Report
to
The LEGISLATIVE FINANCE COMMITTEE

Public Education Department
Promoting Effective Teaching in New Mexico
November 15, 2012

Report #12-12

LEGISLATIVE FINANCE COMMITTEE

Senator John Arthur Smith, Chairman
Representative Luciano "Lucky" Varela, Vice-Chairman
Senator Sue Wilson Beffort
Senator Pete Campos
Senator Carlos R. Cisneros
Representative William "Bill" J. Gray
Senator Stuart Ingle
Representative Rhonda S. King
Representative Larry A. Larrañaga
Senator Carroll H. Leavell
Senator Mary Kay Papen
Representative Henry "Kiki" Saavedra
Representative Nick L. Salazar
Representative Edward C. Sandoval
Senator John Sapien
Representative Don L. Tripp
Representative James P. White

DIRECTOR

David Abbey

DEPUTY DIRECTOR FOR PROGRAM EVALUATION

Charles Sallee

PROGRAM EVALUATION TEAM

Jeff Canney, CGFM
Jon R. Courtney, Ph.D.
Valerie Crespin-Trujillo
Jack Evans
Brenda Fresquez, CICA
Pamela Galbraith
Maria D. Griego
Rachel Mercer-Smith
Matthew Pahl
Michael Weinberg, Ed.D.

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

November 15, 2012

Ms. Hanna Skandera, Secretary-Designate
Public Education Department
Jerry Apodaca Education Building
300 Don Gaspar Avenue
Santa Fe, NM 87501

Dear Ms. Skandera:

On behalf of the Legislative Finance Committee (Committee), I am pleased to transmit the evaluation, *Effective Use of Student Test Data to Assess & Improve Teacher Evaluation.* The program evaluation team assessed the three-tiered licensure system; value-added models; and resource allocation from the state funding formula. The report will be presented to the Committee on November 15th, 2012. Exit conferences were conducted with the Public Education Department on November 7th, 2012 to discuss the contents of the report. The Committee would like a plan to address the recommendations within this report within 60 days from the date of the hearing.

I believe this report addresses issues the Committee asked us to review and hope New Mexico's public education system benefits from our efforts. We very much appreciate the cooperation and assistance we received from your staff.

Sincerely,

David Abbey, Director

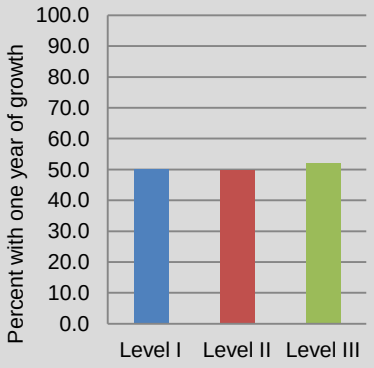
Cc: Senator John Arthur Smith, Chairman, Legislative Finance Committee
Representative Luciano "Lucky" Varela, Vice-Chairman, Legislative Finance Committee
Representative Henry "Kiki" Saavedra, Legislative Finance Committee
Representative Rick Miera, Chairman, Legislative Education Study Committee
Ms. Yolanda Berumen-Deines, Secretary, Children, Youth, and Families Department

Table of Contents

Page No.

Students in high-performing teachers' classrooms gain far more than a year's worth of academic growth, increasing an average of three points on the reading SBA and five points on the math SBA.

Percent of Students With One Year of Growth (SBA Math Gain Scores)



Source: LFC Analysis of PED Data

Nearly 30 percent of level III teachers, the state's highest paid, are in the bottom quartile of teacher performance in math and reading.

Quality teaching is the most influential school factor affecting academic success. States and school districts across the country increasingly recognize this and create incentives to improve teaching quality. One such effort involves using student performance information, through value-added models (VAM), to evaluate teaching effectiveness.

In 2003, New Mexico introduced the three-tiered system to increase the recruitment and retention of quality teachers to improve student achievement. The system created a three-level career ladder for teachers to ascend based on experience, leadership, and skills. Movement up a level results in pay increases of $10 thousand. Previous evaluations of the three-tiered system confirmed the system decreasing widespread teacher shortages, reducing unqualified teachers, and improving teacher pay.

Student performance, however, has not improved with taxpayer investments in teacher pay. A 2009 Legislative Finance Committee (LFC) evaluation using one year of performance data confirmed small differences in performance despite large differences in pay among teachers and offered solutions for improvement. The recommendations were not implemented. Since that time, nearly 6,000 teachers advanced to new license levels, receiving $59 million in mandatory salary increases.

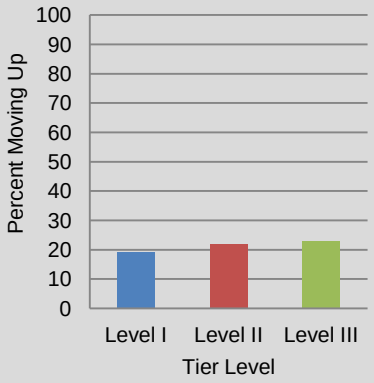
This evaluation assessed the status of the system since the majority of teachers have advanced at least one level and reviewed options for using VAMs to identify effective teaching. The evaluation used multiple years of student and teacher data to assess the performance of New Mexico's fourth through eighth grade teachers and partnered with researchers at the University of New Mexico to model how student populations influence VAM calculations.

Student performance within teacher licensure levels and between licensure levels suggests the local and state evaluation systems are not screening teachers for their effectiveness in the classroom. The difference in performance between teachers of each of the three levels is small. For example, 50 percent of students taught by level I teachers achieved a year's worth of growth in math in 2012, compared to 52 percent of students of level III teachers. Furthermore, each licensure level has high and low performing teachers; in 2012, nearly 30 percent of the lowest performing reading and math teachers in the state had a level III license. These teachers can maintain their level, including those grandfathered into the system, for the rest of their careers because the local evaluation and state license renewal process lacks factors for student achievement.

The three-tiered system continues to offer a solid framework to align resources to performance, but student achievement must be better incorporated into the process. If modified, student achievement could be a data-driven concern for all teachers and serve as a way to reward the state's best teachers and intervene for struggling teachers. While lack of

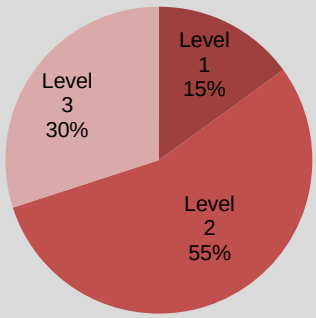
Unlike in 2003, when standardized testing was relatively new, the state now has many years worth of student and teacher data to be incorporated into the three-tiered system and funding formula.

Percent of Students Moving Up At Least One Proficiency Level in SBA Reading Scores, SY12



Source: LFC Analysis of PED Data

License Levels as a Proportion of Low Performing Reading Teachers SY12



Source: LFC Analysis of PED Data

longitudinal data made it difficult to use student performance in teacher evaluations when the three-tiered system was implemented in 2003, the state now has the resources and expertise to incorporate that information. Properly implemented, VAMs can identify teachers for advancement; their complexity, however, limits VAMs role in annual local evaluation of teachers.

PED has sought, through rule, to improve the local evaluation component of the three-tiered system and initiated two task forces to examine how to incorporate student achievement, including using VAMs, into a new system. However, statutory changes not addressed by the new PED rule are necessary to reform local evaluations and the state licensure system.

Finally, the state has not incorporated the three-tiered system into the funding formula. Instead, the formula uses a district-wide training and experience (T&E) factor, even though districts with high T&E values do not regularly achieve better performance than those with low T&E values. As currently structured, T&E widens the achievement gap by providing more funding for more affluent school districts.

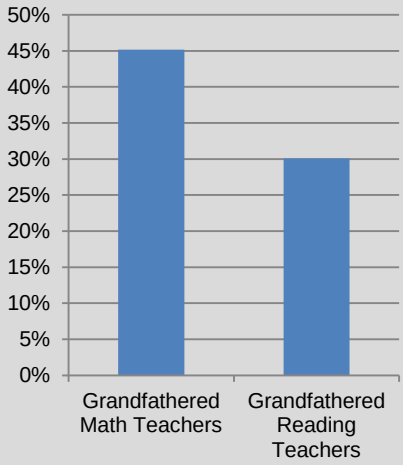
KEY FINDINGS

New Mexico's three-tiered career ladder system does not align pay with student achievement. Student performance within teacher licensure levels and between licensure levels suggests local and state evaluation systems are not screening teachers for effectiveness in the classroom. The difference in performance between teachers of each of the three licensure levels is small, with many high and low-performing teachers at each level. Teachers maintain levels throughout their careers because student achievement is not factored into licensure renewal. Establishing expectations for student achievement in the local and state evaluation systems will better align pay with student achievement.

Improving student achievement was a key policy goal of implementing the three-tiered system. The three-tiered system's founding legislation identifies student success as the fundamental goal of New Mexico's education system. The three-tiered system was designed to help achieve this goal by attracting, retaining, and holding accountable quality teachers.

The state has not established expectations for student achievement in evaluation of level I, II, and III teachers. Competencies used in the state and local evaluations of the three-tiered system include examples of student performance, but the evaluations have no expectations for the performance of all students, particularly on standardized tests. When the three-tiered system was established, the SBA was new and lacked longitudinal information; student performance, therefore, was not incorporated into evaluations. Teachers at different license levels achieve similar student performance, and a majority of New Mexico teachers do not feel the state evaluation process identifies effective teachers.

Grandfathered Teachers with Less than Half of Students Obtaining a Year's Worth of Growth by Subject



Source: LFC Analysis of PED Data

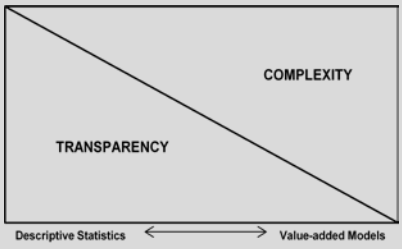
The local evaluation system does not differentiate between high and low-performing teachers or focus on student achievement. Evaluation requirements need strengthening to align with the common core, use student data, assess teacher effectiveness, and improve satisfaction among teachers. PED has not revised the competencies and requirements used in the state's professional development dossier and local evaluation system since implementation and the process could be updated to better reflect current education research.

The professional development dossier (PDD) does not effectively screen teachers for advancement, resulting in ineffective teachers receiving large pay increases. As a result, high and low performing teachers exist at each licensure level. The lack of clear and consistent performance among teachers in each licensure level shows the PDD process does not reward a teacher's impact on student achievement.

The state allows low-performing teachers to keep their license level because the state does not have a rigorous license renewal process. Grandfathered teachers, those that obtained a level II or III licensure without going through the state's PDD, continue to renew their licenses without passing the PDD. Many of these are low-performing teachers with the highest mandated salary in the state.

The three-tiered system offers a framework to align resource allocation to performance, but student achievement must be better incorporated into the process. The system offers significant salary increases and a competency structure; if modified, this system could drive student achievement across the state by setting student performance expectations. A modified three-tiered system could strategically reward the state's best teachers and provide strategic interventions for struggling teachers.

Understandability of Statistical Models



<u>When used appropriately, value-added models (VAMs) can help identify teachers' success levels and drive student achievement.</u> Different VAMs can show volatility among certain teachers. This can be minimized by using two different VAMs to form a composite score. Once these are controlled for, VAMs can be responsibly used to reward outstanding teachers and help those who are ineffective.

Value-added models are increasingly used across the country to evaluate teacher performance. As of October 2012, the U.S. Department of Education granted 33 states, including New Mexico, a waiver from some of the requirements of No Child Left Behind for changing their teacher evaluation systems to incorporate student data. Many proposals included use of VAMs for 50 percent of a teachers' evaluation rating. VAMs have the potential to inform stakeholders about teacher performance, but the volatility in these models warrants caution moving forward because of potential misclassification of teachers.

Depending on the demographic factors used, value-added models produce varied results. Some VAMs control for demographic factors and use multiple years of scores on a handful of different assessments, while others

How Different Value-added Models Using One-year of Data Affect a Teacher with a High Proportion of At-Risk Students

Test Score Only Model (no student demographic factors)	
Math	**Reading**
Needs Improvement	Needs Improvement
Student Demographic Model (includes all available student demographic factors)	
Math	**Reading**
Highly Effective	Highly Effective

Source: UNM

Value-added models effectively identify very-high and very-low performing teachers.

do not. Experts continue to debate about the usefulness of these different models. In 2012, Pearson Education, Inc. published a study comparing five different VAM teacher evaluation approaches and concluded that the results are not definitive and depend on the model used.

Some value-added models adversely affect educators teaching certain populations of students. The Value-added Model Research Group at the University of New Mexico's College of Education used five years of teacher and student data to determine scores for teachers from two different VAMs on teacher scores. One VAM incorporated only test scores (test-score only model), while the other compensated for contextual variables, such as poverty and English language fluency (student demographic model).

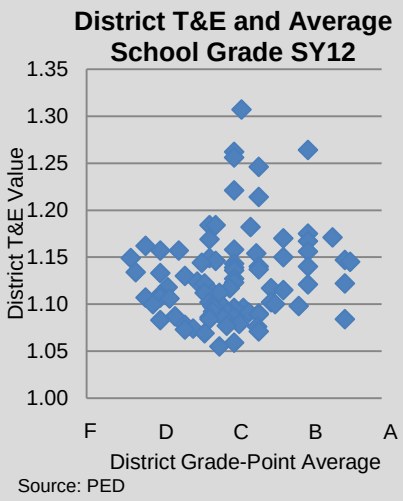
Value-added models are limited in what they can tell educators, the public, and other stakeholders. VAMs compare teachers with one another, making them a relative concept. One set of scores from VAMs do not indicate whether quality instruction is occurring in classrooms, only whether one teacher has students with higher achievement gains than another. Over time, VAM scores across multiple years can create a picture of absolute teacher performance that can be used for licensure advancement.

The use of value-added models can be responsibly integrated into the three-tiered system to identify teachers for advancement and bonus pay. VAM methodologies being developed for teacher evaluation and school grading could be leveraged to measure performance at tier levels and used as benchmarks in advancement between levels. Specifically, VAMs can be used to reward good teachers and identify poor teachers for professional development as part of a larger system of teacher evaluation.

Resource allocation amongst districts and schools creates funding disparities between low-income students and their more affluent peers, without driving student achievement. The funding formula rewards more affluent districts through the T&E index, a multiplier that allocates funds toward districts based on two variables that have not been shown to affect student achievement: teacher education levels and experience. Aligning this factor with a modified three-tiered system and offering a mix of incentives for high-performing teachers in low-income schools will better align resources with student achievement.

New Mexico directs nearly $200 million for higher teacher compensation through the T&E index in the public school funding formula. T&E index values are based on teachers' highest degree earned and years of experience. That index value is multiplied by student demographic and program units already generated in the formula. Funds generated by the T&E index in the funding formula compose up to 20 percent of a district's formula funding.

