improve teacher performance in the classroom could be reviewed. But doing so can be costly. In 2008, the PED reviewed the cost of instituting a model for educators in high-needs schools that would have added 10 non-teaching days to teachers' contracts and up to 20 days to principals' contracts for professional development, curriculum planning and improving instructional technologies. The cost to pilot such a model for 2,000 educators in 60 schools was approximately $7.2 million.

Regarding the LFC recommendation to study a pay-for-performance pilot paid for through competitive federal stimulus money, we are mindful of opportunities made available through the American Recovery and Reinvestment Act (ARRA) to promote new, reform-minded education plans and the strong accountability that accompanies ARRA funds.

Also, the PED has long been aware of national efforts at pay for performance. In 2005, the PED proposed $250,000 for a teacher incentive for priority/rural schools. We believe that such a pilot is worthy of investigation.

However, regarding the issue of the salary incentive, as is noted in the LFC report, the schedule of minimum salaries by level was only fully implemented with school year 2007-2008, with Level III salaries being the last to be phased in.

We also believe that more time is needed to conduct a quantitative evaluation of the three-tier licensure system. As is stated in the evaluation report, students whose teachers were Level III and had gone through the PDD process generally outperformed their peers. But only 6,459 of the teachers in the 2008-09 school year were Level III teachers. There were 10,504 teachers at Level II. Given the gains for Level III teachers, we must strive to bring Level II teachers to Level III status. It is important to note that a Level I teacher must advance to Level II under the system.

A key question in reviewing the LFC report is, "Are we using the right criteria to measure the success of three-tier licensure?" A future need is to identify an assessment tool that will accurately measure student progress. This tool is not the SBA. Short-cycle assessments are one method of accurately assessing and driving instruction. However, if statewide comparisons are going to be made, New Mexico will need to use a statewide tool, which is costly.

I close by agreeing that this a good start to a much-needed larger conversation about educating children in New Mexico. Student learning is a complex process that involves many components that are outside the public schools' control, i.e. poverty, parenting, home environments that are conducive to learning. The three-tier licensure system was not designed to solve all of the problems associated with educating children and is just one step that New Mexico has taken to improve learning. With more highly qualified teachers in the classrooms and innovative programming we are beginning to impact the achievement gap.

Thank you.

VCG/RMW/rmw

cc: The Honorable Lucky Varela, Chairman, Legislative Finance Committee
The Honorable John Arthur Smith, Vice-Chairman, Legislative Finance Committee
David Abbey, Director, Legislative Finance Committee
Frances Maestas, Director, Legislative Education Study Committee
Brian Condit, Chief of Staff, Office of the Governor
Dr. Peter Winograd, Director, Office of Education Accountability
Charles Sallee, Program Evaluation Manager, Legislative Finance Committee
Paul Aguilar, Public Education Analyst, Legislative Finance Committee
Dr. Catherine Cross Maple, Deputy Secretary, Public Education Department
Don Moya, Deputy Secretary, Public Education Department
Dr. Mary Rose CdeBaca, Assistant Secretary, Public Education Department
Dr. Sheila Hyde, Assistant Secretary, Public Education Department
Robert Piro, Chief Information Officer, Public Education Department
Ruth Williams, Manager, Public Education Department

APPENDIX A: STUDENT PERFORMANCE DATA TABLES

Table 9: Classroom Performance Summary

	Level 3 PDD	Level 3 Non-PDD	Level 2 PDD	Level 2 PDD Fail	Level 2 Non-PDD	Level 1	Level 0
N teachers	113	519	321	25	897	431	30
N students	2010	8951	5580	392	15488	7277	501
Score Change (Read)	16.8	15.7	14.7	13.6	15.1	13.9	14.5
Score Change (Math)	23.7	20.4	21.3	18.0	20.6	18.3	15.0
% Improved (Read)	18.6%	19.2%	19.0%	18.1%	19.3%	16.9%	16.1%
% maintained (Read)	64.7%	62.9%	62.1%	59.8%	62.8%	62.7%	66.0%
% declined (Read)	16.7%	17.9%	18.9%	22.2%	18.0%	20.4%	17.9%
% Improved (Math)	18.0%	15.0%	15.6%	9.1%	14.7%	13.0%	10.8%
% maintained (Math)	67.6%	65.9%	66.5%	66.1%	66.6%	65.7%	63.6%
% declined (Math)	14.4%	19.1%	18.0%	24.8%	18.7%	21.3%	25.7%
Avg Performance score	2.8%	-1.3%	-1.1%	-9.9%	-1.3%	-5.9%	-8.3%

Source: LFC analysis of PED data

Table 10: Classroom Demographics Summary

	Level 3 PDD	Level 3 Non-PDD	Level 2 PDD	Level 2 PDD Fail	Level 2 Non-PDD	Level 1	Level 0
N teachers	113	519	321	25	897	431	30
N students	2010	8951	5580	392	15488	7277	501
% Enter below Proficient (Read)	37.5%	37.9%	42.3%	43.6%	41.37%	43.28%	47.5%
% Enter below Proficient (Math)	51.8%	50.5%	55.1%	55.1%	55.45%	56.41%	63.1%
%Free/Reduced Lunch	61.3%	60.5%	70.5%	82.9%	70.0%	72.1%	81.3%
% Hispanic	52.4%	52.1%	61.0%	74.7%	56.1%	61.0%	66.7%
% Native American	6.0%	7.2%	6.1%	3.2%	10.4%	7.6%	10.8%
% Gifted	6.3%	7.1%	5.7%	2.3%	5.5%	5.4%	3.5%
% IEP	8.1%	8.2%	8.6%	7.5%	8.6%	8.3%	10.1%
Avg Salary	$49,755	$54,424	$41,466	$44,056	$44,906	$33,186	$32,369
Avg Yrs Exp	11.0	18.7	5.0	10.2	12.7	2.1	2.4

