

NEW MEXICO
LEGISLATIVE
FINANCE
COMMITTEE

Program
Evaluation
Unit

Program Evaluation: Longitudinal Student
Performance Analysis-Impact of School,
Teacher and Program Interventions

November 17, 2017

Report #17-04

LEGISLATIVE FINANCE COMMITTEE

Representative Patricia A. Lundstrom, Chairwoman
Senator John Arthur Smith, Vice-Chairman
Representative Paul C. Bandy
Senator William F. Burt
Senator Pete Campos
Senator Carlos. R Cisneros
Representative Randal S. Crowder
Representative George Dodge, Jr.
Representative Jimmie C. Hall
Representative Larry A. Larrañaga
Senator Carroll H. Leavell
Senator Howie C. Morales
Senator George K. Munoz
Senator Steven P. Neville
Representative Nick L. Salazar
Representative Jim R. Trujillo

DIRECTOR

David Abbey

DEPUTY DIRECTOR FOR PROGRAM EVALUATION

Charles Sallee

PROGRAM EVALUATION TEAM

Amir Chapel
Jon R. Courtney, Ph.D.
Sarah M. Dinces, Ph.D.
Nathan Eckberg, Esq.
Jenny Felmley, Ph.D.
Michaela Fischer
Brenda Fresquez, CICA
Maria D. Griego
Brian Hoffmeister
Clayton Lobaugh
Travis McIntyre, Ph.D.
Alison Nichols

Representative Patricia A. Lundstrom
Chairwoman

Representative Paul C. Bandy
Representative Randal S. Crowder
Representative George Dodge, Jr.
Representative Jimmie C. Hall
Representative Larry A. Larrañaga
Representative Nick L. Salazar
Representative Jim R. Trujillo

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



Senator John Arthur Smith
Vice-Chairman

Senator William F. Burt
Senator Pete Campos
Senator Carlos R. Cisneros
Senator Carroll H. Leavell
Senator Howie C. Morales
Senator George K. Munoz
Senator Steven P. Neville

November 17, 2017

Christopher Ruszkowski, Secretary Designate
Public Education Department
Jerry Apodaca Education Building
300 Don Gaspar
Santa Fe, New Mexico 87505

Dear Secretary Ruszkowski:

On behalf of the Legislative Finance Committee, I am pleased to transmit the evaluation, *Longitudinal Student Performance Analysis-Impact of School, Teacher and Program Interventions*. The evaluation reviewed long term student outcomes including the impact of poverty, language, and mobility on student performance.

The report will be presented to the Legislative Finance Committee on November 17, 2017. An exit conference to discuss the contents of the report was conducted with the Public Education Department on November 13, 2017. The Committee would like a plan to address the recommendations within this report within 30 days from the date of the hearing.

I believe this report addresses issues the Committee asked us to review and hope your department and New Mexico's school districts will benefit from our efforts. We very much appreciate the cooperation and assistance we received from you and your staff.

Sincerely,

A handwritten signature in blue ink that reads "David Abbey".

David Abbey, Director

Cc: LFC Members

Table of Contents



Executive Summary	1
Key Findings and Recommendations	2
Background	4
Findings and Recommendations	8
In general, New Mexico schools produce a year of growth for each grade.....	8
Student mobility may undermine student performance over time and undercut accountability systems.....	13
A more coordinated use of state funded interventions would help close achievement gaps.....	17
Teacher effectiveness and school grades vary by schools' rates of at-risk students	22
Agency Response	28
Appendices	29
Appendix A: Evaluation Scope and Methodology	29
Appendix B: Detailed Research Methodology	30
Appendix C: Effects of Student Proficiency and Growth Due to Special Education Status.....	32
Appendix D: School Average Percent of K-3 Plus Enrollment by District, Summer 2017	34
Appendix E: Eighth Grade SY13 SBA and SY16 PARCC Reading Score Distributions.....	35
Appendix F: Selected Bibliography.....	36





Student achievement improves over time but achievement gaps remain

Successful school systems start with high performance expectations for *all* students. Teachers and schools can have a positive impact on student achievement, particularly when implementing research-based practices such as data driven decision making. Non-school factors, such as student poverty and language status, can influence achievement as well, which requires even stronger implementation of best practices by teachers and schools.