The T&E index directs more funding to more affluent school districts and produces a questionable return on investment after factoring in poverty. The T&E does not recognize better performance by teachers and higher pay, but instead rewards relatively affluent districts for keeping teachers and sometimes requiring them to meet higher education requirements. Aligning



Source: PED

Average Salaries at High and Low-Poverty Elementary Schools in Selected Districts SY12

Dist.	Elementary School	Avg. Salary	FRL
APS	Emerson	$42,900	96%
	Georgia O'Keefe	$46,923	15%
SFPS	R.M. Sweeney	$41,503	87%
	Wood-Gormley	$49,506	19%
LCPS	Booker T. Washington	$43,714	90%
	White Sands	$46,385	30%

Source: PED

the T&E index to a modified three-tiered system that focuses on student performance will allow the state to send resources to high-performing teachers and schools. High-poverty, rural districts with the greatest needs generally have difficulty hiring experienced teachers with advanced degrees who increase the T&E index. As FRL levels increase, T&E values drop.

The T&E index is not aligned to the three-tiered system. The three-tiered system provides large salary changes not accounted for in the T&E index. For several years the LFC has noted, in its present form, the T&E index is not aligned to the three-tiered system. An evaluation of the public school funding formula conducted jointly by the LFC and the Legislative Education Study Committee, the New Mexico Effective Teaching Task Force final report, and the AIR funding formula study recommended better alignment of the T&E index with the three-tiered system.

Level III teachers are more likely to teach in more affluent districts and schools. While districts do not have explicit policies to move more experienced and educated teachers to more affluent schools, students in poverty are more likely to have a less experienced, poorer performing teacher. Recruiting and retaining high quality teachers in low-income schools is integral to ensuring students in poverty achieve academically.

The state system can offer a mix of incentives to recruit and retain good teachers in high-poverty schools. Research shows financial incentives can recruit high-quality teachers and slightly decrease turnover in the short-term, but money does not work in the long-term to keep teachers at low-income schools: "Even when bonuses succeeded in drawing teachers to the poorest schools, such incentives could not compensate for the lack of support they encountered in these schools, which in turn contributed to the departure of many of these teachers."

KEY RECOMMENDATIONS

The Legislature should:

Replace the PDD and establish the effective teaching portfolio (ETP) as part of the licensure advancement application with new requirements and competencies. The ETP will have new requirements strengthening expectations for student achievement, requiring satisfactory annual evaluations, and allowing the most effective teachers, as measured by a statewide VAM, to bypass the ETP process;

Create licensure terms for level I, II, and III licenses. Level I licenses should have a five-year license term; and teachers must submit for renewal after three years; level II and Level III licenses should have an eight-year term and teachers must submit for renewal after six years;

Create new requirements for level II and level III licensure renewal, including meeting student performance expectations through the ETP or statewide VAM, and allowing teachers not meeting those expectations extra time to show competency before being denied renewal of a teaching license;

Require PED to annually rank the performance of licensed teachers providing instruction in tested grades and subjects through two different value-added models;

Change the T&E index to an effective teacher index that rewards districts based on the number of teachers they have in each license level;

Consider a mechanism, possibly through the funding formula, to provide additional compensation to effective teachers (as measured by the new aforementioned teacher evaluation and three-tiered licensure system) to teach in high-poverty schools.

PED should:

Establish updated basic competency and effectiveness indicators for teachers, as part of the ETP, including setting new student achievement expectations for Level II teachers and more rigorous student achievement expectations Level III teachers;

Establish an effectiveness evaluation for career teachers (level II and III), occurring every three years that incorporates student achievement and professional development goals. Public schools may use the results of the effectiveness evaluation to make employment decisions, in accordance with other provisions of law;

Require a professional development plan by the 40th day establishing the current year's performance goals, including measurable objectives for student performance. The goals should be based on updated basic competency and effectiveness indicators, the previous year's annual evaluation, and a previous year's students' performance;

Create and use a statewide VAM that uses two different calculations to obtain a composite score to help eliminate VAM biases for teachers of certain populations;

Provide that a performance evaluation be conducted annually for all teachers, and be based on whether a teacher exceeds, meets or, does not meet expectations on basic competency and effectiveness indicators, professional development goals, and satisfaction from parents. All teachers must be observed by principals 3 times a year.

Overview of the three-tiered system. In 2003, the Legislature passed comprehensive education reform, including the establishment of the three-tiered system and corresponding new minimum salaries. The School Personnel Act of the Public School Code outlines the three-tiered system certification and compensation schedules. The minimum salaries established in law were phased in between 2003 and 2008:

- Level I, Provisional Teacher: $30,000 in SY04;
- Level II, Professional Teacher: $35,000 in SY05 and $40,000 in SY06; and
- Level III-A, Master Teacher: $45,000 in SY07 and $50,000 in SY08.

The three-tiered system requires teachers to submit a professional development dossier (PDD) for level advancement. The PDD is a collection of evidence of teacher performance assessed by external reviewers, and is intended to provide sufficient evidence that a teacher is qualified to advance to a higher licensure level. The PDD evaluates teachers on nine teacher competencies.

Fast Facts of the three-tiered system and the PDD process

The PDD process has three submittal periods per year: February, June, and November, and takes about three to four months to complete

Each PDD submittal costs a teacher approximately $100.

2012 PDD pass rate:

- Teachers moving from level I to level II: 92 percent
- Teachers moving from level II to level III: 90 percent

Not all teachers at higher licensure levels submitted a PDD for advancement, as over 2,700 teachers advanced to Level III 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. This gap in timing provided a window for certain teachers to qualify for $50 thousand salaries without submitting a PDD. Many of these "grandfathered" teachers are still actively teaching in New Mexico public schools and hold a level II or III license without submitting a PDD.

The three-tiered system has substantially increased teacher pay in New Mexico. According to a report issued by the National Education Association (NEA), salaries for New Mexico public school teachers increased 38.8 percent from the SY01 to SY11. This is the eighth highest increase among states in the nation during the ten year time span. In SY11, New Mexico ranked 40th among states, with an average public school teacher salary of $46,888, according to the NEA report. The report did not factor in cost of living into its rankings.

Evaluations as part of the three-tiered system. Local and state evaluations are both components of the three-tiered system. School administrators conduct local evaluations every one to three years and focus on evaluating classroom practice. State evaluations are conducted through the PDD, and teachers submit local evaluations as a part of the PDD when applying to advance a license level. While the local and state evaluations use the same competencies to evaluate teachers, only local evaluations are conducted at school sites.

Local evaluations are also considered for teachers renewing their licensure. Based on local annual evaluations, the superintendent of the district (or governing authority of other institution or school) verify the applicant has demonstrated the competencies for the current level of licensure and has met other requirements of the state's highly objective uniform standard of evaluation (HOUSE).

Three-tiered system and improved student performance. One of the primary purposes of the three-tiered system was to facilitate student 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 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."

Given 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. The standards-based assessment (SBA) is a statewide assessment given annually to third through eighth-grade students and again to eleventh-graders. The SBA meets the requirements of No Child Left Behind and is based on New Mexico state standards. New Mexico has four levels of performance used by the SBA: beginning step, nearing proficiency, proficient, and advanced. Proficient performance is expected of New Mexico students.

Other reports have explored the connection between the three-tiered system and student performance, including a Legislative Education Study Committee (LESC) memo in 2006 describing the extent to which the three-tiered system requires documentation of student achievement. In 2007, a joint evaluation by the Office of Educational Accountability (OEA), the LESC, and the LFC suggested further study into the links between advanced licensure and student academic performance. A 2009 joint report by the Legislative Finance Committee, Legislative Education Study Committee, and the Office of Education Accountability studied links between licensure and student performance. This report serves as a follow-up to the 2009 report.

Data used for LFC three-tiered analysis

The LFC used teacher records linked to student math and reading SBA scores to analyze the relationship between licensure level and student achievement. The total numbers of 3-8 grade teachers with at least one math or reading score ranged between 6,900 and 7,200 teachers. Teachers with fewer than 10 students were eliminated from the dataset along with level 0 teachers and students who did attend the same school for the full academic year (FAY). Non-FAY students were included in subsequent VAM analysis. Teachers with only third grade students were eliminated from the dataset as the students for these teachers did not have prior-year data to calculate gain. Data below reflect the populations after these data cleaning techniques were applied.

Selected Descriptive Statistics for New Mexico Teachers Used in LFC three-tiered Analysis

Year	Number of Teachers	Average Salary	Average Years of Experience	Percent Hispanic	Age in years at start of year
SY10	4,629	$45,612	9.7 years	31.3%	44.3
SY11	4,608	$45,531	10.1 years	36 %	44.1
SY12	4,595	$44,788	10.1 years	35.8%	44.3

Source: LFC Analysis of PED Data

Selected Descriptive Statistics for New Mexico Teachers Used in LFC three-tiered Analysis

Year	Number of Students	Percent FRL	Percent ELL	Percent Hispanic	Percent SPED	Percent Gifted
SY10	98,378	68.5%	13.7%	57.7%	11.6%	7.9%
SY11	101,029	69.5%	13.7%	61.1%	11.3%	8%
SY12	102,152	69.8%	14.6%	61.5%	11.7%	8.7%

Source: LFC Analysis of PED Data

The three-tiered system and the training and experience (T&E) multiplier has a major impact on a district's allocation from the State Equalization Guarantee (SEG). This multiplier increases funding allocated to a district by as much as 20 percent based 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 and receive less funding from the T&E index. For several years the LFC has noted that the T&E index is not aligned to the three-tiered system.

Evaluation and value added modeling. States are increasingly relying on a statistical procedure known as value added modeling (VAM) to evaluate teacher performance. VAMs have the potential to inform stakeholders when student achievement data exists (reading and math) but not other subject areas. There is not one widely accepted VAM among education researchers or administrators. Value added models use data from students' past test scores to predict subsequent scores and then subtracts that prediction from current year scores to provide an estimate for teachers. This estimate is the "value added" and the models themselves can range from simple statistical procedures to more complex, multi-level models. Models can be run in basic statistical software, but more complex models require custom programming of statistical formulas and are increasingly being run by private, for-profit corporations.

Development and reliance on VAMs for teacher evaluation is controversial. Teachers in Chicago Public Schools went on strike in part because of reliance of the city on VAMs accounting for 45 percent of teacher evaluations. After seven lost school days, a renegotiated contract reflected a lesser reliance on VAMs, accounting for 30 percent of teacher evaluations.

Formed by Executive Oder in April 2011, the Effective Teacher Task Force's purpose was to determine how best to measure the quality of teachers and school leaders. The group publicly met 10 times and issued recommendations in August 2011. The task force recommended replacing the current pass/fail teacher evaluation system with five effectiveness levels determined, in part, by the results of VAMs. During the 2012 Legislative Session, House Bill 249, instituting a similar teacher evaluation system failed.

In April 2012, PED formulates and implemented a new teacher evaluation system making academic growth a key factor in teacher and principal evaluation. The new evaluation system was required as part of PED's waiver from No Child Left Behind. The NMTEACH workgroup, made up of teachers, administrators, union representatives, and other stakeholders, has since helped finalize a rule creating a new evaluation system. Fifty percent of the evaluation is based on a teacher's student achievement growth (35 percent on the SBA and 15 percent on other assessments); 25 percent on locally-adopted, PED-approved measures such as student surveys and short-cycle assessments; and 25 percent on observations of teaching. Fourteen school districts and 68 schools are piloting this system during the 2013 school year with statewide implementation scheduled to begin in SY14.

NEW MEXICO'S THREE-TIERED CAREER LADDER SYSTEM DOES NOT ALIGN PAY WITH STUDENT ACHIEVEMENT

Improving student achievement was a key policy goal of implementing the three-tiered system. The three-tiered system's founding legislation states student success for every child is the fundamental goal of New Mexico's education system. The three-tiered system was designed to help achieve this goal by attracting, retaining, and holding accountable quality and diverse teachers.

The three-tiered system is meant to be a progressive career system for teachers in which license level is a reflection of teacher ability, performance, and leadership. A level I license is a provisional license that allows beginning teachers to develop, whereas level II and level III represent teachers who meet and exceed department-adopted academic content and performance standards. Minimum salaries of $30 thousand, $40 thousand, and $50 thousand are tied to each licensure level to compensate teachers for their performance and abilities, as measured by the nine teacher competencies, as well as leadership roles taken at the school level.

Table 1. The Three-Tiered System's Licensure Levels

Licensure Level	Description	Minimum Salary
Level I	A provisional license that gives a beginning teacher the opportunity for additional preparation to be a quality teacher.	$30,000
Level II	A license for a fully qualified professional who is primarily responsible for ensuring that students meet and exceed department-adopted academic content and performance standards.	$40,000
Level III	A license for the highest level; for teachers that advance as instructional leaders in the teaching profession and undertake greater responsibilities such as curriculum development, peer intervention and mentoring.	$50,000

Source: NMSA 1978 22-10A-4

State law requires teacher evaluations to use a highly objective uniform statewide standard of evaluation (HOUSE). The Public Education Department developed nine key teaching competencies covering three areas of practice: instruction, student learning, and professional learning. Through the three-tiered system, teachers are evaluated against these competencies using local and state evaluation processes. A complete list of all competencies can be found in **Appendix B**.

State and local evaluations are essential pieces of the three-tiered system. Ascending levels within the three-tiered system is contingent upon evidence of satisfactory annual evaluations at the local level, as well as a satisfactory score on the Professional Development Dossier (PDD). While the local and state evaluations use the same competencies to evaluate teachers, only the local is conducted at the school site; state evaluations are conducted through the PDD, a collection of evidence of teacher performance that is reviewed externally.

Table 2. The Three-Tiered System's Evaluation System

Local Evaluation	State Evaluation (PDD)
Occurs regularly - every 1-3 years	Occurs when teacher applies for new license level
Teachers are evaluated on 3 strands of teacher competencies	Teachers are evaluated on 3 strands of teacher competencies
Evaluations are conducted by school administrators	Evaluations are conducted by independent reviewers
Focused on evaluating classroom practice	Focused on evaluating whether artifacts meet competencies
	Local Evaluations are taken into account

Source: NMSA and NMAC

Since 2009, nearly 6,000 teachers advanced to new licensure levels in the three-tiered system, receiving $59 million in mandatory salary increases. During that time 3,877 thousand teachers advanced from level I to level II, and 1,980 thousand advanced from level II to level III. Each advancement results in a minimum salary change of $10 thousand, or between a 25 percent to 33 percent increase in base pay.

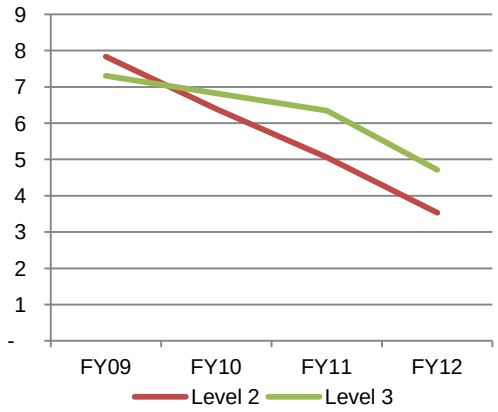
Table 3. SY09-SY12 Number of Teachers Ascending Licensure Levels

	SY09	SY10	SY11	SY12	Total
From level I to level II	904	1,278	786	909	3,877
From level II to level III	497	637	384	462	1,980
Total	1,401	1,915	1,170	1,371	5,587

Source: LFC Analysis of PED Data

Teachers advancing through the three-tiered system increasingly make up a larger proportion of classroom teachers than those grandfathered into their licensure level. As grandfathered teachers retire, the impact of the three-tiered system and PDD will become more apparent. The number of teachers grandfathered into the three-tiered system has declined steadily due to retirements and level II grandfathered teachers going through the PDD process to obtain a level III license.