Source: LFC analysis of PED data

TABLE 11: NMSBA Cut Score Intervals

	GRADE	Beginning	Nearing	Proficient	Advanced
MATH	3	0-555	556-610	611-659	660-999
	4	0-583	584-635	636-677	678-999
	5	0-608	609-657	658-695	696-999
	6	0-628	629-676	677-713	714-999
READING	GRADE	Beginning	Nearing	Proficient	Advanced
	3	0-591	592-620	621-669	670-999
	4	0-599	600-639	640-682	683-999
	5	0-608	609--655	656-695	696-999
	6	0-619	620-668	669-709	710-999

Source: NMSBA 2008 Technical Report

Table 12: Change in Math Proficiency Status

	% Beginning	% Nearing	% Proficient	% Advanced	% Below Proficient	% Proficient and Above	Students	Teachers
Level 0								
2007	10.6%	52.5%	30.3%	6.6%	63.1%	36.9%	501	30
2008	14.4%	55.9%	25.3%	4.4%	70.3%	29.7%		
Level 1								
2007	8.3%	48.1%	37.1%	6.5%	56.4%	43.6%	7274	431
2008	9.9%	52.7%	31.7%	5.8%	62.6%	37.4%		
Level 2 PDD PASS								
2007	7.3%	47.7%	38.0%	6.9%	55.1%	44.9%	5562	321
2008	7.7%	50.6%	33.7%	8.0%	58.3%	41.7%		
Level 2 PDD DID NOT PASS								
2007	7.4%	47.7%	40.1%	4.8%	55.1%	44.9%	392	25
2008	8.4%	58.7%	29.8%	3.1%	67.1%	32.9%		
Level 2 Non-PDD								
2007	7.4%	48.0%	37.5%	7.0%	55.5%	44.5%	15491	897
2008	8.3%	51.2%	33.0%	7.6%	59.4%	40.6%		
Level 3 PDD								
2007	6.2%	45.6%	39.5%	8.7%	51.8%	48.2%	2010	113
2008	6.6%	44.8%	40.1%	8.4%	51.4%	48.5%		
Level 3 Non-PDD								
2007	6.0%	44.5%	41.0%	8.4%	50.5%	49.5%	8951	519
2008	7.0%	48.0%	35.3%	9.6%	55.0%	45.0%		

Source: LFC analysis of PED data

Table 13: Change in Reading Proficiency Status

	% Beginning	% Nearing	% Proficient	% Advanced	% Below Proficient	% Proficient and Above	Students	Teachers
Level 0								
2007	14.4%	33.1%	45.1%	7.4%	47.5%	52.5%	501	30
2008	11.8%	41.3%	39.9%	7.0%	53.1%	46.9%		
Level 1								
2007	11.6%	31.6%	48.9%	7.8%	43.3%	56.7%	7274	431
2008	9.7%	38.6%	44.6%	7.1%	48.3%	51.7%		
Level 2 PDD PASS								
2007	11.5%	30.9%	49.6%	8.1%	42.3%	57.7%	5562	321
2008	9.2%	35.5%	46.5%	8.8%	44.7%	55.3%		
Level 2 PDD DID NOT PASS								
2007	12.2%	31.4%	51.3%	5.1%	43.6%	56.4%	392	25
2008	12.0%	37.2%	45.4%	5.4%	49.2%	50.8%		
Level 2 Non-PDD								
2007	10.5%	30.9%	50.2%	8.4%	41.4%	58.6%	15491	897
2008	8.5%	34.8%	47.0%	9.7%	43.3%	56.7%		
Level 3 PDD								
2007	9.6%	27.9%	52.0%	10.5%	37.5%	62.5%	2010	113
2008	6.5%	32.2%	50.4%	10.8%	38.7%	61.2%		
Level 3 Non-PDD								
2007	9.0%	28.9%	52.5%	9.6%	37.9%	62.1%	8951	519
2008	7.4%	32.5%	49.0%	11.2%	39.8%	60.2%		

Source: LFC analysis of PED data

TABLE 14: Math Scale Score Change by Proficiency Levels

License (Content)	08 Status	N Students	Score Change	% Declining	% Improving
Level 3 PDD	Beginning	119	9.1	56.3%	NA
Level 3 NonPDD	Beginning	626	8.6	58.1%	NA
Level 2 PDD	Beginning	428	10.3	55.6%	NA
Level 2 Did not pass	Beginning	33	14.5	54.5%	NA
Level 2 NonPDD	Beginning	1279	8.8	56.8%	NA
Level 1	Beginning	717	7.8	57.6%	NA
Level 0	Beginning	72	9.5	56.9%	NA
Level 3 PDD	Nearing	910	20.5	17.7%	7.7%
Level 3 NonPDD	Nearing	4293	17.0	24.5%	6.4%
Level 2 PDD	Nearing	2812	18.8	21.3%	7.5%
Level 2 Did not pass	Nearing	230	16.0	29.1%	5.7%
Level 2 NonPDD	Nearing	7930	17.9	21.8%	7.3%
Level 1	Nearing	3833	16.8	23.7%	7.7%
Level 0	Nearing	280	15.2		
Level 3 PDD	Proficient	766	24.6	8.2%	22.7%
Level 3 NonPDD	Proficient	3163	24.1	8.5%	20.7%
Level 2 PDD	Proficient	1877	24.5	8.6%	22.1%
Level 2 Did not pass	Proficient	117	22.7	10.3%	17.1%
Level 2 NonPDD	Proficient	5106	24.1	9.0%	22.0%
Level 1	Proficient	2303	22.3	9.6%	20.3%
Level 0	Proficient	127	20.4	22.8%	12.6%
Level 3 PDD	Advanced	213	32.1	NA	47.9%
Level 3 NonPDD	Advanced	863	32.8	NA	45.8%
Level 2 PDD	Advanced	445	33.8	NA	50.3%
Level 2 Did not pass	Advanced	12	28.1	NA	41.7%
Level 2 NonPDD	Advanced	1176	34.2	NA	25.3%
Level 1	Advanced	421	29.3	NA	44.4%
Level 0	Advanced	22	23.9	NA	31.8%