New Mexico consistently ranks low nationally on student performance. These metrics assess student performance as a snapshot—one group of 3rd graders' performance in one year compared to another groups' performance the next year. Most students perform below grade level, and are low-income or are learning English. Previous Legislative Finance Committee evaluations found most students enter the school system behind on the first day of kindergarten. Public schools are designed to produce one year's worth of academic growth each year. This report seeks to understand student performance over time and whether schools produce academic growth as designed. The report generally examines two groups of students, following their progress over multiple years.

In general, students gained slightly more than a year's worth of academic growth for each grade—exactly what the system is designed to do. Despite the positive academic gains, New Mexico's large achievement gaps among students remain in place over time. To boost annual achievement levels, New Mexico's schools must produce significantly more academic growth to help students “catch up.” Some school districts demonstrated strong long-term academic growth for students, boosting proficiency levels over ten percentage points, while students in other districts fell behind.

High levels of student mobility negatively impacts achievement and most students never experience their school's full academic intervention. Only half of kindergarteners stay at the same school through 3rd grade, and less than a quarter through 5th grade. As mobility increases, both school and student performance decreases. However, some interventions may boost student performance and offer the opportunity to close achievement gaps. Students participating in New Mexico's prekindergarten (PreK) programs show lasting academic gains through 8th grade, and combining PreK and K-3 Plus appears to close achievement gaps.

The state needs changes in three key areas— access, administration, and accountability— to improve student achievement over the long-term. More students need *access* to PreK and K-3 Plus, schools implementing best practices, and highly effective teachers. The school system needs to ensure consistent, proper *administration* of best practices and evidence-based interventions. Finally, the state needs to continue to refine its *accountability* systems using many types of data to ensure students have access to properly administered programs so all students can achieve at high levels.



KEY FINDINGS AND RECOMMENDATIONS

In general, New Mexico schools produce a year of growth for each grade

In general, for one cohort from third to eighth grade, on average students in New Mexico gained approximately 3.5 additional days of learning, or a standardized difference of .006 on test scores, over the five years than expected. However, districts vary widely, with some districts losing a large number of days of learning, while others gain additional days of learning.

Students who qualify for free or reduced lunch, a measure of low-income, for all four years from kindergarten through third grade have a significantly lower rate of reading and math proficiency compared to those students who were enrolled in the program for fewer years, or never classified as low-income. Most students qualified for free or reduced lunch for all four years.

Low-income student growth is lower than that of non-low-income students. However, low-income students need to have higher academic growth rates than their peers in order to close achievement gaps.

Student mobility may undermine student performance over time and undercut accountability systems

Students in the third, sixth, and tenth grade in SY16 who changed schools more often had lower reading and math PARCC scores than their peers.

Schools with high rates of student mobility have higher proportions of low-income and English learner students. As schools implement academic interventions to help at-risk students, at-risk students may not be able to benefit fully from these interventions due to frequent school changes and high mobility rates.

PED should consider publishing mobility rates so schools, school districts, and the Legislature can identify schools or school districts with high mobility rates that are successfully improving student outcomes as well as those who are not.

State policy allows for additional student mobility, potentially affecting student performance.

A more coordinated use of state funded interventions would help close achievement gaps

In FY18, the state appropriated \$88.2 million from the general fund for “below-the-line” programs. However, the state does not prioritize funding for schools or school districts implementing a number of these below-the-line programs with a coordinated approach.

Prekindergarten (PreK), has positive long-term impacts on reading and math test scores through eighth grade and may mitigate some of the negative academic effects that occur when students change schools.

The state and school districts have not aligned PreK and K-3 Plus as only four percent of all students participated in both programs. When they do, good things happen.

Currently, the majority of students at K-3 Plus schools do not participate in the program. Of almost 13 thousand students who took the third grade PARCC test in K-3 Plus schools, only about 35 percent of students were ever enrolled in K-3 Plus. If school districts or schools implemented K-3 Plus schoolwide,

it may be easier to place students with their same teacher, likely improving the effect of the program.

State law targets K-3 Plus funding to the most challenged schools, however it does not prioritize funding to schools with other support interventions.

Schools with a higher percentage of teachers rated as effective or above and schools with better A-F school grades have relatively higher percentages of students achieving academic proficiency than other schools.

Schools with an A grade and schools with a higher percentage of teachers rated as effective or above have relatively fewer at-risk students than other schools. However, regardless of high proportions of at-risk students, some schools can still achieve high levels of academic proficiency and A or B school grades.