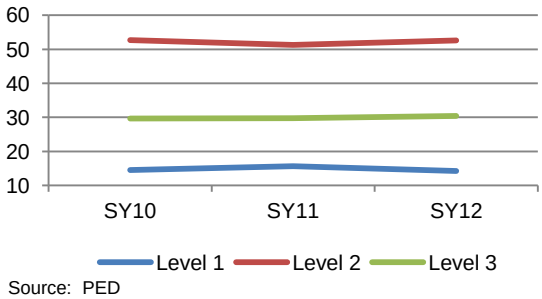
Chart 1. Grandfathered Teachers by Licensure Level
(in thousands)



Source: LFC Analysis of PED Data

Level III teachers comprise a larger and growing share of all teachers in the system. Between 2009 and 2012, the percent of level I teachers dropped from 15 percent to 14 percent of all teachers, while the percent of level III teachers increased nearly one percent, from 29.7 percent to 30.4 percent of all teachers.

Chart 2. Percent of Teachers by Licensure Level



Source: PED

In 2009, LFC staff recommended strengthening the three-tiered system to ensure teachers receiving large salary increases were producing better results for students. The LFC's 2009 evaluation of the three-tiered system and achievement gap found teacher licensure level did not significantly raise student achievement. Level III teachers generally had higher student achievement than teachers of other licensure levels, but they also were more likely to have a population of students more likely to succeed. The report recommended PED consider developing a bonus pay-for-performance pilot program. The report also called on PED to form a workgroup to evaluate proposals such as requiring more evidence of student performance in PDD submissions and teacher evaluations and establishing goals for expected gain on the SBA in grade levels and content areas.

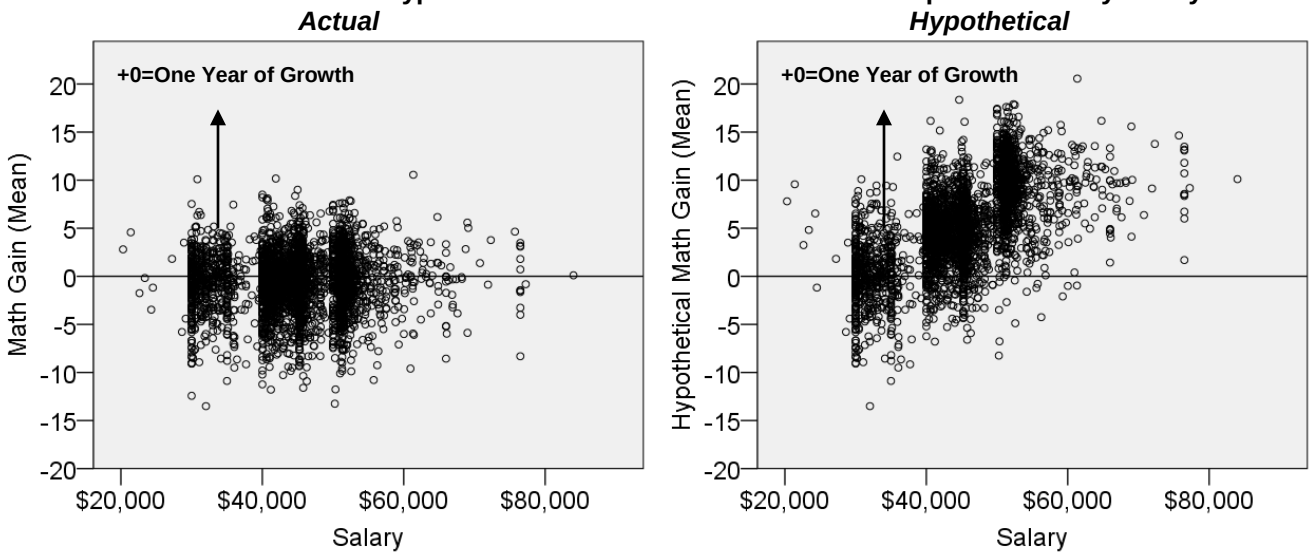
The state has not established expectations for student achievement across level I, II, and III teachers as part of evaluation systems. Competencies used in the state and local evaluations of the three-tiered system include examples of student performance, but the evaluations have no expectations for the performance of all students, particularly on standardized tests. When the three-tiered system was established, the SBA was new and lacked longitudinal information. As a result, student data was not incorporated into the evaluations. Teachers with different license levels achieve similar student performance, and a majority of New Mexico teachers do not feel the state evaluation process identifies effective teachers.

Student achievement is not a robust element of the current three-tiered system. To advance tiers through the PDD, teachers must submit and analyze student work. Teacher completing the PDD select examples of low, medium, and high-level student work and submit written reflections on that work. However, the 2007 joint report on the three-tiered system by the LFC, LESC and OEA stated, "These requirements focus primarily on describing or documenting student achievement, while involving no direct, explicit consequences – whether rewards or sanctions – for teacher based on the achievement of their students." Furthermore, the requirements in PDD only provide a picture of student performance for a few students, which are not easily compared across classrooms, schools, and districts. Student achievement data from statewide assessments like the SBA are comparable across the state.

As a result of no expectations for student performance, teachers across licensure levels produce similar student achievement results, despite large differences in pay and cost to taxpayers. Teachers in higher tiers generally produce better outcomes for students, but these differences are small and can often be accounted for by other factors. For example, after accounting for rates of English language learners (ELL), special education students (SPED), students participating in free and reduced-priced lunch (FRL), and ethnicity, the differences between tiers are further diminished.

The following scatter plots show how New Mexico teachers perform based on the average test score gain of their students in math as measured by the SBA and a hypothetical situation if more highly paid teacher were providing significant gains for students. Results for reading SBA scores are similar. A gain of zero represents a full year of academic growth for a student. For example, if a student scored a 40 on the 3rd grade reading SBA in 2011, and the following year scored a 40 on the 4th grade reading SBA, that student has maintained their proficiency level and met the challenge the 4th grade test provided. In chart 4, the scatter plot entitled *Actual* shows clear clusters of leveled teachers based on their minimum salaries. Each level achieves similar student achievement. If higher licensure levels were associated with higher student achievement, the scatter plot would look like the scatter plot entitled *hypothetical* in chart 4.

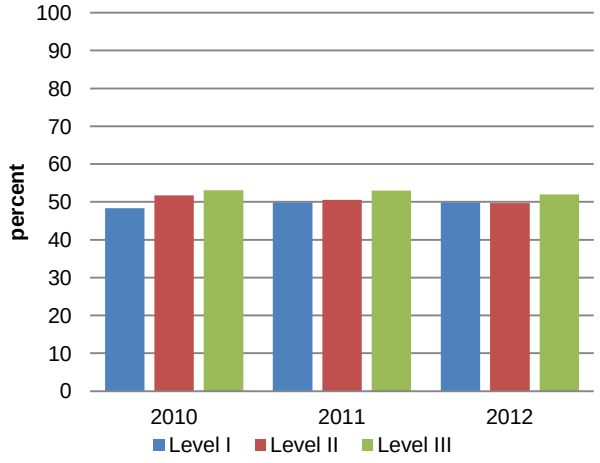
Chart 3. Actual and Hypothetical Math Student Academic Gain per Teacher by Salary



Source: LFC analysis of PED data

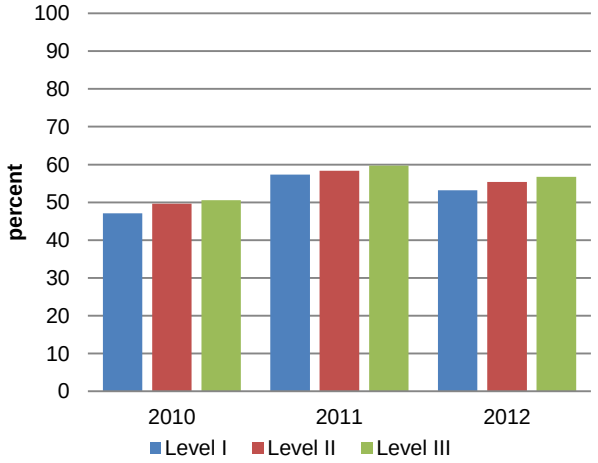
Based on student gains on the math and reading SBA, small differences exist between licensure levels. For example, in 2012, 52 percent of level III math teachers' students obtained a year's worth of growth, while 50 percent of level I and level II teachers' students accomplished the same feat. Even in years when the teachers with different licensure levels produced more disparate results, such as 2010, the difference between the percent of students who gained in math and reading was 4 to 5 percentage points, and significant differences did not exist among all tier levels. In 2010, 48 percent level I teachers' math students gained on the SBA and 53 percent of level III students gained on the assessment.

Chart 4. Percent of Math Students Obtaining One Year of Growth Per Tier (as measured by SBA gain scores)



Source: LFC Analysis of PED data

Chart 5. Percent of Reading Students Obtaining One Year of Growth Per Tier (as measured by SBA gain scores)



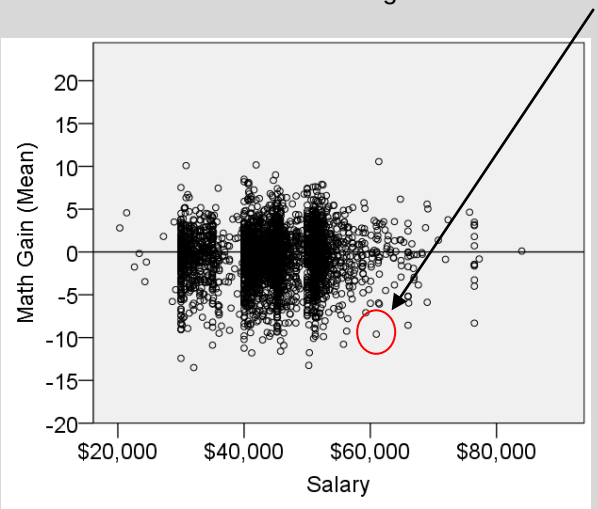
Source: LFC Analysis of PED data

Student demographics can explain part of the difference between licensure level performances. Level I teachers are more likely to teach students in poverty. Nationwide research and studies by the LFC found the obstacles facing students in poverty can affect achievement. For example, according to the LFC's program evaluation on developing early literacy in New Mexico, "Gaps persist in achievement between ethnicities, but the biggest differences are strongly associated with socioeconomic status and English language acquisition levels." Analysis from that study showed a 50 percentage point difference in reading proficiency levels between ELL and FRL students and their non-ELL, non-FRL peers. Teachers with a higher proportion of these students, such as level I teachers in New Mexico, may have lower test scores and fewer students showing gains on the SBA as a result.

Teachers in each licensure level perform similarly in moving their students up a proficiency level. Helping students increase a proficiency level on the SBA, such as an increase from a nearing proficiency rating one year to a proficient rating the following year, is uncommon: fewer than 20 percent of students moved up at least one proficiency level for math, and fewer than 25 percent moved up at least one proficiency level for reading during the last three years. Like the percentage of teachers' students who gain on the SBA, the difference between each licensure level of teachers is small, varied, and is not always statistically significant between tiers.

An example of a consistent poor return on investment

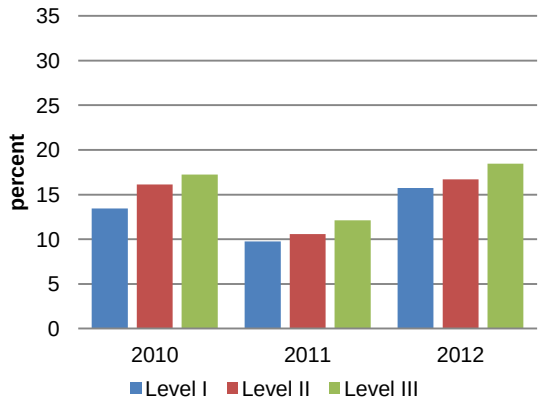
In 2012 an elementary teacher with over 30 years of experience was paid $60 thousand dollars. Their class of 15 students averaged a loss of 9.6 math SBA points compared to the previous year with over 70 percent of the class going down at least one proficiency level. The previous year the same teacher's class lost over 10 SBA points compared to the previous year. Losses for this teacher were also seen in reading SBA scores.



Source: LFC Analysis of PED data

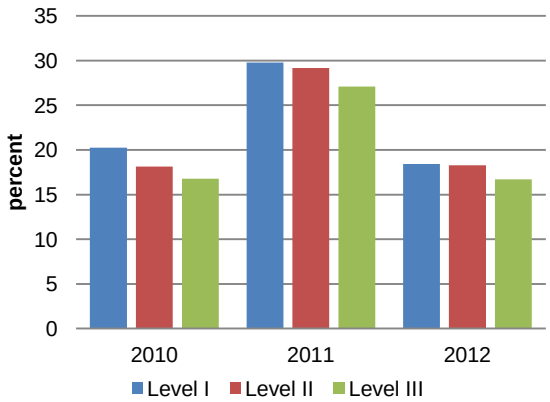
In math, students are losing more ground than they are gaining. Percentages for students moving down at least one proficiency level in math are higher than the percent of students moving up at least one proficiency level. Up to 30 percent of students lose at least one proficiency level in math, while less than 20 percent moved up at least one proficiency level in the last three years. Students are likely losing a proficiency level due to ineffective teaching and a lack of interventions.

Chart 6. Percent of Students Moving Up One Proficiency Level (Math)



Source: LFC Analysis of PED data

Chart 7. Percent of Students Moving Down One Proficiency Level (Math)

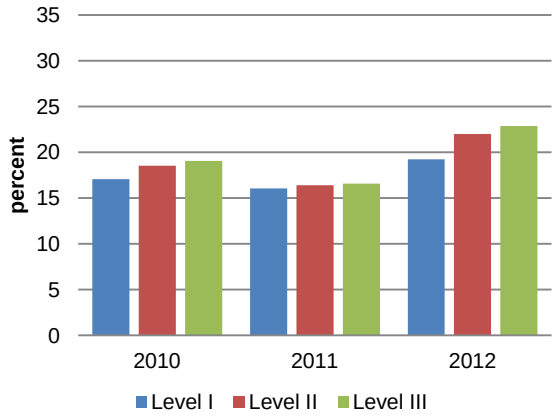


Source: LFC Analysis of PED data

In 2012, level III teachers were able to drive 2 percent more of their math students and 4 percent more of their reading students to higher proficiency levels. In 2011, level III teachers were able to move 1 percent more of their reading students up a proficiency level and 2 percent more of their math students up a proficiency level.

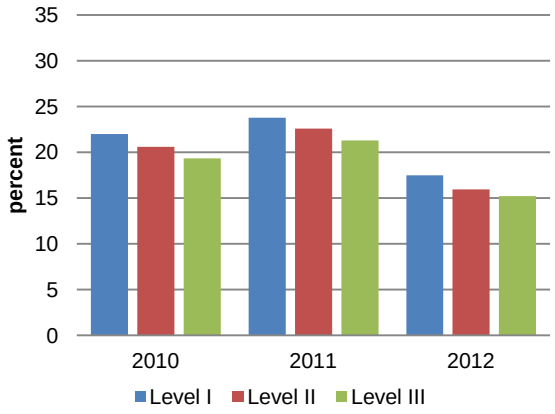
A smaller share of level III teachers' students move down a proficiency level, but the difference between them and level I teachers' students is small. In 2012, 17 percent of math teachers' students went down a proficiency level, while 18 percent of level I teachers' students moved down a level. Level II teachers had the same percentage of students decrease a proficiency level as level I teachers. Amongst all licensure levels and all years, more students move down a proficiency level than up. Almost 30 percent of a teacher's students moved down at least one proficiency level for math, and almost 25 percent of a teacher's students moved down at least one proficiency level for reading, during the last three years.