Source: LFC analysis of PED data

TABLE 15: Reading Scale Score Change by Proficiency Levels

License (Content)	08 Status	N Students	Score Change	% Declining	% Improving
Level 3 PDD	Beginning	130	2.6	34.6%	NA
Level 3 NonPDD	Beginning	658	1.0	36.0%	NA
Level 2 PDD	Beginning	513	-0.6	38.4%	NA
Level 2 Did not pass	Beginning	47	-5.9	51.1%	NA
Level 2 NonPDD	Beginning	1323	-0.5	35.8%	NA
Level 1	Beginning	707	-0.4	35.4%	NA
Level 0	Beginning	59	0.1	30.5%	NA
Level 3 PDD	Nearing	648	12.7	29.5%	15.0%
Level 3 NonPDD	Nearing	2903	10.5	32.4%	11.9%
Level 2 PDD	Nearing	1973	11.5	30.8%	14.9%
Level 2 Did not pass	Nearing	146	11.6	35.6%	16.4%
Level 2 NonPDD	Nearing	5385	10.4	30.4%	13.3%
Level 1	Nearing	2807	9.9	32.7%	13.0%
Level 0	Nearing	207	11.8		
Level 3 PDD	Proficient	1012	18.4	10.0%	16.3%
Level 3 NonPDD	Proficient	4386	18.2	9.6%	17.9%
Level 2 PDD	Proficient	2588	17.8	9.3%	19.0%
Level 2 Did not pass	Proficient	178	18.4	6.2%	19.1%
Level 2 NonPDD	Proficient	7278	18.3	9.1%	18.8%
Level 1	Proficient	3245	17.7	17.8%	9.9%
Level 0	Proficient	200	18.2		
Level 3 PDD	Advanced	218	28.4	NA	50.5%
Level 3 NonPDD	Advanced	999	29.0	NA	57.8%
Level 2 PDD	Advanced	488	27.6	NA	56.8%
Level 2 Did not pass	Advanced	21	32.4	NA	61.9%
Level 2 NonPDD	Advanced	1505	29.8	NA	58.5%
Level 1	Advanced	515	28.4	NA	55.0%
Level 0	Advanced	35	28.7	NA	51.4%

Source: LFC analysis of PED data

Table 16: T-test for Statistical Significance of Average Scale Score Changes
80-100% of Students Eligible for Free or Reduced Lunch

≥80 <100 Quintile	N	Avg	StandDev	P-Value	Statistically Significant
Level 3 PDD (MATH)	43	24.94	13.7	0.0014	Yes
Level 3 GF (MATH)	159	18.50	10.94		
Level 3 PDD (MATH)	43	24.94	13.7	0.2556	No
Level 2 PDD (MATH)	144	22.45	12.21		
Level 3 PDD (MATH)	43	24.94	13.7	0.0186	Yes
Level 2 GF (MATH)	392	20.69	10.90		
Level 3 PDD (MATH)	43	24.94	13.7	0.0087	Yes
Level 1 (MATH)	209	19.88	10.90		
Level 3 GF (MATH)	159	18.50	10.94	0.0032	Yes
Level 2 PDD (MATH)	144	22.45	12.21		
Level 3 GF (MATH)	159	18.50	10.94	0.0332	Yes
Level 2 GF (MATH)	392	20.69	10.90		
Level 3 GF (MATH)	159	18.50	10.94	0.2305	No
Level 1 (MATH)	209	19.88	10.90		
Level 2 PDD (MATH)	144	22.45	12.21	0.1095	No
Level 2 GF (MATH)	392	20.69	10.90		
Level 2 PDD (MATH)	144	22.45	12.21	0.039	Yes
Level 1 (MATH)	209	19.88	10.90		
Level 2 GF (MATH)	392	20.69	10.90	0.3859	No
Level 1 (MATH)	209	19.88	10.90		

Source: LFC analysis of PED data

Table 17: T-test for Statistical Significance of Average Scale Score Changes
60-80% of Students Eligible for Free or Reduced Lunch

≥60 <80 Quintile	N	Avg	StandDev	P-Value	Statistically Significant
Level 3 PDD (MATH)	16	22.09	10.15	0.7402	No
Level 3 GF (MATH)	103	21.05	11.85		
Level 3 PDD (MATH)	16	22.09	10.15	0.8888	No
Level 2 PDD (MATH)	63	21.67	10.82		
Level 3 PDD (MATH)	16	22.09	10.15	0.358	No
Level 2 GF (MATH)	194	19.28	11.84		
Level 3 PDD (MATH)	16	22.09	10.15	0.0483	Yes
Level 1 (MATH)	82	17.17	8.77		
Level 3 GF (MATH)	103	21.05	11.85	0.7359	No
Level 2 PDD (MATH)	63	21.67	10.82		
Level 3 GF (MATH)	103	21.05	11.85	0.2212	No
Level 2 GF (MATH)	194	19.28	11.84		
Level 3 GF (MATH)	103	21.05	11.85	0.0143	Yes
Level 1 (MATH)	82	17.17	8.77		
Level 2 PDD (MATH)	63	21.67	10.82	0.1566	No
Level 2 GF (MATH)	194	19.28	11.84		
Level 2 PDD (MATH)	63	21.67	10.82	0.0064	Yes
Level 1 (MATH)	82	17.17	8.77		
Level 2 GF (MATH)	194	19.28	11.84	0.1473	No
Level 1 (MATH)	82	17.17	8.77		

Source: LFC analysis of PED data

Table 18: T-test for Statistical Significance of Average Scale Score Changes
40-60% of Students Eligible for Free or Reduced Lunch

≥40 <60 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (MATH)	19	26.99	10.98	0.1025	No
Level 3 GF (MATH)	128	22.43	11.33		
Level 3 PDD (MATH)	19	26.99	10.98	0.0006	Yes
Level 2 PDD (MATH)	59	18.80	7.81		
Level 3 PDD (MATH)	19	26.99	10.98	0.0215	Yes
Level 2 GF (MATH)	156	21.03	10.52		
Level 3 PDD (MATH)	19	26.99	10.98	0.0001	Yes
Level 1 (MATH)	74	14.91	11.23		
Level 3 GF (MATH)	128	22.43	11.33	0.0271	Yes
Level 2 PDD (MATH)	59	18.80	7.81		
Level 3 GF (MATH)	128	22.43	11.33	0.2821	No
Level 2 GF (MATH)	156	21.03	10.52		
Level 3 GF (MATH)	128	22.43	11.33	0.0001	Yes
Level 1 (MATH)	74	14.91	11.23		
Level 2 PDD (MATH)	59	18.80	7.81	0.1403	No
Level 2 GF (MATH)	156	21.03	10.52		
Level 2 PDD (MATH)	59	18.80	7.81	0.0255	Yes
Level 1 (MATH)	74	14.91	11.23		
Level 2 GF (MATH)	156	21.03	10.52	0.0001	Yes
Level 1 (MATH)	74	14.91	11.23		

Source: LFC analysis of PED data

Table 19: T-test for Statistical Significance of Average Scale Score Changes
20-40% of Students Eligible for Free or Reduced Lunch

≥20 <40 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (MATH)	21	21.87	7.85	0.4885	No
Level 3 GF (MATH)	74	20.24	9.88		
Level 3 PDD (MATH)	21	21.87	7.85	0.5846	No
Level 2 PDD (MATH)	31	20.64	7.95		
Level 3 PDD (MATH)	21	21.87	7.85	0.8212	No
Level 2 GF (MATH)	96	21.38	9.20		
Level 3 PDD (MATH)	21	21.87	7.85	0.2678	No
Level 1 (MATH)	38	19.30	8.75		
Level 3 GF (MATH)	74	20.24	9.88	0.8421	No
Level 2 PDD (MATH)	31	20.64	7.95		
Level 3 GF (MATH)	74	20.24	9.88	0.4391	No
Level 2 GF (MATH)	96	21.38	9.20		
Level 3 GF (MATH)	74	20.24	9.88	0.6216	No
Level 1 (MATH)	38	19.30	8.75		
Level 2 PDD (MATH)	31	20.64	7.95	0.6885	No
Level 2 GF (MATH)	96	21.38	9.20		
Level 2 PDD (MATH)	31	20.64	7.95	0.5121	No
Level 1 (MATH)	38	19.30	8.75		
Level 2 GF (MATH)	96	21.38	9.20	0.2339	No
Level 1 (MATH)	38	19.30	8.75		

Source: LFC analysis of PED data

Table 20: T-test for Statistical Significance of Average Scale Score Changes <20% of Students Eligible for Free or Reduced Lunch

<20 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (MATH)	14	20.99	9.50	0.8706	No
Level 3 GF (MATH)	55	20.41	12.35		
Level 3 PDD (MATH)	14	20.99	9.50	0.9954	No
Level 2 PDD (MATH)	24	21.01	10.49		
Level 3 PDD (MATH)	14	20.99	9.50	0.6871	No
Level 2 GF (MATH)	59	22.15	9.68		
Level 3 PDD (MATH)	14	20.99	9.50	0.3594	No
Level 1 (MATH)	28	18.01	9.97		
Level 3 GF (MATH)	55	20.41	12.35	0.8362	No
Level 2 PDD (MATH)	24	21.01	10.49		
Level 3 GF (MATH)	55	20.41	12.35	0.4026	No
Level 2 GF (MATH)	59	22.15	9.68		
Level 3 GF (MATH)	55	20.41	12.35	0.3759	No
Level 1 (MATH)	28	18.01	9.97		
Level 2 PDD (MATH)	24	21.01	10.49	0.6362	No
Level 2 GF (MATH)	59	22.15	9.68		
Level 2 PDD (MATH)	24	21.01	10.49	0.296	No
Level 1 (MATH)	28	18.01	9.97		
Level 2 GF (MATH)	59	22.15	9.68	0.0684	No
Level 1 (MATH)	28	18.01	9.97		

Source: LFC analysis of PED data

Table 21: T-test for Statistical Significance of Average Scale Score Changes
80-100% of Students Eligible for Free or Reduced Lunch

≥80 <100 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (READ)	43	16.79	9.82	0.227	No
Level 3 GF (READ)	159	14.69	10.15		
Level 3 PDD (READ)	43	16.79	9.82	0.5028	No
Level 2 PDD (READ)	144	15.69	9.31		
Level 3 PDD (READ)	43	16.79	9.82	0.1246	No
Level 2 GF (READ)	392	14.28	10.19		
Level 3 PDD (READ)	43	16.79	9.82	0.1182	No
Level 1 (READ)	209	14.14	10.15		
Level 3 GF (READ)	159	14.69	10.15	0.3738	No
Level 2 PDD (READ)	144	15.69	9.31		
Level 3 GF (READ)	159	14.69	10.15	0.6685	No
Level 2 GF (READ)	392	14.28	10.19		
Level 3 GF (READ)	159	14.69	10.15	0.6069	No
Level 1 (READ)	209	14.14	10.15		
Level 2 PDD (READ)	144	15.69	9.31	0.147	No
Level 2 GF (READ)	392	14.28	10.19		
Level 2 PDD (READ)	144	15.69	9.31	0.1457	No
Level 1 (READ)	209	14.14	10.15		
Level 2 GF (READ)	392	14.28	10.19	0.8724	No
Level 1 (READ)	209	14.14	10.15		