Chart 8. Percent of Students Moving Up One Proficiency Level (Reading)



Source: LFC Analysis of PED data

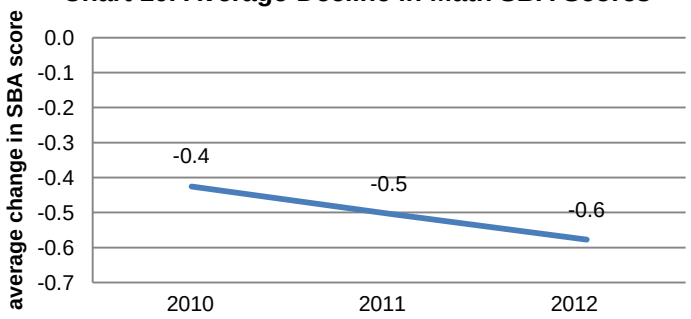
Chart 9. Percent of Students Moving Down One Proficiency Level (Reading)



Source: LFC Analysis of PED data

Math teachers' students across all teachers decline year-over-year. Math students in the state averaged a decline of 0.4 points on the math SBA in 2010 and a decline of 0.6 points on the assessment in 2012.

Chart 10. Average Decline in Math SBA Scores



Source: LFC Analysis of PED Data

Without more differences in student outcomes, the three-tiered system acts as an expensive proxy for paying teachers based on their education and experience. Licensure level, education level, and years of experience are strongly related amongst New Mexico teachers. This relationship is driven by PED's requirements for submitting a PDD, which include minimum years of experience and education level. A bachelor's degree is the highest degree obtained by 69 percent of level I, while 54 percent of level II teachers have a bachelor's as their highest degree completed. Of level III teachers, over half have earned a master's degree or higher. Similarly, level I teachers have fewer years of experience, 1.6 years, compared with level III teachers, 12.3 years.

The local evaluation system does not differentiate between high and low-performing teachers or focus on student achievement. Evaluation requirements need strengthening to align with common core, use student data, assess teacher effectiveness, and improve satisfaction among teachers. The competencies and requirements used in the state's PDD and local evaluation system have not been reviewed and changed since implementation and could be updated to better reflect current education research. Much research has been published regarding effective teaching since 2003, and a review and update of these competencies is needed to ensure the evaluation's effectiveness.

The PED, through their Effective Teaching Task Force and NMTEACH Task Force, has developed a new rule that changes the local evaluation structure for teachers. The new evaluations are broken down into three major components: 50 percent of the evaluation is based on a teacher's student achievement growth; 25 percent on locally-adopted, PED-approved measures; and 25 percent on observations of teaching. The new evaluations will be fully incorporated for all teachers during the 2014 school year.

Effective and explicit use of student data is only a small part of the current local evaluation structure. The performance evaluation system does not include a clear standard of practice for data use, particularly assessment data. Strand 2, competency 5 states "Uses information gained from ongoing assessment for remediation and instructional planning," but does not suggest how a teacher goes about using data effectively, and more precise expectations are needed to drive teacher development.

Not all teachers are annually observed and evaluated on their classroom practice. While level I teachers must have their classroom practice evaluated each year, level II and level III teachers only receive such evaluations every three years. Between these evaluations of their classroom practice, level II and III teachers use progressive documentation to provide formative information on performance. This progressive documentation, which does not require classroom observation, results in a summative performance evaluation every three years, which closely resembles a level I annual evaluation and includes classroom observations.

The current pass-fail evaluation system does not provide nuanced feedback or identify of teachers' ability. New Mexico uses a binary evaluation system that rates teachers as satisfactory or unsatisfactory. A recent study by the New Teacher Project, entitled *The Widget Effect*, analyzed 12 districts in four states to find 99 percent of teachers in districts using binary evaluation systems are rated effective. The study concludes this is problematic because excellence goes unrecognized and professional development opportunities and support cannot be properly targeted toward teachers who need it.

The New Mexico Effective Teaching Task Force made a similar finding in its final report; "Research indicates that multiple levels of effectiveness are needed in order to provide a mechanism for distinguishing average work performance from truly outstanding work performance."

Other measures not included in the current teacher evaluation system or PED's new teacher evaluation system are needed to assess complete teacher performance. While student achievement is integral to determining the effectiveness of a teacher, other measures must be incorporated into teacher evaluation systems to fully capture a teacher's performance. The Bill and Melinda Gates Foundation's Measures of Effective Teaching (MET) study suggests a fair and reliable teacher evaluation system includes the following five measures:

- Student achievement gains on assessments,
- Classroom observations and teacher reflections,
- Teachers' pedagogical content knowledge,
- Student perceptions of the classroom instructional environment, and
- Teachers' perceptions of working conditions and instructional support at their schools.

The New Mexico Effective Teaching Task Force came up with similar recommendations for multiple measures to be included in a new teacher evaluation system, noting "Effectiveness levels should only be assigned after careful consideration of multiple measures, including student achievement data, observations, and other proven measures selected by local districts from a list of options approved by New Mexico's Public Education Department." Because local districts select many of their multiple measures from a menu of options, some of measures might not be included in the evaluation.

Competencies need to be updated to align with the common core standards. PED calls the standards, which will be phased-in over three years and in full effect by 2015, "A different approach to learning, teaching and testing engenders a deeper understanding of critical concepts and the practical application of knowledge." Given the substantial change to the standards, testing, and expectations for teaching, the current competencies will need to be adjusted accordingly.

PED has recognized the shortcomings of the local evaluation system; however the proposed replacement requires a statutory change and is not linked to the three-tiered system. Through two task forces, the PED has designed and begun implementation plans for a new local evaluation system. The system includes the use of multiple measures and allows districts to determine which assessments they will use to evaluate teachers in addition to the SBA. This conflicts with HOUSE, which articulates the way teachers are evaluated must be uniform across the state. If each district to choose part of its own evaluation system, teacher evaluation will not be uniform across the state. A change in legislation is necessary to implement their rule.

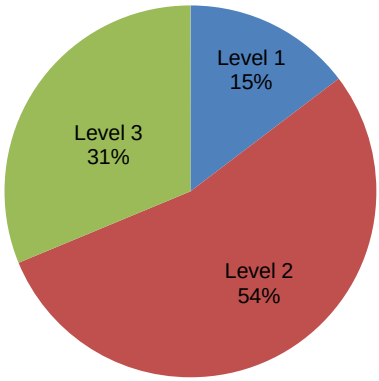
The professional development dossier does not effectively screen teachers for advancement, resulting in ineffective teachers receiving large pay increases. Teacher performance within licensure levels varies greatly because the state evaluation's PDD process does not effectively screen for high performing teachers. As a result, high and low performing teachers exist at each licensure level. The lack of clear and consistent performance among teachers in each licensure level shows the PDD process rewards teacher experience and education and not a teacher's impact on student achievement.

The PDD fails to effectively differentiate performance among teachers advancing licensure and pay levels; PED has not fixed this important part of the system. A 2009 study conducted by the LFC noted little difference between grandfathered teachers and teachers in licensure levels who passed the PDD. The report recommended using student achievement to drive the PDD process, but PED has yet to make student achievement a major factor in ascending licensure levels. The findings from the 2009 LFC report remain relevant to the current situation in the three-tiered system.

Over 90 percent of teachers who submit a PDD advance a level and receive large pay increases. Passage rates for teachers seeking level II and level III licenses indicate the system is providing raises for a vast majority of teachers who meet the experience and education criteria, rather than reaching an expected level of student performance. In 2012, 90 percent of the level II teachers who applied to move up to level III licensure were successful.

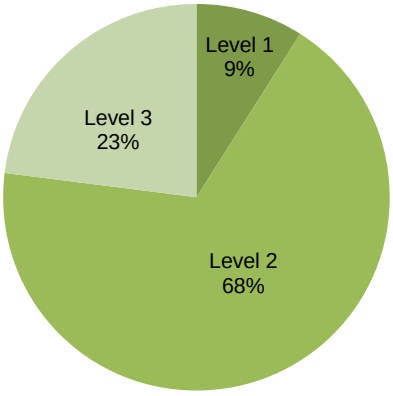
The differences in teacher performance within licensure levels vary greatly, suggesting the PDD process and competencies do not properly screen for teacher effectiveness. The PDD process does not do enough to focus and distinguish teachers based on student achievement, and as a result large amounts of high and low-performing teachers exist in each licensure level. In SY12, over two-thirds of high performing reading teachers had a level II license. Level II teachers made up the majority of underperforming reading teachers, but 30 percent of underperforming reading teachers were level III teachers. Underperforming teachers were identified as those performance is ranked in the lowest 16 percent of all teachers based on student performance on the SBA.

Chart 11. Proportion of Licensure Level Teachers as Percent of All Teachers



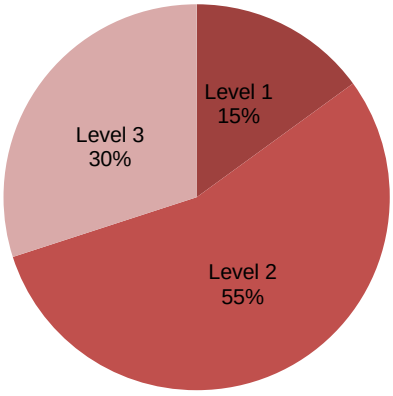
Source: PED

Chart 12. FY12 License Levels as a Proportion of High Performing* Reading Teachers



Source: LFC Analysis of PED Data
*High-performing teachers are those ranked in the top 16th percentile of all teachers

Chart 13. FY12 License Levels as a Proportion of Low Performing* Reading Teachers



Source: LFC Analysis of PED Data
*Low-performing teachers are those ranked in the bottom 16th percentile of all teachers

Only level II teachers were over-represented among high and low-performing teachers when compared with the proportion of level II teachers statewide. Fifty-four percent of teachers have a level II license statewide, but 68 percent of high-performing teachers and 55 percent of low-performing teachers have a level 2 license. Level I and Level III teachers are under-represented given their statewide percentages of 16 and 40 percent, respectively.

In SY12, the state awarded nearly $3 million in mandatory minimum salaries to low performing level-III teachers. Two-hundred eighty level III teachers ranked in the lower third of all valid math and reading teachers in moving students to a year or more of growth. In reading, this meant less than 40 percent of students achieved a year of growth; in math, ineffective level III teachers moved less than 30 percent of students a year's worth of growth. Seventy of the nearly 300 teachers were ranked in the lower third for both reading and math. Further, nearly 30 percent of the lowest performing teachers in the state had a level III license in SY12. Twenty-nine percent, or 271, of the bottom quartile of reading teachers have a level III license.

Table 4. Share of Teachers in Each Quartile of Teacher Performance, Based on SY12 SBA Reading Results

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	Percent	Count	Percent	Count	Percent	Count	Percent	Count
Level 1	15%	142	13%	123	13%	119	11%	106
Level 2	55%	506	55%	534	54%	501	54%	507
Level 3	29%	271	33%	317	33%	306	35%	327

Source: LFC Analysis of PED Data

Two hundred sixty level III teachers are paid the maximum mandated salary but perform amongst the lowest quartile of math teachers. Twenty-six percent of level II teachers and 25 percent of level I teachers are in the lowest quartile of performers.

Table 5. Share of Teachers in Each Quartile for Teacher Performance, Based on SY12 SBA Math Results

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	Percent	Count	Percent	Count	Percent	Count	Percent	Count
Level 1	15%	131	14%	124	16%	142	14%	125
Level 2	56%	495	56%	498	52%	462	52%	464
Level 3	29%	260	30%	266	32%	283	34%	298

Source: LFC Analysis of PED Data

Thirty-six percent of teachers disagree or strongly disagree the PDD successfully identifies highly effective teachers. In 2009, over half of teachers disagreed with the same statement. Since 2009, fewer teachers believe the PDD process positively impacts their ability to improve student performance or the materials required to submit a dossier adequately measure and reflect their skills as a teacher. The percentage of respondents who 'agree' or 'strongly agree' decreased by 9 percent and 4 percent, respectively. Many teachers felt the PDD process needed to include observations and should be include teacher observation.

> *"PDD is an OK tool, but not completely effective. Consistently random observation of classes is a quick way to see if teachers are at least managing and doing what is required. SBA is an effective tool for math, science and English, but not to grade the whole school. Each subject needs a diagnostic for each subject and grade area that reflects the student's knowledge of that grade-level subject area."*
>
> ***-Respondent from LFC Survey***

Despite a 92 percent rate of passage, only 25 percent of teachers believe the PDD is scored objectively and consistently. The respondents expressed concern with the honesty of the teachers submitting the dossier. After explaining they were encouraged to cheat on the dossier, one survey respondent noted, "The process relies on the integrity of the individual. Some are more honest than others. The artifacts used for evidence are good indicators of teacher effectiveness if they are authentic." Other respondents noted some teachers used "fake data" to successfully pass the PDD.

PED rule allows out of state teachers to bypass the PDD process for advancement to higher licensure levels. Whereas teachers within the state are required to go through the PDD process for advancement to a higher licensure level, a teacher who moves to New Mexico after teaching in another state or country can be placed into level II or level III licensure without passing the PDD. The presentation of a dossier is not required for these teachers and the basis of placement is total amount of years required for level placement.

The state allows low-performing teachers to keep their license level because the state does not have a rigorous license renewal process. Grandfathered teachers, those who obtained a level II or III licensure without going through the state's PDD, continue to renew their licenses without passing the PDD. Many of these teachers are low-performing teachers with the highest mandated salary in the state.

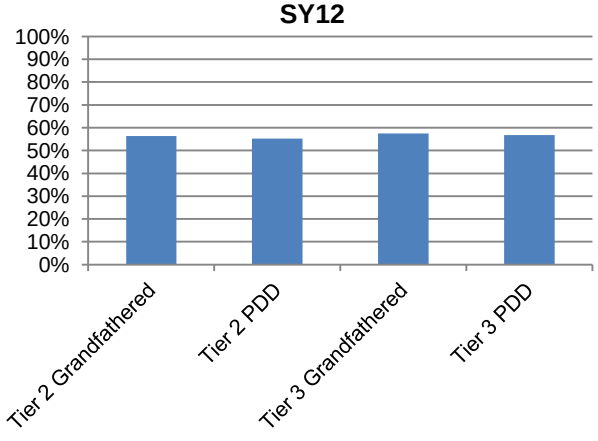
The requirement to renew licenses does not include demonstrated effectiveness. The requirement to renew licensure for New Mexico teachers is outlined in rule. Requirements include having demonstrated the competencies for the current level of licensure and meeting other requirements evidenced by annual evaluations. Therefore, a teacher can stay at level II for their career without completing the PDD. Although evaluations are taken into consideration for renewal, current evaluations do not use student performance measures based on SBA scores or other standardized student outcomes and therefore could allow underperformance over time.