Source: LFC analysis of PED data

Table 22: T-test for Statistical Significance of Average Scale Score Changes

60-80% of Students Eligible for Free or Reduced Lunch

≥60 <80 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (READ)	16	15.59	10.46	0.8571	No
Level 3 GF (READ)	103	15.14	9.09		
Level 3 PDD (READ)	16	15.59	10.46	0.4256	No
Level 2 PDD (READ)	63	13.21	10.65		
Level 3 PDD (READ)	16	15.59	10.46	0.5588	No
Level 2 GF (READ)	194	14.09	9.80		
Level 3 PDD (READ)	16	15.59	10.46	0.3428	No
Level 1 (READ)	82	12.75	10.98		
Level 3 GF (READ)	103	15.14	9.09	0.2157	No
Level 2 PDD (READ)	63	13.21	10.65		
Level 3 GF (READ)	103	15.14	9.09	0.3684	No
Level 2 GF (READ)	194	14.09	9.80		
Level 3 GF (READ)	103	15.14	9.09	0.107	No
Level 1 (READ)	82	12.75	10.98		
Level 2 PDD (READ)	63	13.21	10.65	0.545	No
Level 2 GF (READ)	194	14.09	9.80		
Level 2 PDD (READ)	63	13.21	10.65	0.8004	No
Level 1 (READ)	82	12.75	10.98		
Level 2 GF (READ)	194	14.09	9.80	0.3177	No
Level 1 (READ)	82	12.75	10.98		

Source: LFC analysis of PED data

Table 23: T-test for Statistical Significance of Average Scale Score Changes

40-60% of Students Eligible for Free or Reduced Lunch

≥40 <60 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (READ)	19	17.71	9.24	0.9517	No
Level 3 GF (READ)	128	17.85	9.40		
Level 3 PDD (READ)	19	17.71	9.24	0.1024	No
Level 2 PDD (READ)	59	13.77	8.97		
Level 3 PDD (READ)	19	17.71	9.24	0.5046	No
Level 2 GF (READ)	156	16.14	9.71		
Level 3 PDD (READ)	19	17.71	9.24	0.0737	No
Level 1 (READ)	74	12.75	10.98		
Level 3 GF (READ)	128	17.85	9.40	0.0057	Yes
Level 2 PDD (READ)	59	13.77	8.97		
Level 3 GF (READ)	128	17.85	9.40	0.1352	No
Level 2 GF (READ)	156	16.14	9.71		
Level 3 GF (READ)	128	17.85	9.40	0.0006	Yes
Level 1 (READ)	74	12.75	10.98		
Level 2 PDD (READ)	59	13.77	8.97	0.1046	No
Level 2 GF (READ)	156	16.14	9.71		
Level 2 PDD (READ)	59	13.77	8.97	0.5653	No
Level 1 (READ)	74	12.75	10.98		
Level 2 GF (READ)	156	16.14	9.71	0.0186	Yes
Level 1 (READ)	74	12.75	10.98		

Source: LFC analysis of PED data

Table 24: T-test for Statistical Significance of Average Scale Score Changes

20-40% of Students Eligible for Free or Reduced Lunch

≥20 <40 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (READ)	21	18.08	8.97	0.3222	No
Level 3 GF (READ)	74	15.68	9.96		
Level 3 PDD (READ)	21	18.08	8.97	0.1599	No
Level 2 PDD (READ)	31	14.55	8.61		
Level 3 PDD (READ)	21	18.08	8.97	0.8247	No
Level 2 GF (READ)	96	17.57	9.65		
Level 3 PDD (READ)	21	18.08	8.97	0.0806	No
Level 1 (READ)	38	14.20	7.46		
Level 3 GF (READ)	74	15.68	9.96	0.5828	No
Level 2 PDD (READ)	31	14.55	8.61		
Level 3 GF (READ)	74	15.68	9.96	0.2135	No
Level 2 GF (READ)	96	17.57	9.65		
Level 3 GF (READ)	74	15.68	9.96	0.4216	No
Level 1 (READ)	38	14.20	7.46		
Level 2 PDD (READ)	31	14.55	8.61	0.1228	No
Level 2 GF (READ)	96	17.57	9.65		
Level 2 PDD (READ)	31	14.55	8.61	0.857	No
Level 1 (READ)	38	14.20	7.46		
Level 2 GF (READ)	96	17.57	9.65	0.0552	No
Level 1 (READ)	38	14.20	7.46		

Source: LFC analysis of PED data

Table 25: T-test for Statistical Significance of Average Scale Score Changes

<20% of Students Eligible for Free or Reduced Lunch

<20 Quintile	N	Avg	StandDev	P-Value	Statistically Significant?
Level 3 PDD (READ)	14	15.27	13.26	0.7812	No
Level 3 GF (READ)	55	16.28	11.80		
Level 3 PDD (READ)	14	15.27	13.26	0.9475	No
Level 2 PDD (READ)	24	15.03	9.06		
Level 3 PDD (READ)	14	15.27	13.26	0.6624	No
Level 2 GF (READ)	59	16.65	9.89		
Level 3 PDD (READ)	14	15.27	13.26	0.542	No
Level 1 (READ)	28	17.26	7.75		
Level 3 GF (READ)	55	16.28	11.80	0.6452	No
Level 2 PDD (READ)	24	15.03	9.06		
Level 3 GF (READ)	55	16.28	11.80	0.856	No
Level 2 GF (READ)	59	16.65	9.89		
Level 3 GF (READ)	55	16.28	11.80	0.6921	No
Level 1 (READ)	28	17.26	7.75		
Level 2 PDD (READ)	24	15.03	9.06	0.4913	No
Level 2 GF (READ)	59	16.65	9.89		
Level 2 PDD (READ)	24	15.03	9.06	0.3439	No
Level 1 (READ)	28	17.26	7.75		
Level 2 GF (READ)	59	16.65	9.89	0.7749	No
Level 1 (READ)	28	17.26	7.75		

Source: LFC analysis of PED data

Table 26. 2007-08 Percentage of Students Scoring Proficient or Above and Associated Achievement Gaps

	2007-08 Fourth Grade Reading	2007-08 Fourth Grade Math	2007-08 Eighth Grade Reading	2007-08 Eighth Grade Math
Non Economically Disadvantaged	70.68%	58.05%	76.78%	53.18%
Economically Disadvantaged	42.56%	31.04%	55.01%	26.56%
Gap	28.12%	27.01%	21.77%	26.62%
Caucasian	68.71%	55.60%	76.59%	54.72%
Hispanic	44.85%	33.30%	57.73%	29.23%
Gap	23.87%	22.30%	18.86%	25.49%
White Non-Free/Reduced Lunch	77.30%	65.29%	82.00%	62.22%
White Free/Reduced Lunch	56.17%	41.44%	65.88%	39.86%
Gap	21.13%	23.85%	16.12%	22.36%
Hispanic Non-Free/Reduced Lunch	62.46%	48.58%	70.52%	42.89%
Hispanic Free/Reduced Lunch	40.73%	29.73%	53.31%	24.51%
Gap	21.73%	18.85%	17.21%	18.38%
Am Indian Non-Free/Reduced Lunch	51.14%	35.51%	68.04%	35.89%
Am Indian Free/Reduced Lunch	32.57%	22.89%	50.11%	19.98%
Gap	18.56%	12.63%	17.93%	15.91%
African American Non Free/Reduced Lunch	61.76%	46.47%	74.32%	39.30%
African American Free/Reduced Lunch	41.77%	26.49%	54.57%	25.13%
Gap	20.00%	19.98%	19.75%	14.17%

Source: PED

Table 27. Western States 4th Grade Math Socioeconomic Achievement Gap (NAEP Proficient or Above)

STATE	2003 F/R INELIGIBLE	2003 F/R ELIGIBLE	2003 GAP	2007 F/R INELIGIBLE	2007 F/R ELIGIBLE	2007 GAP	03-07 GAP CHANGE
NM*	31%	11%	20%	43%	16%	27%	7%
UT*	37%	20%	17%	48%	25%	23%	6%
CO*	43%	14%	29%	55%	21%	34%	5%
OR*	40%	19%	21%	47%	21%	26%	5%
ID*	38%	20%	18%	50%	27%	23%	5%
NV*	33%	11%	22%	42%	16%	26%	4%
AZ*	39%	12%	27%	46%	15%	31%	4%
WA*	48%	20%	28%	56%	26%	30%	2%
CA	41%	11%	30%	46%	16%	30%	0%
WY	47%	25%	22%	51%	32%	19%	-3%

*this western state increased students scoring proficient and the achievement gap

Source: NCES

Table 28: Nine Teacher Competencies and Associated PDD Strands

Strand A: Instruction (Competencies 1, 2, and 5)
Competency 1: Teacher accurately demonstrates knowledge of the content area and approved curriculum.
Competency 2: Teacher utilizes a variety of teaching methods and resources for each area taught.
Competency 5: Teacher effectively utilizes student assessment techniques and procedures.
Strand B: Student Learning (Competencies 3, 4, 6, and 7)
Competency 3: Teacher communicates with and obtains feedback from students in a manner that enhances student learning and understanding.
Competency 4: Teacher comprehends the principles of student growth, development, and learning, and applies them appropriately.
Competency 6: Teacher manages the educational setting in a manner that promotes positive student behavior and a safe and healthy environment.
Competency 7: The teacher recognizes student diversity and creates an atmosphere conducive to the promotion of positive student involvement and self concept.
Strand C: Professional Learning (Competencies 8 and 9)
Competency 8: Teacher demonstrates a willingness to examine and implement change, as appropriate.
Competency 9: Teacher works productively with colleagues, parents and community members.

Source: PED

2003

Several public school reforms (HB212) passed. Key provisions included:

- The three-tiered licensure system which bases minimum salary levels on demonstrated aptitude
- Requirements for standards based tests designed to meet the requirements of the federal No Child Left Behind Act
- Creating a fund to help schools
- Minimum salary levels for principals
- Creating the Office of Education Accountability within the Department of Finance and Administration

To pay for these reforms, in September 2003, voters narrowly approved a constitutional amendment to increase the annual distribution from the Land Grant Permanent Fund from 4.7% of a five year rolling average to 5% in 2004, 5.8% from 2005 to 2012, 5.5% from 2013 to 2016, and back to 5% in 2017.

2004

The Legislature increased appropriations to public schools to $2 billion and set aside $120 million in the 'Education Lockbox' to fund reforms. Initiatives included creating the Public Education Department (PED), an upgraded data system, truancy and dropout prevention, charter schools, and a career ladder for educational assistants.

2005

Educational reforms were focused on holistic educational systems, the P-20 concept. The Legislature increased public school funding to $2.1 billion, funded the next phase of the three-tiered licensure system, passed the Pre-Kindergarten Act, expanded the assessment program to include an 11th grade test, directed PED to assess districts' technology needs, and passed student health related initiatives.

2006

Reform efforts focused on college and career preparedness. The Legislature increased appropriations to public schools to $2.3 billion, funded the next phase of the three-tiered licensure system, and provided funds for salary increases for all school employees. Other actions included creating a Charter School Division and a Math and Science Bureau at PED, revising capital outlay, and extending the Kindergarten Plus Pilot program.

2007

Among other changes, the Legislature dealt with High School Redesign which increased the legal drop-out age to 18 and increased graduation requirements by requiring four years of math and English and an Advanced Placement, dual-credit, or online course. SB 211 creates a standard student identification number to be used by public schools and higher education, established a statewide teacher accountability reporting system, and allowed PED to develop rules regarding the use of unlicensed content experts. Other significant projects included passing the Cyber Academy Act to provide online courses, increases in principal salaries and evaluations tied to student achievement, the six-year K-3 pilot project, and joining the American Diploma Project. The Legislature also funded Pre-K with $14 million.

2008

Funding for public schools has increased to about $2.6 billion. While student achievement remains the focus, the Legislature placed more emphasis on complete funding of existing reforms rather than creation of additional new reforms. Appropriations included $40 million to fund an average 2% raise for all educational employees, $12.3 million to increase employer contributions into the educational retirement fund, $7.2 million to implement the K-3 plus program, $4 million for assessment and test development, $5.5 million to the school improvement framework, $2 million to the teacher professional development fund, $2 million for beginning teacher mentorship, and $2.5 million for summer institutes. The Legislature also provided additional funds for elementary physical education and school breakfast, Regional Educational Cooperatives (REC's), and the STARS system.