A renewed level III license for a low-performing teacher

PED's inability to identify ineffective teachers and the lack of measures of effectiveness in license renewal has a detrimental effect on student achievement. For example, a level III teacher in a medium-sized school district in eastern New Mexico had over 20 percent of their class lose a proficiency level in math for SY10 and the class averaged a two-point loss on the SBA. In SY11 this teacher's performance worsened significantly as over 60 percent of their class lost a proficiency level in math and on average lost eight points on the SBA. This teacher's level III license was renewed in 2012 despite this poor performance. Unfortunately, in SY12 this teacher's performance worsened again with over 70 percent of their class losing at least one proficiency level and the class, on average, losing 11 points on their SBA score compared with the previous year.

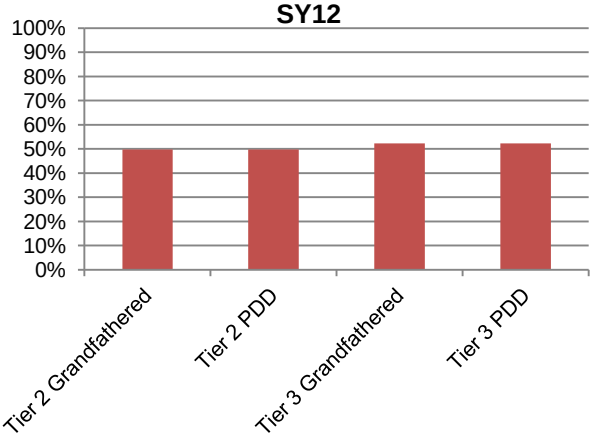
The state grandfathered thousands of teachers into higher salaries without a dossier, and does not require demonstrated effectiveness upon renewal of license. Students of PDD passers and those grandfathered into the system perform similarly, suggesting the three-tiered system has been unsuccessful in driving student achievement. The PDD is intended to provide sufficient evidence a teacher is qualified to advance to a higher licensure level and costs teachers hundreds of dollars to apply. However, in SY11 and SY12 there were no significant differences in student achievement between PDD and grandfathered teachers.

Chart 14. Students Obtaining One Year of Reading Growth Per Level In SY12



Source: LFC Analysis of PED Data

Chart 15. Students Obtaining One Year of Math Growth Per Level In SY12

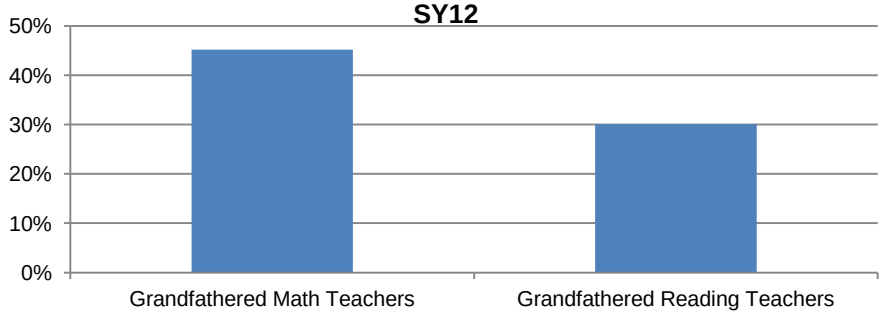


Source: LFC Analysis of PED Data

Grandfathered teachers show no significant differences in students moving up or down in proficiency or in raw gain score when compared to teachers who passed the PDD. Similarly, comparing the percentage of students with a year of growth on the reading and math SBA showed no significant difference between PDD passers and grandfathered teachers. On average level II and level III PDD teachers have 11 years of experience whereas grandfathered teachers at the same levels have 14 years of experience. After accounting for years of experience PDD and grandfathered non-PDD teachers perform similarly.

Nearly 50 percent of grandfathered teachers are in the lowest half of teachers when measured by student achievement. Forty-five percent of grandfathered reading teachers and 50 percent of grandfathered math teachers fall into the lowest half of teachers when measured by student performance. Many students in the classrooms of grandfathered teachers do not grow one year on the SBA in math or reading. Less than half of 126 grandfathered math teachers' students obtained a year's worth of growth, and 182 grandfathered reading teachers achieved the same feat.

Chart 16. Grandfathered Teachers with Less than Half of Students Obtaining a Year's Worth of Growth by Subject SY12



Source: LFC Analysis of PED Data

The three-tiered system offers a framework to align resource allocation to performance, but student achievement must be better incorporated into the process. The three-tiered system offers a framework for the state to align resources with results. The system offers significant salary increases and a competency structure; if modified, this system could drive student achievement across the state. A modified three-tiered system could serve as a way to strategically reward the state's best teachers and provide strategic interventions for struggling teachers.

The three-tiered system successfully retained teachers in New Mexico schools. A 2007 LFC, LESC, and OEA study found nearly 64 percent of teachers believed the three-tiered system helped with recruiting and retaining teachers. The study, which compared data from 2001 and 2007, found fewer teachers were leaving the profession within their first three years and fewer teachers overall were leaving to take positions in other states or outside of the teaching profession.

PED now captures student achievement data that could be incorporated into the PDD to make it more robust. PED now has many years worth of student achievement data linked to teachers to articulate their effectiveness. For example, SBA data can now connect students, demographic information, and teachers. Student achievement should be central to any changes to the three-tiered and teacher evaluation systems, but the state should move forward knowing that restrictions to the way data is currently collected and flaws in statistical models measuring student achievement will need to be addressed before the system is implemented.

Opportunity exists to incorporate demonstrated effectiveness into passage between tiers and renewal and better align funding and results. Incorporating demonstrated effectiveness in the form of student achievement could help provide incentives for better performance and deliver a higher return on investment for teachers as higher levels in the system. Both systems currently make student achievement a part of their systems but fail to make clear expectations about the amount of student achievement expected and the metrics to be used to measure student

achievement. Teacher evaluation systems should evaluate teachers on clear criteria such as student testing gain scores, whereas the three-tiered system could incorporate value-added models to identify the highest performing teachers in the state as one criterion for advancement. For example, forty-six percent of teachers with level II licenses who have been teaching for the last three years have averaged at least a year of growth for students in two of the last three years. A system that prioritizes these teachers for movement up to level III or renewal of their level II license would begin to align resources with results. This is just one example of how student achievement could be used to make student achievement a more robust element of the three-tiered system.

Student performance could then become the foundation for the way teachers and districts are compensated. By making licensure level more dependent on student performance and aligning the Training and Experience (T&E) index to the three-tiered system, the state can allocate resources to districts and teachers making the most impact in driving student achievement.

The system could allow exceptionally impactful teachers to ascend to level III licensure without submitting a PDD. Allowing high-performing teachers to move more freely in the three-tiered system will improve student achievement amongst teachers. Teachers who have proven themselves among the best educators in the state consistently and are not already a level III teacher could have a way to bypass the traditional PDD process and become a level III teacher based on their success in driving student achievement. A strengthened PDD is essential for other teachers in non-tested grades and subjects.

Recommendations:

The state should set performance expectations for licensure levels.

The Legislature should:

- Replace the PDD and establish the effective teaching portfolio (ETP) as part of the licensure advancement application with new requirements. The ETP will have new requirements strengthening expectations for student achievement. The legislature should adopt the following requirements for licensure advancement:
  Advancement to any licensure level: three years of classroom teaching at Level I before advancement; three years of satisfactory annual local evaluations; and superintendent approval of advancement and verification of submittal information.
  Advancement from license Level I to Level II: require one year of mentor program; meet performance expectations as demonstrated through an ETP – OR – qualify through Performance Ranking: Level I teachers ranked in the highest 50 percent of all level II teachers for three consecutive years can bypass ETP requirements and be promoted to level II.
  Advancement from license Level II to Level III: meet performance expectations as demonstrated through ETP – OR – qualify through Performance Ranking: Level II teachers ranked in the highest 50 percent of all level III teachers for three consecutive years can bypass ETP requirements and be promoted to level III;
- Modify statute to clarify that public school remedy for non-performance includes non-renewal of contract, or other action (suspension or termination) in accordance with other existing due process laws.

PED should:

- Create a Value-added model to estimate teacher effects on student performance.

The state should have a licensure renewal process that uses student achievement as a primary determining factor in the process.

The Legislature should:

- Create the following licensure terms for Level I, II and III licenses.
  Level I licenses should have a five-year license term and teachers must submit for advancement after three years; Level II and Level III licenses should have an eight-year term and teachers must submit for renewal after six years.

- Create the following requirements for license level renewals for Level II and III licensure holders:
  Level II requirements for renewal: satisfactory score on effectiveness evaluation for most recent three year period; satisfactory score on student achievement portion of ETP –OR – VAM ranking of 'meets expectations' within licensure level (Level II); a teacher failing to meet renewal requirements within license term may apply for a provisional Level II license and demonstrate satisfactory performance within two years.
  Level III requirements for renewal: satisfactory score on effectiveness evaluation for most recent three year period; satisfactory score on student achievement portion of ETP –OR- VAM ranking of meets expectations within licensure level (Level III) statewide; a teacher failing to meet renewal requirements within license term may apply for a provisional Level III license and demonstrate satisfactory performance within two years.

The state should update teacher competencies to incorporate student achievement, and reflect recent research and common core standards.
The Legislature should:

- Replace the PDD and establish the effective teacher portfolio (ETP) as part of the licensure advancement application. The ETP should include overhauled competencies that reflect current research, are updated to reflect the new common core standards, and provide evidence of effective teaching practice. New teacher competencies should focus on three areas:
  1) instruction, professional development and student learning: instruction includes evidence of instructional plans, assessment techniques, use of data to inform practice, adaptation of teaching for diverse learners, classroom management, and implementation of state content standards;
  2) professional development includes evidence of meeting professional development goals, collaborating with other educators, parent involvement, or research and publication;
  3) student learning includes evidence of improved student achievement on PED-approved assessments using at least three years of data. Evidence of student learning should constitute at least 50 percent of the overall ETP score, which should be heavily rooted in student year-over-year gain scores on the SBA math and reading assessments.

PED should:

- Establish updated basic competency and effectiveness indicators for teachers, as part of the ETP, including setting new student achievement expectations for Level II teachers and more rigorous student achievement expectations Level III teachers.

The state should strengthen local evaluations to better drive student performance.
The Legislature should:

- Require principals to receive training at least once every two years to improve evaluation skills;
- Strengthen statutory requirements for a highly objective uniform standard of evaluation (HOUSE) for teachers by requiring the following:
  - Professional Development Plan by 40th day establishing the current year's performance goals, including measurable objectives for student performance. The goals should be based on Basic Competency and Effectiveness Indicators, the previous year's annual evaluation, and previous year's students' performance;
  - Performance Evaluation: Annual evaluations should be based on whether the teacher met or exceeded expectations on Basic Competency and Effectiveness Indicators, made satisfactory progress on professional development goals, and received satisfactory ratings from students and parents. Classroom observations from principals;
  - Performance Improvement Plan: Establish a structure to provide assistance to teachers not meeting expectations.

- Local Schools: Create policies and procedures to implement this section and authorize PED to approve additional options and measures for a local system of data collection for the annual teacher performance evaluation, including the use of peer observations.

PED should:

- Establish an effectiveness evaluation for career teachers (level II and III):
  - After three years of classroom teaching require an effectiveness evaluation to be conducted no later than the 40th day the following school year and include three-year summaries of progress meeting Basic Competency and Effectiveness Indicators; improving student achievement component should count for no less than 50 percent an overall rating;
  - The summative effectiveness evaluation includes a cumulative assessment of a teachers' effectiveness at improving student achievement over time, as measured by PED expected student performance growth targets on the ETP. Performance expectations should be aligned with the three-tiered licensure levels, and subject and grade level standards; public schools may award teachers with successful effectiveness evaluations multi-year contracts not to exceed the equivalent term of a contract of the district's superintendent. Public schools may use the results of the effectiveness evaluation to make employment decisions, in accordance with other provisions of law.
- Require professional development plan by the 40th day establishing the current year's performance goals, including measurable objectives for student performance. The goals should be based on updated basic competency and effectiveness indicators, the previous year's annual evaluation, and a previous year's students' performance.
- Provide that a performance evaluation be conducted annually, and be based on whether a teacher exceeds, meets, or does not meet expectations on basic competency and effectiveness indicators, professional development goals, and satisfaction from parents. All teachers must be observed by principals three times a year.

WHEN USED APPROPRIATELY, VALUE-ADDED MODELS CAN IDENTIFY EFFECTIVE TEACHERS AND DRIVE STUDENT ACHIEVEMENT

Value-added models are increasingly used across the country to evaluate teacher performance. As of October 2012, the U.S. Department of Education granted 33 states, including New Mexico, a waiver from some of the requirements of No Child Left Behind for changing their teacher evaluation systems to incorporate student data. Many proposals included use of VAMs for 50 percent of a teachers' evaluation rating. VAMs have the potential to inform stakeholders about teacher performance, but the volatility in these models warrants caution moving forward because of potential misclassification of teachers.

Depending on the demographic factors used, value-added models produce varied results. Some VAMs attempt to control for demographic factors and may use multiple years of scores on a handful of different assessments, while others do not. Value-added experts debate the meaning of these different models.

In 2012, Pearson Education, Inc. published a study comparing five different VAM teacher evaluation approaches and concluded VAM results are not definitive, and will depend on the model used. **Appendix C** provides a comparison of these five models. The study used data from a large school district in Texas and included data on demographic variables such as gender, ethnicity, English proficiency, special education status, and FRL.

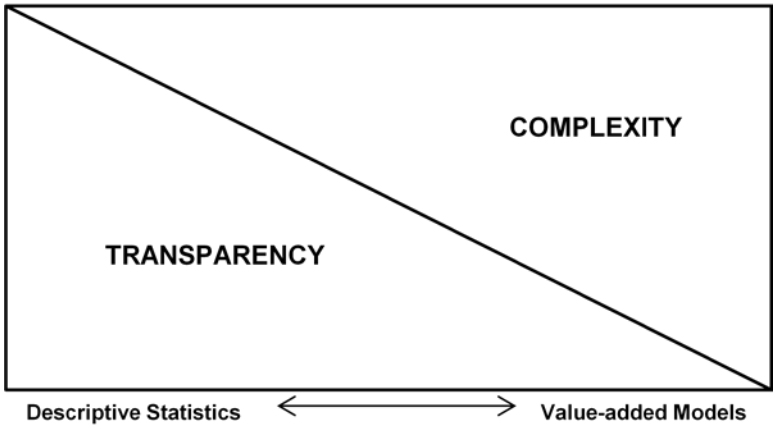
Chicago Teacher's Union Strike

In the fall of 2012, teachers in Chicago Public Schools, the nation's third largest school system, struck. Several reasons were cited for the strike, but, importantly, they included opposition to a system that would use student achievement data for 45 percent of teachers' evaluations. During the strike, researchers from 16 Chicago-area universities wrote an open letter to Chicago Mayor Rahm Emanuel warning against relying too heavily on VAMs in teacher evaluations. The New York Times later ran a rebuttal piece advocating for their use as prescribed. Chicago's students lost seven school days of school to the strike, which ended with a new contract agreement. The major tenants of this contract included a 17 percent pay raise for teachers and an evaluation system using VAMs for 30 percent, the minimum required by Illinois law.