2009

While the economic downturn imposed limitations on state government programs and services, the Legislature managed to appropriate about $2.4 billion for public schools. With $164.7 million in stimulus funds from the American Recovery and Reinvestment Act of 2009, reductions to education amounted to less than one percent. The Legislature considered, but did not pass, a bill creating a new funding formula and a related bill to generate about $380 million in new revenue for public education. Despite the decline in revenues, the Legislature provided $8.5 million for pre-kindergarten, $8.5 million for K-3 plus, $3.5 million for school improvement, $1.8 million for advanced placement, $1 million for afterschool enrichment, $1.5 million for instructional materials for dual credit, $3.4 million for elementary breakfast, $1.5 million for beginning teacher mentorship, $2.5 million for summer institutes, and $200 thousand for a school leadership institute.

APPENDIX C: SUMMARY OF LIMITED SCHOOL VISITS

School Tours
In an effort to identify reforms or initiatives that improve student performance, I visited nine schools across the state last week. I'm glad that Sen Asbill was able to join me in Carlsbad and Sen Kernan was able to join me in Hobbs. I still have one more school to visit in Santa Fe. Many of the themes observed so far were difficult to measure precisely but were clearly evident when conversing with principals.

NMSBA data can be used to identify the best teachers.
While the current three-tiered program evaluation is not meant to evaluate individual teachers, I did ask principals if their best teachers are also the best at improving test scores. Essentially, I asked them if our narrowly defined methods of using one year of growth in the NMSBA data would produce the same list of most effective teachers as a broader teacher evaluation method that might include other concepts such as creating good citizens, inspiring students to learn, communicating with parents, and principal observations, etc. Almost all of the principals stated that the teachers with greatest success in NMSBA data would, in fact, be the best teachers overall. Many principals stated that if a teacher is effective at the broader measures of teacher effectiveness, this would translate into increases in NMSBA scores. Principals agree that effective and ineffective teachers can be found at all licensure levels.

High Performing Schools have High Expectations/Goals for all students.
Common among the top performers was the insistence that goals be set very high. Top performers expressed high expectations for **all** students in **all** subgroups. Top performers were willing to put in extra effort to improve performance for EVERY child. Top performers involve every student in setting individual performance goals. High performing schools develop multiple, more specific, high achieving goals. Their goal is not simply to make AYP. One goal for a top performing school included having all children reading at grade level. This principal stated that "we set our expectations for all children as if they are gifted."

High Performing Schools are guided by assessment data.
High performing schools were characterized by the ability to do more than collect data, but to use data to track and target individual students, plan instruction and interventions, and make grade or school wide plans for improvement.

High performing schools demonstrated greater teacher involvement in the analysis and use of the data. High performing schools met frequently (monthly if not bi-weekly) to review data for each student and develop action plans (often interventions) to address individual student needs for improvement. The analysis and use of data is routine and institutionalized in high performing schools. High performing schools also use data from multiple assessments or programs (for example Accelerated Reader, SME Lab, Dibbles, DRA, NWEA) to validate results and make predictions on the NMSBA. For example, many of the schools use the Accelerated Reader program (AR), but in the high performing schools, each student has an AR goal for the year and librarians track how many points have been earned by grade and how many books have been read and tested on.

High Performing Schools have Leaders with "Can Do" attitudes.
Principals at high performing schools focus on aspects they control rather than what they can not control. These principals did not accept common objections from teachers like 'the parents just aren't involved'. In general, high performing schools are more likely to discuss higher goals for the future while low performing schools often offered explanations for not meeting goals in the past. Principals at high performing schools described their efforts to accomplish goals, whereas low performing schools were more likely to speak in terms of resources they were lacking. Decisions at top performing schools were student driven, not teacher driven.

Principals at high performing schools were in the classrooms frequently. While these principals are hugely important, they are not the only leaders in the school. These schools look to hire teachers who are willing to put forth extra effort. For example Barry opens the computer lab early in the morning for students to practice and teachers will stay late when needed. Monterrey operates a summer reading program where teachers volunteer time. Teacher leaders were critical to ensure that common planning time was used effectively.

High Performing Schools get maximum benefit from teacher collaboration.
High performing schools take teacher collaboration seriously and it is structured to help accomplish school wide goals, not just a time for teachers to talk to one another. (When asked, "What do teachers discuss during common planning time?" one principal at a low performing school responded that the teachers were discussing what they watched on TV last night) High performing schools are more likely to have teachers meet in content areas to identify skills that need improvements, and grade levels to identify effective instructional practices.

High Performing Schools are obsessive about instructional time.
High performing schools made better use of common planning time, often during early release on Wednesday. High performing schools also arranged schedules to maximize instructional time.

Other noteworthy observations:
Unions:
In general, top performing schools were less supportive of strong unions. Comments suggested strong union districts had more difficulty in getting teachers to put in extra effort beyond the typically school day.

Reading and Math teacher specialists/Instructional coaches:
It was NOT the case that all top performers had Reading and Math specialists. Some top performers (Carlsbad) did not employ them while some of the low performers (Hobbs) did.

Professional Development (PD) typically district driven:
Principals noted that PD tends to be district driven; the decisions about how to use PD days is often made for principals. There may be limited opportunities for principals or teachers to assess themselves and request professional development to address their needs.

EPSS not effective:
General consensus that the EPSS is not a very effective accountability tool, format is repetitive and unwieldy.

Technology:
Schools are using a variety of new "Smart Board" type Technologies:

Interwrite pads (http://www.interwritelearning.de/products/pad/detail.html)

Mimio (http://www.mimio.com/index.asp)

Prometheon (http://www.prometheanworld.com/server.php?show=nav.16)

Programs:
Hobbs schools adopted the Core Knowledge System which seems to have a greater emphasis on arts and culture, etc: http://www.coreknowledge.org/CK/about/index.htm
The extent to which this impacts their NMSBA scores is uncertain.