Some VAMs produce unstable scores across years. All VAMs return different scores for the some teachers in different years. This could reflect a change in the effectiveness of a teacher between school years or the way a VAM score is calculated. Averaging VAM scores across a set number of years reduces this volatility, but such an approach does not help new teachers, teachers moving schools or grade levels, or teachers with small numbers of students from special populations.

More complex models are more consistent from year-to-year, but they can be difficult to interpret and expensive to implement. Multi-level VAM models, like the one PED uses for school grading, are complex and hard to explain to policymakers and stakeholders. The Pearson study indicates the importance of communicating the model and interpreting results to stakeholders and identifies this as an issue when choosing a model to use.

Figure 1. Understandability of Statistical Models



Additionally, the more complex the model, the more likely it will require custom programming; as a result, VAMs are increasingly being run by private for-profit corporations. For example, the software company SAS has gained the rights to a VAM approach developed by Dr. William Sanders in Tennessee and is now marketing the hosts data along with VAM analysis for districts and states.

Some value-added models adversely affect educators teaching certain populations of students. The Value-added Model Research Group at the University of New Mexico's College of Education used teacher and student data to determine scores for teachers from two different VAMs on teacher scores. One VAM incorporated only test scores (test-score only model), while the other compensated for contextual variables, such as poverty and English language fluency (student demographic model).

The two models placed teachers into one of five performance categories based on their students' performance.

Table 6. Performance Categories and Percentile Ranges of VAM Scores

Performance Level	Percentile Range
Ineffective	2nd percentile and below
Needs Improvement	2nd to 16th percentiles
Meets Expectations	16th to 84th percentiles
Highly Effective	84th to 98th percentiles
Exceptionally Effective	98th percentile and above

See **Appendix D** for a methodology behind the value-added models and the performance categories.

Mr. Wilson – Gifted Education Teacher

Mr. Wilson teaches at a large, urban district. In 2011, 97 percent of Mr. Wilson's students were classified as gifted. Twenty percent of his students qualified FRL and 42 percent were Hispanic. As a teacher of mostly gifted students, Mr. Wilson's VAM rating would vary depending on whether student demographic factors were included in the model.

How Different VAMs Affect Mr. Wilson's Status

	Test Score Only Model (no student demographic factors)		Student Demographic Model (includes all available student demographic factors)	
	Math	Reading	Math	Reading
1 year of data	Exceptionally Effective	Highly Effective	Ineffective	Ineffective
2 years of data	Highly Effective	Meets Expectations	Needs Improvement	Needs Improvement

Because Mr. Wilson specializes in teaching a high-performing group of students and improves their student achievement, models that do not incorporate student demographics reflect his effectiveness. Models that do incorporate student demographics penalize Mr. Wilson because his students are gifted and a relatively lower proportion of them are in poverty.

Teachers with high proportions of gifted students, students in special education, or poor students are affected by the model applied to them. Depending on whether the test-score only model or the student demographic model is used, some teachers with these populations of students have value-added scores showing them as high-performing or low-performing. For example, in a model without student demographics, teachers with high levels of FRL students have lower scores than a model including student demographics.

Table 7. Special Populations Adversely Affected When the Test Score Only and Student Demographic Models are Applied

Test Score only-model	Student Demographic Model
English-language learners	Gifted
Free Lunch	
Special Education C and D	

Source: UNM

Many teachers' ratings change depending on the value-added models used to estimate their impact on student achievement. After applying and comparing the results of both models, 77 percent of the teachers evaluated, stayed in the same performance category. The choice of VAM used affects the rating of 23 percent of teachers who moved from at least one performance category to another. Twelve percent increased at least one performance category and 11 percent decreased at least one performance category.

Mrs. Martinez – Special Education Specialist

Mrs. Martinez teaches 4th grade in a western New Mexico school district. In 2012, 68 percent of Mrs. Martinez's class was FRL. Mrs. Martinez has experience as a special education teacher and has shown the ability to drive their improvement, so nearly all of the students assigned to her class have IEPs. Half of her students are Caucasian. Mrs. Martinez's value-added scores are inconsistent across models.

How Different VAMs Affect Mrs. Martinez's Status

	Test Score Only Model (no student demographic factors)		Student Demographic Model (includes all available student demographic factors)	
	Math	Reading	Math	Reading
1 year of data	Ineffective	Needs Improvement	Meets Expectations	Highly Effective
2 years of data	Meets Expectations	Meets Expectations	Meets Expectations	Highly Effective

Teachers like Mrs. Martinez, with a high proportion of their students in special education, can be significantly affected by the type of model selected to evaluate them. Because of the population of students she teaches (nearly all special education), Mrs. Martinez could be judged as one of the state's very best teachers or one of the worst.

Value-added models are limited in what they can tell educators, the public, and other stakeholders. VAMs compare teacher performance to one another, making them a relative concept. One set of scores from VAMs do not indicate whether quality instruction is occurring in classrooms, only whether one teacher has students with higher achievement gains than another. Over time, VAM scores across multiple years can create a picture of absolute teacher performance that can be used for licensure advancement.

Value-added models used in other states can distinguish high-performing teachers from low-performing ones with some certainty, but cannot reliably distinguish between the middle-performing teachers. In New York, 77 percent of teachers earned a rating of "effective," meaning their students grew "equal to the average for similar students." The state's evaluation system gives teachers one of four overall ratings: highly effective, effective, developing and ineffective. Seven percent of teachers earned a rating of "highly effective," and 6 percent earned a rating of "ineffective."

Ms. Campos – Teacher of At-risk Students

Ms. Campos teaches 3rd grade in small, rural school district. Over 90 percent of her students are FRL and are classified as English language learners. Additionally, all of her students participate in special education and 100 percent of her students are Native American. As a teacher of this highly at-risk group of students, Ms. Campos' value-added rating in a VAM depends heavily on which model is applied.

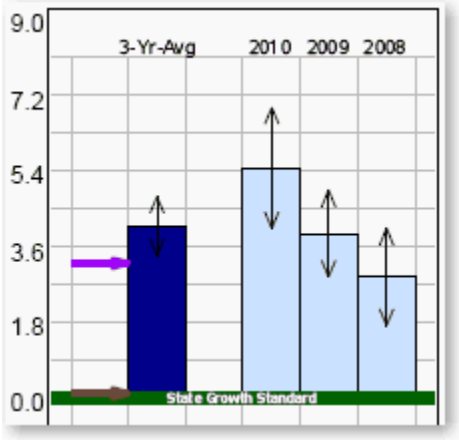
How Different VAMs Affect Ms. Campos' Status

	Test Score Only Model (no student demographic factors)		Student Demographic Model (includes all available student demographic factors)	
	Math	**Reading**	**Math**	**Reading**
1 year of data	Needs Improvement	Needs Improvement	Highly Effective	Highly Effective
2 years of data	Needs Improvement	Meets Expectations	Needs Improvement	Meets Expectations

Ms. Campos moves between one of the lowest performance categories to the highest. Using a model with no student demographic factors could discourage effective teachers from accepting positions in low-income schools.

A large proportion of scores from value-added models have a high degree of uncertainty. VAMs are unable to pinpoint the exact ranking of a teacher and instead provide wide estimates of performance. For example, in the sample below taken from the Tennessee Value-Added Assessment System, the teacher received a VAM score of 5.4 in 2010. However, the teacher's real score could be anywhere between approximately four and seven after taking statistical significance into account.

Chart 17. Example of Value-added Model Scores from Tennessee

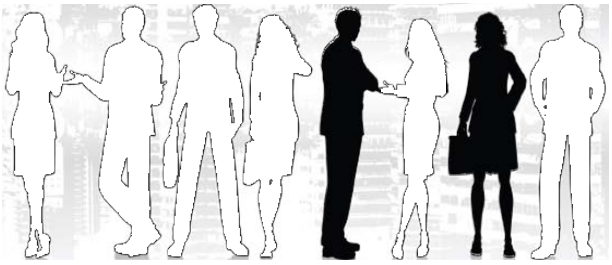


Source: Tennessee Department of Education

Using the SBA, currently the only assessment appropriate for statewide value-added models, approximately 20 percent of teachers can be evaluated. Standardized tests are necessary inputs for VAMs. Because the SBA assesses only reading, math, and science, most students and teachers in the state cannot be evaluated under the model. This includes teachers of social studies, vocational education programs, art, and more. Even science teachers could be excluded from a model because the science SBA is only administered in grades 4, 7, and 11, and consecutive years may be needed to compute VAM scores. A district using additional standardized assessments could use a VAM on teachers within the district. Some VAMs use a school's aggregate score to evaluate these teachers.

Figure 2. Teachers in New Mexico

Figure 3. Teachers in New Mexico for whom we can compute VAM scores



Principal evaluations and some VAMs are identifying the same high-performing teachers. One study of principals in New York City found teachers rated more effective by their principals were more likely to also have high value-added test scores.

Some VAMs may not include many students due to mobility, test exemption, and absenteeism. VAMs use data from multiple years. If a student moves from one school to another during the school year or between school years or is absent for much of the school year, that student will not meet requirements to be considered as attending a full academic year. Furthermore, students who do not take the SBA or receive exemptions will not count in a VAM. Alternative methods exist for filling in missing scores in statistical modeling.

The use of VAMs can be responsibly integrated into the three-tiered system to identify teachers for advancement and bonus pay. VAM methodologies being developed for teacher evaluation and school grading could be leveraged to measure performance at tier levels and could be used as benchmarks in advancement between levels. Specifically, VAMs can be used to reward good teachers and identify poor teachers for professional development as part of a larger system of teacher evaluation.

VAMs can do a good job of identifying highly effective and highly ineffective teachers for rewards and interventions. Once teachers of certain populations that experience variability are identified and controlled for, research has shown VAMs do a good job of identifying very good and very poor teachers, but do not differentiate between teacher scores in the middle. Districts and states can reliably use this data to reward very good teachers and put them in roles allowing them to share expertise, such as a mentor. Similarly, identifying low-achieving teachers allows districts to strategically align resources.

School leaders have shown they use value-added data to drive professional development and improvement in classroom instruction. Research at Columbia University showed principals given teachers' VAM scores and performance data used the data to make personnel decisions and plan their teacher professional development. There were no formalized expectations for use of the data but principals, after being provided training on how to interpret scores, used the data in ways they found most advantageous for the school.

VAMs should not be used in annual local evaluations because of their limitations and complexity. Because VAMs do a good job of identifying very good and very poor teachers but do not do a good job of differentiating between teacher scores in the middle, the use of VAM in annual local evaluations could provide inaccurate reflections of teacher performance. Additionally, the use of multiple years of VAM data to boost reliability would hide recent changes in teacher performance. For example, if a teacher made a large gain between the previous and current year, this progress could be diminished by averaging the most recent score with multiple previous years of data. Additionally, a VAM score is a relative measure which provides information on how a teacher performs compared with other teachers. Teachers working toward definitive goals such as SBA proficiency scores of 40, an absolute measure, would not be provided useful information by a relative VAM score.

VAMs should be used cautiously. VAMs should not be applied to all teachers of students with test scores, because of special circumstances. For example, some special education teachers specializing in work with severely high-needs special education populations might not be appropriate to score on a VAM because of their students' limited ability to take the assessment. Similarly, teachers of exceptionally high-scoring students are less able to raise students' scores, and should not be subjected to VAM scores or their associated consequences. Exceptions for the small minority of teachers in these circumstances should be made when considering how best to implement VAMs.

Recommendations:

The state should take advantage of value-added models to

The Legislature should:

- Require PED to annually rank the performance of licensed teachers providing instruction in tested grades and subjects through a valid value-added model. Results will be provided only to public schools and the individual teachers. The results should benchmark performance relative to teachers statewide, within the district, the school, and license level by grade taught and overall. The results should be part of a teacher's personnel file, confidential, and only available for review by authorized personnel.
- Create a fast-track licensing process for teachers that receive the very highest statewide value-added scores; teachers consistently (3 years or more) receiving the highest scores could apply for level III licensure regardless of their current licensure status.
- Allow findings from future research by the Value-added Model Research Group at UNM, which aims to validate the accuracy of the VAM rankings and conduct research on special education population's effects on the state, to guide future decisions regarding the use of VAMs.

PED should:

- Use two different VAMs to obtain a composite score to help eliminate VAM biases for teachers of certain populations;
- Allow findings from future research by the Value-added Model Research Group at UNM, which aims to validate the accuracy of the VAM rankings and conduct research on special education population's effects on the state, to guide future decisions regarding the use of VAMs.

RESOURCE ALLOCATION AMONGST DISTRICTS AND SCHOOLS CREATES FUNDING DISPARITIES BETWEEN LOW-INCOME STUDENTS AND THEIR MORE AFFLUENT PEERS, WITHOUT DRIVING STUDENT ACHIEVEMENT

New Mexico directs nearly $200 million for higher teacher compensation through the T&E index in the public school funding formula. T&E index values are based on teachers' highest degree earned and years of experience. That index value is multiplied by student demographic and program units already generated in the formula. Funds generated by the T&E index in the funding formula compose up to 20 percent of a district's formula funding.

Table 8. Percentage of Formula Funding from the T&E Index

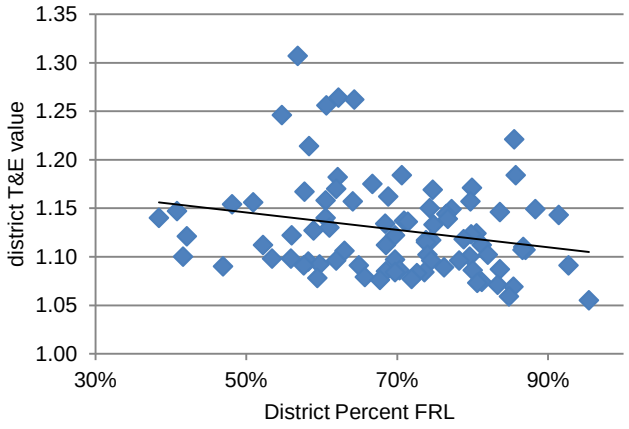
District	Share of Total Formula Funding
Carlsbad	19.8%
Tatum	16.6%
Texico	15.2%
Silver City	14.2%
Animas	13.7%
Ruidoso	12.9%
Cobre	12.8%
Artesia	12.7%
Tularosa	12.7%

Source: PED

The T&E index directs more funding to more affluent school districts and produces a questionable return on investment after factoring in poverty. The T&E does not recognize better performance by teachers and higher pay, but instead rewards relatively affluent districts for keeping teachers and sometimes requiring them to meet higher education requirements. Aligning the T&E index to a modified three-tiered system that focuses on student performance will allow the state to send resources to high-performing teachers and schools.

High-poverty, rural districts with the greatest needs generally have difficulty hiring experienced teachers with advanced degrees who increase the T&E index. As FRL participation levels increase, T&E values drop. Because poverty is also highly related to lower SBA scores, districts with lower T&E values sometimes tend to have lower SBA scores than their peers.

Chart 18. District T&E and Free and Reduced-Priced Lunch Levels



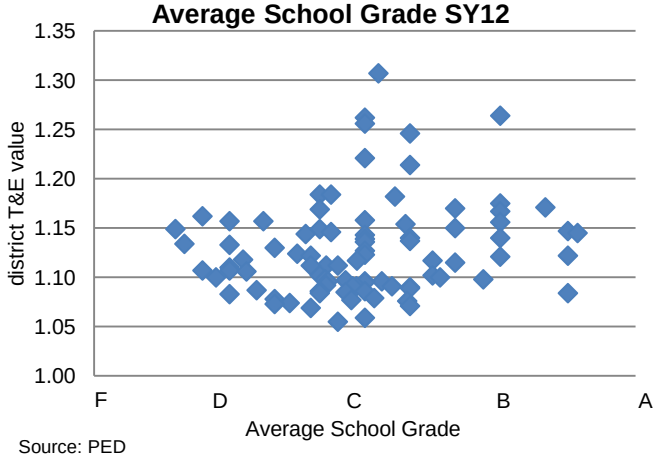
Source: PED

An independent study conducted for the Maddox Foundation also found "districts with a high T&E index tend to have relatively fewer at-risk students." The report recommended the state drop the T&E index from the formula and develop a categorical aid program focused on providing funds for high-need districts to improve teacher qualifications and recruit teachers in hard-to-staff schools and subjects.

The T&E index encourages higher education levels and more years of service, despite inconclusive evidence these factors increase student achievement. For example, some school districts, including those with some of the highest T&E values, mandate teachers earn a master's degree before a certain number of years serving with the district. No clear body of research links higher education level or more years of service with better student outcomes or achievement. A 2009 report from the Urban Institute states characteristics such as graduate education and experience are at best weak predictors of a teacher's contribution to student achievement. However many states, including New Mexico, continue to use education and experience as the basis for teacher compensation.

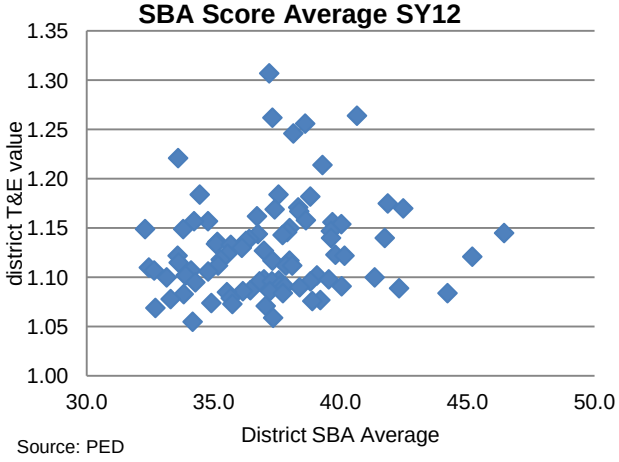
After accounting for district free and reduced-price lunch levels, T&E has no effect on district school grades, average test scores, or percent of students proficient. T&E has no relationship with district achievement as measured in numerous ways including using PED's school grading system and district SBA score average.

Chart 19. District T&E and Average School Grade SY12



Source: PED

Chart 20. District T&E and SBA Score Average SY12



Source: PED

Even districts with the very highest T&E values generally have average school grades. Of the five districts that earn the highest T&E values in the public school funding formula, half of them average C grades or less, and only one district averaged a B grade. Many districts that claim lower-than average T&E values have better school grades than many of the highest T&E earners; for example, Rio Rancho and Moriarty school districts, which each claim a 1.1 on the T&E index, had average school grades of 2.6 (C) and 2.9 (C), respectively.

Table 9. Districts with the Highest and Lowest T&E Index Values with Average School Grades and SBA SY12

District	Average School Grade	T&E Index	SBA Average
Tatum	2.1 (C)	1.31	38.4
Animas	3.0 (B)	1.26	40.9
San Jon	2.0 (C)	1.26	38.7
Carlsbad	2.0 (C)	1.26	39
Texico	2.3 (C)	1.25	38.6
Tucumcari	2.3 (C)	1.07	38.1
Jemez Mountain	1.6 (D)	1.07	34.1
Santa Rosa	2.0 (C)	1.07	38.8
Hatch	1.8 (D)	1.06	34.2
Lake Arthur	2.0 (C)	1	38.5

Source: PED

The T&E index is not aligned to the three-tiered system. The three-tiered system provides large salary changes not accounted for in the T&E index. For several years the LFC has noted, in its present form, the T&E index is not aligned to the three-tiered system. An evaluation of the public school funding formula conducted jointly by the LFC and the LESC, the New Mexico Effective Teaching Task Force final report, and the AIR funding formula study recommended better alignment of the T&E index with the three-tiered system. For example, in 2011, the LFC and LESC recommended the T&E index be replaced with an effective teacher index that assigns values to teachers based on licensure level, not education level and experience.

Forty percent of teachers in New Mexico charter schools are level I teachers, more than double the percent in Albuquerque Public Schools (APS), the district with the most charters. While the proportion of level I teachers varies throughout the state's regions, charters have between two to three times the proportion of level I teachers. Subsequently, their level II and level III teachers are lower than all regions throughout the state.

Table 10. Ratio of Teacher Licensure Level by Region SY12

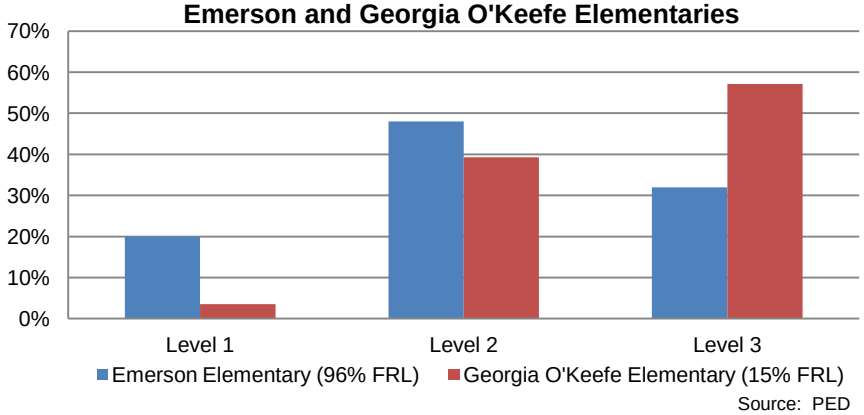
Region	Level I	Level II	Level III
APS	19%	44%	37%
NW	18%	47%	35%
NE	12%	52%	36%
SE	17%	44%	38%
SW	17%	44%	39%
Charter	40%	33%	28%

Source: PED

Level III teachers are more likely to teach in more affluent districts and schools. While districts do not have explicit policies to move more experienced and educated teachers to more affluent schools, students in poverty are more likely to have a less experienced, poorer-performing teacher. Recruiting and retaining high quality teachers in low-income schools is integral to ensuring students in poverty achieve academically.

Schools with higher poverty rates have teachers with lower licensure levels than more affluent schools. Based on an analysis of four school districts, Albuquerque, Las Cruces, Rio Rancho and Santa Fe, schools with higher proportions of FRL students have staffs with lower licensure levels. For example, in APS, Emerson Elementary and Georgia O'Keefe Elementary have similar staff sizes but different student populations: 96 percent of Emerson's students are FRL compared with 15 percent at O'Keefe. Over half of Georgia O'Keefe's teachers have a level III license and the school has only one level I teacher. Emerson's teaching staff is mostly level I and level II teachers.

Chart 21. Comparison of Staff Licensure Levels at Emerson and Georgia O'Keefe Elementaries



Source: PED

Teachers at high-poverty schools are paid less than those at low-poverty schools. Because high-poverty schools have more teachers with lower licensure levels, the average pay is lower. These districts do not have explicit policies placing level III teachers in more affluent schools.

Table 11. SY12 Average Salaries at High and Low-Poverty Elementary Schools in Selected Districts

District	Elementary School	Average Salary	Free and Reduced-Price Lunch Rate
APS	Emerson	$42,900	96%
	Georgia O'Keefe	$46,923	15%
Santa Fe	R.M. Sweeney	$41,503	87%
	Wood-Gormley	$49,506	19%
LCPS	Booker T. Washington	$43,714	90%
	White Sands	$46,385	30%

Source: PED

Research shows teachers at high-poverty schools are less effective than their counterparts in more affluent schools. A Duke University study of high and low-poverty schools in North Carolina found students in the high poverty schools are served by teachers with lower qualifications than those in lower poverty schools. Researchers found these qualifications were connected to higher student achievement. The study considered competitiveness of the teachers' undergraduate institution, teacher scores on licensure exams, national board certification, and years of experience.

Offering a mix of incentives to recruit and retain good teachers in high-poverty schools can work. Research finds creating incentives to get highly qualified and effective teachers to teach in high-poverty schools can work, but keeping effective teachers is more challenging. Financial incentives can recruit high-quality teachers and slightly decrease turnover in the short-term, but money does not work in the long-term to keep teachers at low-income schools: "Even when bonuses succeeded in drawing teachers to the poorest schools, such incentives could not compensate for the lack of support they encountered in these schools, which in turn contributed to the departure of many of these teachers."

Financial incentives attack part of the problem, but do not solve working conditions. Districts must find ways to incentivize the best administrators to lead high-poverty schools and give them added support.

Good results in challenging environments and poor results in advantageous settings

In SY12, Teacher X made $50 thousand a year and had a 100 percent of their students qualify for free and reduced-price lunch. Teacher X is an effective teacher moving more than half of the class up in proficiency level with no children losing a proficiency level. Students averaged a five point gain in SBA scores across students.

The same year, Teacher Z made $53 thousand a year in another district and has a 0 percent participation in the free and reduced-price lunch program. Teacher Z is a less effective teacher, having twice as many students in their class losing proficiency than gaining and with students averaging a two point loss in SBA scores across students.

Recommendations:

The state should align resources to allocate funding to districts in a way consistent with teacher pay scales that incorporate performance.
The Legislature should:

- Change the T&E index to an effective teacher index that rewards districts based on the number of teachers they have in each license level. This should be accomplished over a two-year time period: in year one, the current T&E index should be multiplied by only membership units in the formula, and in year two, the effective teacher index should be fully implemented.

The state should incentivize high-performance among teachers and provide incentives for teaching in high-need schools.
The Legislature should:

- Require only teachers meeting or exceeding expectations on annual performance evaluations receive state or district funded salary increases the subsequent year; and
- Consider a mechanism, possibly through the funding formula, to provide additional compensation to effective teachers (as measured by the new aforementioned teacher evaluation and three-tiered licensure system) to teach in Title I 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

HANNA SKANDERA
SECRETARY-DESIGNATE OF EDUCATION

SUSANA MARTINEZ
Governor

November 13, 2012

Mr. David Abbey, Director
Legislative Finance Committee
325 Don Gaspar, Suite 101
Santa Fe, NM 87501

RE: Teacher Effectiveness Evaluation

Dear Director Abbey:

Thank you for the opportunity to respond to the draft evaluation on Effective Use of Test Data to Assess and Improve Teacher Evaluation in New Mexico. Please accept my compliments to your staff for their professionalism and collaborative approach throughout the evaluation process. As always, the Public Education Department (PED) is committed to continuous quality improvement, best practices, and positively impacting outcomes for all of our students in New Mexico.

The evaluation appears to be thorough and objective and points to a number of issues that we believe are necessary to implement a robust teacher evaluation system that will lead to a more effective workforce focused on ensuring students receive the education they need to thrive and survive in the 21st century. We are encouraged that the evaluation aligns with so many of the components of the newly developed evaluation system and in those minor areas where PED had concerns. We are pleased that the LFC has taken comment and made minor changes accordingly.

A key component of the LFC evaluation is the need to reward highly effective teachers, those that through thick and thin, make a difference in each student's life. I encourage the LFC to work toward changes in the funding formula that lead to these rewards, while holding these teachers to the highest standards for the benefit of all.

The exit conference between LFC and PED was held Wednesday November 7, 2012 and the draft report was discussed. The department does not have any recommended changes at this time. However, we note that several of the recommendations made in the LFC report are already being addressed as part of the teacher evaluation pilot currently operating in 68 schools across 21

districts. We look forward to working with your staff as the new evaluation system unfolds to ensure an evaluation process that is robust, fair and truly focuses on improving the teaching skills of all teachers.

Thank you again for the opportunity to comment on the evaluation.

Warm regards,

Hanna Skandera
Secretary-Designate
Public Education Department

HS/at

APPENDIX A: Project Information

Evaluation Objectives.

- Follow-up on LFC's 2009 three-tiered system report.
- Investigate T&E index and return on investment.
- Evaluate existing value-added model (VAM) methods and outputs.

Evaluation Procedures.

- Interviewed district and state-level administrators.
- Reviewed state, district, and school-level student performance data and student demographic data.
- Conducted an online survey of teachers that have gone through the three-tiered system.
- Reviewed and determined the impact of the current three-tiered system. Additional descriptive and inferential statistics, along with specific methodologies will be made available in a separate publication.
- Reviewed applicable laws and regulations; previous research reports; LFC file documents, including all available project documents; relevant performance reviews from other states; and performance measures.
- Evaluated VAM methodologies in collaboration with the Value-added Research Group at the University of New Mexico.

Evaluation Team.

Matthew Pahl, Lead Program Evaluator
Dr. Jon Courtney, Program Evaluator
Elaine Romero, Program Evaluator

The Value-added Research Group at the University of New Mexico calculated value-added scores for teachers using teacher and student data. The group conducted analyses using two value added models, which are described in detail in **Appendix C.** The Value-added Research Group included the following members:

Dr. Richard Howell, Dean of the College of Education, University of New Mexico
Dr. Richard Bowman, Chief Accountability and Strategy Officer, Santa Fe Public Schools
Dr. Kristin Umland, Associate Professor, Department of Mathematics and Statistics, University of New Mexico
Dr. James Selig, Assistant Professor, College of Education, University of New Mexico
Dr. Laura Kapitula, Assistant Professor, Department of Statistics, Grand Valley State University

Authority for Evaluation. LFC is authorized under the provisions of 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. LFC is also authorized to make recommendations for change to the Legislature. In furtherance of its statutory responsibility, LFC may conduct inquiries into specific transactions affecting the operating policies and cost of governmental units and their compliance with state laws.

Exit Conferences. The contents of this report were discussed with Secretary-designate Skandera and Senior PED staff on November 7, 2012.

Report Distribution. This report is intended for the information of the Office of the Governor; the 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.

Charles Sallee
Deputy Director for Program Evaluation

New Mexico Teacher Competencies for Licensure Levels I, II, and III
Assessment Criteria

New Mexico is one of the most diverse states in the nation, and this diversity is reflected in the strengths and needs of New Mexico's students. The ability of a highly qualified teacher to address the learning needs of all New Mexico's students, including those students who learn differently as a result of disability, culture, language, or socioeconomic status, forms the framework for the New Mexico Teacher Competencies for Licensure Levels I, II, and III-A Assessment Criteria Benchmarks.

1. The teacher accurately demonstrates knowledge of the content area and approved curriculum

Provisional Teacher - LEVEL I	Professional Teacher - LEVEL II	Master Teacher - LEVEL III
A. Utilizes and enhances approved curriculum.	A. Enhances and extends approved curriculum.	A. Contributes to the refinement and development of the approved curriculum.
B. Gives clear explanations relating to lesson content and procedures.	B. Gives clear explanations relating to lesson content and procedures.	B. Provides clear explanations relating to lesson content and procedures in multiple ways and is aware of knowledge and preconceptions that students can bring to the subject.
C. Communicates accurately in the content area.	C. Communicates accurately in the content area.	C. Communicates accurately in the content area and can create multiple paths to the subject matter.
D. Shows interrelatedness of one content area to another.	D. Integrates other subjects into the content curriculum.	D. Can articulate to students the interrelatedness of the disciplines.