Barry in Clovis has seen positive results from their Imagine it Reading program: http://www.imagineitreading.com/flash/index.php

Further study:
Some of the top performing schools received Title 1 funds and some of the low performing schools did not receive Title 1. I would like to further study how these funds are used and assess the impact of receiving these funds, or not receiving these funds, on student performance.

Conclusion:
Much of the success of the top performers could be attributable to the positive and professional attitude of staff, strong leadership, and effective use of assessment data. Please contact me if you would like more information or examples of concepts discussed.

APPENDIX D: SYSTEMS TO EVALUATE TEACHER PERFORMANCE

Recent discussion about funding public education and educational reforms nationally and in New Mexico has centered on teacher effectiveness. There is a growing focus in education to think of teacher effectiveness in terms of outputs, as well as inputs. A number of groups, most recently the National Title I Association, are calling on states to investigate output-based measures of teacher effectiveness, rather than input-based proxies, such as the existing "highly qualified" definition in the NCLB law.[5]

The recently enacted economic-stimulus bill requires every state to take steps to improve teacher effectiveness, as well as to tackle one of the most pervasive problems in K-12 education: inequities in access to top teaching talent for poor and minority children. The U.S. Department of Education cited California, Illinois, Michigan, Nebraska, New Mexico, North Dakota, Pennsylvania, Puerto Rico, and Wyoming for not updating, monitoring, or reporting progress on their state plans.[6] The federal Teacher Incentive Fund, a performance-pay program, has promoted interest in using test scores to estimate teacher effectiveness.

Teacher resource allocation has become a greater concern as policy makers are made aware that our best teachers are often not employed where they are most needed. Often more experienced teachers move to schools with a lower percentage of poverty students. Quality instruction is vital to increasing student achievement. A balanced accountability system includes outputs as well as inputs. Ideally, an accountability system would support instruction by evaluating teachers on measures of student learning and achievement as well as on their knowledge, skills, and abilities.

Today, there is not a method with unanimous support that is used to measure the impact of an individual teacher on student learning. There is, however, general agreement that a snapshot of standardized test scores alone do not provide a comprehensive representation of how a teacher contributes to student achievement. A single test score should not be used as the sole measure of teacher effectiveness, but a well designed and easily understood growth model that measures a teacher's contribution to students' progress increases fairness. This study explores how standardized assessments can be used to promote teacher accountability.

This is not a holistic picture of teacher impact. This evaluation employs a simple growth model that measures only one year of growth, only a few grades, and only two subject areas. However, this statewide growth data is the first study of this kind using data to identify variation between licensure levels. While the study does, infrequently, present data for individual teachers, the focus is on the evaluation of the licensure system itself. For the purpose of this evaluation, quality instruction is viewed as contributions to student learning as measured by changes from assessment scores at the end of one year to assessment scores at the end of the next year.

For content areas where the NMSBA is administered and growth information is available, teachers could be accountable for the students in their classrooms making at least one year's worth of growth each year from the students' starting point. The teachers' performance could be calculated using a running multi-year average of the teacher's classes. In subjects where the NMSBA is not given, teachers and principals should develop student learning goals that incorporate short-cycle assessments or other measures of student performance.

[5] Education Week article. 3/11/09

[6] SOURCES: U.S. Department of Education; Education Week

APPENDIX E: RESEARCH METHODS

The New Mexico Standards Based Assessment (NMSBA) is the state assessment test given annually to third- through eighth- grade students and again to eleventh-graders. The analysis eliminated test scores of those students who take a different test than the English version of the NMSBA, most notably the New Mexico Alternate Performance Assessment (NMAPA) and Spanish versions of the NMSBA because of different cut scores which limit comparisons. Excluding these students from the analysis resulted in approximately 4% of the student population being excluded in each SY.

Statewide test scores of all students and selected subgroups with teachers who have passed the PDD process are compared to scores of students with teachers who have not gone through the process. The objective focuses on growth in student performance for teachers who have gone through the PDD process to advance from Level 2 to Level 3. Student scores for these teachers will be benchmarked against students' scores from five other groups: 1) Level 2 teachers who did not pass or go through the PDD process, 2) level two teachers who went through the PDD process, 3) Level 3 teachers who did not go through the PDD process. 4) Level 1 teachers, and 5) State averages. The evaluation looks at improvement in scale scores from year to year as well as percentage of students meeting proficiency. The evaluation will attempt to control for other variables, such as student low income status and other demographics, by comparing the students' score from 2007 to their score in 2008. Essentially, the objective will identify if teachers become more effective at improving test scores as they advance through the three-tiered licensing system. The study is an initial attempt to produce a verifiable analysis that indicates the extent to which the Three-Tiered system and the PDD process reward and reflect quality instruction.

The achievement gap update focuses on the performance of low income students as compared to non-economically disadvantaged students. Other demographic variables are analyzed to compare the gaps in performance between low income minorities and their non-economically disadvantaged peers. For example, data is analyzed to determine if, and to what extent, non-low income Hispanics outperform low income Hispanics. The results demonstrate the extent to which economically disadvantaged students struggle relative to other economic and/or ethnic subgroups and have implications for PED and districts in determining the content areas and subgroups where resources can be best targeted to close the gaps.

A history of the achievement gap since the three-tiered teacher licensure system was implemented is provided. This evaluation includes a comparison of the achievement gap in SYs 2004-05 and 2007-08 and also an analysis of longitudinal data dating back to 2005 to illustrate achievement gap changes among cohorts. The history of the achievement gap since the 2004-05 SY also includes disaggregate data on the achievement gap including economically disadvantaged and non-economically disadvantaged Hispanic/Latino, Native American, and white students. Current student achievement is compared to achievement in the 2004-05 SY to reveal changes in each subgroup of students.