2. The teacher appropriately utilizes a variety of teaching methods and resources for each area taught.

Provisional Teacher - LEVEL I	Professional Teacher - LEVEL II	Master Teacher - LEVEL III
A. Provides opportunities for students to work independently, in small groups, and in large groups.	A. Designs appropriate opportunities for large group, small group, and independent student learning experiences.	A. Designs and engages students in large group, small group, and independent work activities.
B. Uses a variety of methods, including demonstrations, lectures, student initiated work, group work, questioning, and independent practice.	B. Selects from a variety of teaching methods (demonstrations, lectures, student projects, group work, independent practice) for specific instructional goals and purposes.	B. Demonstrates effective selection and use of a variety of methods to make knowledge accessible to all students.
C. Uses a variety of resources such as field trips, supplemental printed materials, manipulatives, and technology.	C. Integrates a variety of resources into instruction, including field trips, supplemental printed materials, manipulatives, and technology.	C. Demonstrates effective integration of a variety of resources and learning experiences into the curriculum.

Assessment Criteria *Benchmarks* for New Mexico Teacher Competencies for Licensure Levels I, II, and III

II. The teacher appropriately utilizes a variety of teaching methods and resources for each area taught (continued)		
D. Provides opportunities for students to apply, practice, and demonstrate knowledge and skills learned through various modalities.	D. Demonstrates understanding and appropriate application of learning styles, modalities, and intelligences theories.	D. Designs opportunities for students to apply, practice, and demonstrate knowledge and skills based on knowledge of learning modalities, style preferences, and intelligences.
E. Implements necessary modifications and adaptations in instruction and curriculum so that students with disabilities have access to the general education curriculum in the least restrictive environment.	E. Designs and implements necessary modifications and adaptations in instruction and curriculum so that students with disabilities have access to the general education curriculum in the least restrictive environment.	E. Engages with colleagues and parents to collaboratively design and implement necessary modifications and adaptations in instruction and curriculum so that students with disabilities have access to the general education curriculum in the least restrictive environment.

3. The teacher communicates with and obtains feedback from students in a manner that enhances student learning and understanding.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Explains and/or demonstrates the relevance of topics and activities.	A. Effectively explains, demonstrates or communicates the relevance of topics and activities.	A. Engages students in explaining and/or demonstrating the relevance of topics and activities.
B. Communicates to students the instructional intent, directions, or plan.	B. Consistently communicates to students the instructional intent, directions, and plans.	B. Involves students in establishing instructional direction and plans.
C. Establishes and states expectations for student performance.	C. Establishes and states expectations for student performance	C. Establishes and states expectations for student performance.
D. Clarifies actions, directions, and explanations when students do not understand.	D. Presents directions and explanations in a variety of ways to insure student understanding.	D. Presents directions and explanations in a variety of ways to insure student understanding.
E. Actively solicits communication from students about their learning.	E. Solicits communication from students about their learning for the purposes of ongoing instructional planning.	E. Engages students in the analysis and evaluation of their learning and adjusts instruction based on student feedback.
F. Communicates regularly with students about their progress.	F. Communicates regularly with students about their progress.	F. Communicates regularly with students about their progress.

Assessment Criteria *Benchmarks* for New Mexico Teacher Competencies for Licensure Levels I, II, and III

4. The teacher comprehends the principles of student growth, development and learning, and applies them appropriately.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Instructs students in the use of cognitive thinking skills such as critical thinking, problem-solving, divergent thinking, inquiry, and decision-making.	A. Consistently integrates the use of cognitive thinking skills such as critical thinking, problem-solving, divergent thinking, inquiry, and decision-making into instruction.	A. Consistently integrates the use of cognitive thinking skills such as critical thinking, problem-solving, divergent thinking, inquiry, and decision-making into instruction.
B. Uses teaching techniques that address student learning levels, rates, and styles.	B. Adapts teaching techniques to accommodate a range of student learning levels, rates, styles and special needs.	B. Selects the most effective teaching techniques to address a variety of student learning levels, rates, styles and needs as well as diverse interests and backgrounds.
C. Uses materials and media that address student learning levels, rates, and styles.	C. Adapts materials and media to address a range of student learning levels, rates, styles and special needs.	C. Selects the most effective materials and media to address a variety of student learning levels, rates, styles and needs.
D. Uses resources such as community service agencies, school personnel, and parents to meet student learning levels, rates and styles.	D. Selects from a variety of community service agencies, specialized school personnel, and parents to address different learning levels, rates, styles, and needs.	D. Integrates community resources, service agencies, other school personnel, parents, and community members into the curriculum.

5. The teacher effectively utilizes student assessment techniques and procedures.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Uses a variety of assessment tools and strategies.	A. Selects appropriate assessment tools and strategies for specific learning outcomes.	A. Designs and uses multiple methods of measuring student understanding and growth.
B. Uses information gained from ongoing assessment for remediation and instructional planning.	B. Uses formative and summative assessment for remediation and instructional planning.	B. Integrates assessment data from multiple sources into instructional planning and improvement.
C. Maintains documentation of student progress.	C. Maintains documentation of student progress.	C. Maintains documentation of student progress.
D. Communicates student progress to students and families in a timely manner.	D. Consistently maintains communication with students and families about student progress.	D. Develops a two-way system of communicating with students and families about student progress.

Assessment Criteria *Benchmarks* for New Mexico Teacher Competencies for Licensure Levels I, II, and III

6. The teacher manages the educational setting in a manner that promotes positive student behavior and a safe and healthy environment.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Serves as a model for constructive behavior patterns.	A. Identifies, explains, and models constructive behavior patterns.	A. Integrates the teaching of constructive, pro-social behaviors into regular instruction.
B. Executes routine tasks effectively and efficiently.	B. Establishes and teaches effective and efficient routines.	B. Establishes and teaches effective and efficient routines.
C. Establishes and states expectations for student behavior.	C. Establishes and reinforces expectations for student behaviors that promote citizenship in a classroom community.	C. Engages students in establishing expectations for building a learning community in the classroom.
D. Handles transitions effectively.	D. Maintains smoothness and momentum during classroom transitions.	D. Maintains smoothness and momentum during instructional transitions.
E. Has materials and media ready for student use.	E. Prepares and arranges material in advance for easy student accessibility.	E. Establishes an environment where materials and media are available and ready for student use.
F. Minimizes distractions and interruptions.	F. Minimizes distractions and interruptions.	F. Minimizes distractions and interruptions.
G. Manages student behavior effectively and appropriately.	G. Monitors and directs student behavior effectively and appropriately.	G. Develops a classroom management system that promotes acceptable and appropriate student behavior.
H. Identifies hazards, assesses risks, and takes appropriate action.	H. Identifies hazards, assesses risks, and takes appropriate action.	H. Identifies hazards, assesses risks and takes appropriate action.

Assessment Criteria *Benchmarks* for New Mexico Teacher Competencies for Licensure Levels I, II, and III

7. The teacher recognizes student diversity and creates an atmosphere conducive to the promotion of positive student involvement and self-concept.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Demonstrates sensitivity and responsiveness to the personal ideas, learning needs, interests, and feelings of students with disabilities and/or from culturally and linguistically diverse backgrounds (e.g., Native Americans, Hispanic Americans, African Americans, Asian Americans, as well as other recent immigrant groups).	A. Acknowledges and validates the ideas, learning needs, interests, and feelings of students with disabilities and/or from culturally and linguistically diverse backgrounds (e.g., Native Americans, Hispanic Americans, African Americans, Asian Americans, as well as other recent immigrant groups.	A. Adjusts practice based on observation and knowledge of students with disabilities and/or from culturally and linguistically diverse groups (e.g., Native Americans, Hispanic Americans, African Americans, Asian Americans, as well other recent immigrant groups).
B. Acknowledges student performance and achievement.	B. Consistently recognizes student performance and achievements.	B. Creates curriculum designs that include student performance and acknowledgment of achievement.
C. Acknowledges that every student can learn.	C. Understands how students differ in their approaches to learning and adjusts instruction to meet diverse needs.	C. Demonstrates an awareness of the influences of context, disability, language, and culture on student learning.
D. Provides opportunities for each student to succeed and understands how students differ in their approaches to learning based on diverse cultural and linguistic backgrounds and exceptionalities.	D. Designs opportunities for each student to succeed, based on individual learning needs.	D. Provides accommodations and interventions that allow each student to succeed based on individual learning needs.
E. Provides students with opportunities for active involvement and creativity.	E. Designs specific activities that require active involvement and creativity.	E. Engages students in learning experiences that promote creativity, critical and divergent thinking.
F. Provides opportunities for students to be responsible for their behavior and learning.	F. Designs opportunities that require and reinforce student responsibility for learning.	F. Designs opportunities that require and reinforce student responsibility for learning.
G. Promotes positive student/teacher relationships.	G. Develops students' self-esteem, motivation, character, and sense of civic responsibility.	G. Fosters the development of respect for individual, cultural, linguistic, disability, and religious differences.
H. Encourages high student expectations.	H. Establishes and communicates high expectations for all students.	H. Engages students in setting high standards for performance.
I. Demonstrates an awareness and respect for each student's background, experience, learning ability, language, and culture.	I. Demonstrates knowledge of different student backgrounds, experiences, learning abilities, languages, and cultures and incorporates this knowledge into curricular decisions and instructional methodology.	I. Treats all students equitably, recognizing and planning for individual differences in cultures, languages, learning abilities, backgrounds, and experiences.

Assessment Criteria *Benchmarks* for New Mexico Teacher Competencies for Licensure Levels I, II, and III

8. The teacher demonstrates a willingness to examine and implement change, as appropriate.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Seeks out information on methodology, research and current trends in education to enhance and improve the quality of learning.	A. Seeks out information on methodology, research and current trends in education to enhance and improve the quality of learning.	A. Demonstrates the ability to reason, take multiple perspectives, be creative, and take reasoned risks to improve teaching.
B. Implements a variety of strategies to enhance learning.	B. Demonstrates knowledge of best practices that enhance learning.	B. Collaborates with colleagues in the research and design of improved instructional strategies
C. Recognizes that change entails risk and modifications may be needed.	C. Participates in instructional improvement and school reform initiatives.	C. Assumes a leadership role in the study and implementation of instructional improvement and school reform initiatives.

9. The teacher works productively with colleagues, parents and community members.		
Provisional Teacher - LEVEL I	**Professional Teacher - LEVEL II**	**Master Teacher - LEVEL III**
A. Collaborates with colleagues.	A. Actively promotes collegial relations with other school personnel.	A. Serves as a role model for collaborative working relations across the profession.
B. Communicates with parents on a regular basis.	B. Provides a system for interactive communication between teacher and parents.	B. Demonstrates knowledge of specific school, family, and community resources that can support student learning.
C. Uses conflict resolution strategies when necessary.	C. Uses conflict resolution strategies as appropriate.	C. Assists colleagues in the use of conflict resolution strategies.
D. Involves parents and community in the learning environment.	D. Promotes active roles for parents and community members in student learning.	D. Engages parents and community members productively in the work of the school.
E. Communicates in a professional manner with colleagues, parents, and community members regarding educational matters.	E. Communicates in a professional manner with colleagues, parents, and community members regarding educational matters.	E. Works collaboratively and creatively with colleagues, parents, and community members regarding educational matters.

APPENDIX C: Comparisons Among Value-Added Models

Different Value-added Model types and Associated Advantages and Disadvantages

Model Type	Advantages	Disadvantages
Percent Passing Change Model	1. Familiar to policymakers and stakeholders 2. Simple to calculate	1. Not technically a value-added model 2. Produces teacher measures that are confounded with differences between cohorts
Average Score Change Model	1. Familiar to policymakers and stakeholders 2. Simple to calculate	1. Requires vertically scaled scores across grades 2. Does not control for student characteristics that are unrelated to teacher effectiveness
Multiple Regression Model	1. Estimates teacher effectiveness as the residualized gain in student's current score after controlling for student's prior performance and demographic characteristics 2. Does not require vertically scaled scores across grades	1. Moderately complex and hard to explain to policymakers and stakeholders 2. Does not account for grouping effects
Hierarchical Linear Regression Model	1. Accounts for grouping of students within teachers 2. Estimates teacher effectiveness as the residualized gain in student's current score after controlling for student's prior performance and demographic characteristics 3. Does not require vertically scaled scores across grades	1. Highly complex and hard to explain to policymakers and stakeholders
Layered Mixed Effects Model	1. Apportions credit for student score gains to individual teachers 2. Does not require vertically scaled scores across grades	1. Highly complex and hard to explain to policymakers and stakeholders 2. Has stringent data requirements

Source: Wei, Hembry, Murphy & McBride, 2012

APPENDIX D: Value-added Model Methodology

VAM estimates were calculated using a dataset with all teachers with at least 10 tested students for five school years. Student test scores were normalized within year and test, and calculations were made using the resulting Z-scores. The estimates were the Best Unbiased Linear Predictors of score gains after the mixed-effects model was run. Two types of models were run: one type included the teaching context (student demographic model), and one type did not (test score only model). Models with one and two years of prior student scores were run. Students who did not have either prior test score or demographic information were excluded from the models where those covariates were required.

Teacher and school-level contextual covariates were the averages of their individual level variables.

The reduced form equations associated with the models are as follows:

Two year student demographic model: $Z_{i,t} = Z_{i,t-1}\alpha_g + Z_{i,t-2}\delta_g + \boldsymbol{X_{i,t}\beta} + \boldsymbol{\bar{X}_{c,t}\gamma_c} + \boldsymbol{\bar{X}_{s,t}\gamma_s} + (\mu_{t,t} + \varepsilon_{i,t})$

Two year test score only model: $Z_{i,t} = Z_{i,t-1}\alpha_g + Z_{i,t-2}\delta_g + (\mu_{t,t} + \varepsilon_{i,t})$

One year student demographic model: $Z_{i,t} = Z_{i,t-1}\alpha_g + \boldsymbol{X_{i,t}\beta} + \boldsymbol{\bar{X}_{c,t}\gamma_c} + \boldsymbol{\bar{X}_{s,t}\gamma_s} + (\mu_{t,t} + \varepsilon_{i,t})$

One year test score only model: $Z_{i,t} = Z_{i,t-1}\alpha_g + (\mu_{t,t} + \varepsilon_{i,t})$

Subscripts in the models are as follows: i is the individual student, g is the grade, t is the teacher. When preceded by a comma, t is the time period, where t -1 is the prior time period, and t -2 is the time period two periods prior.

Covariates included in the contextual model at the individual level were: FRL status, gender, ethnicity, sped level/gifted, ELL status, FAY status, old/young/repeating grade.

Covariates included in the contextual model at the school and teacher level were: FRPL status, gender, ethnicity, sped level/gifted, ELL status, FAY status, old/young/repeating grade, average and standard deviation of prior math and reading scores.

The teacher effect estimate is the Best Linear Unbiased Predictor of $\mu_{t,t}$. It is calculated using a random effects specification. $\boldsymbol{X_{i,t}}$,$\boldsymbol{\bar{X}c, t}$, **and** $\boldsymbol{\bar{X}_{s,t}}$ are the student, teacher, and school-level vectors of covariate averages, respectively.