Teacher effectiveness and school grades vary by schools' rates of at-risk students

Key Recommendations

The Legislature should consider:

Continued expansion of prekindergarten programs, including extended day services.

The Public Education Department should:

Continue to identify, implement, and monitor specific interventions that may be helpful in improving low-income and English learner student outcomes.

The Public Education Department and school districts should:

Examine mobility patterns and their impact on school and student performance.

Align and coordinate below-the-line funding for school districts and schools.

Prioritize funding to schools willing to implement K-3 Plus schoolwide and who maintain fidelity to the program by keeping students with their same teacher and who end K-3 Plus programs within two weeks of the start of the school year.

Continue to identify and implement interventions to attract and retain highly effective teachers in high risk schools.

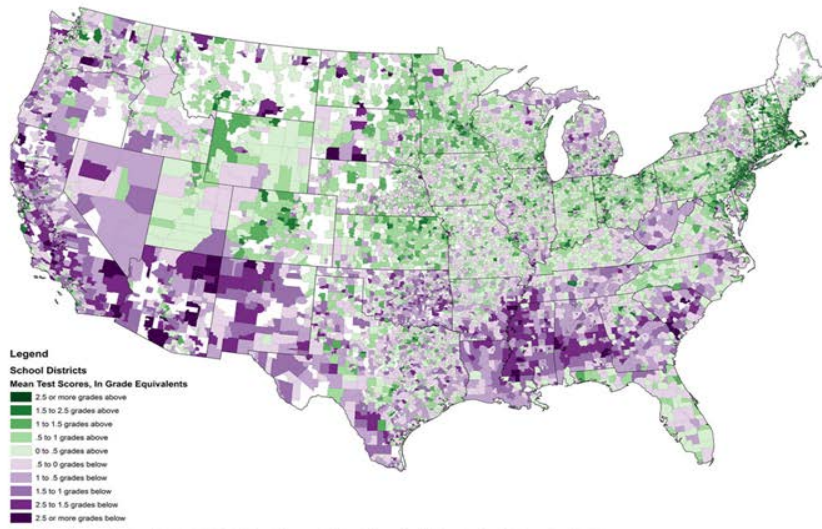
Overview

System Performance

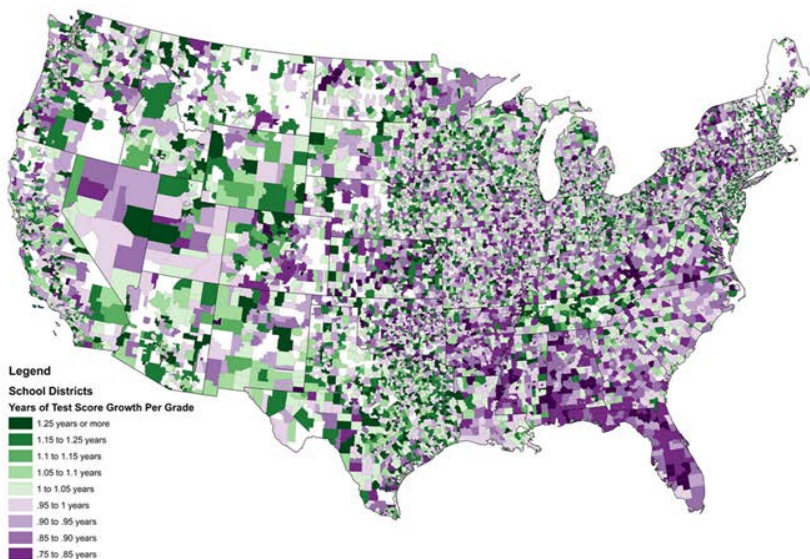
Historically, New Mexico student outcomes lag below the national average, with lower performance for low-income students. As of 2017, 71 percent of New Mexico students graduate high school, and 25 percent of third graders can read at grade level. Risk factors, such as low-income, being an English learner (EL) student, or changing schools may negatively affect student achievement, leading to these outcomes. These student outcomes are exaggerated when examining the achievement gap, as the third grade proficiency rate drops by five percent for low-income students. Furthermore, in a national study of 2015 high school graduation rates, New Mexico had the lowest graduation rate at 69 percent, with only 64 percent of low-income students graduating on time.

Figure 1. National Student Average Test Scores and Growth, Grades 3-8, 2009-2013
(Green = Positive, Purple = Negative)

National Student Test Scores, 2009-2013



National Average Student Growth, 2009-2013

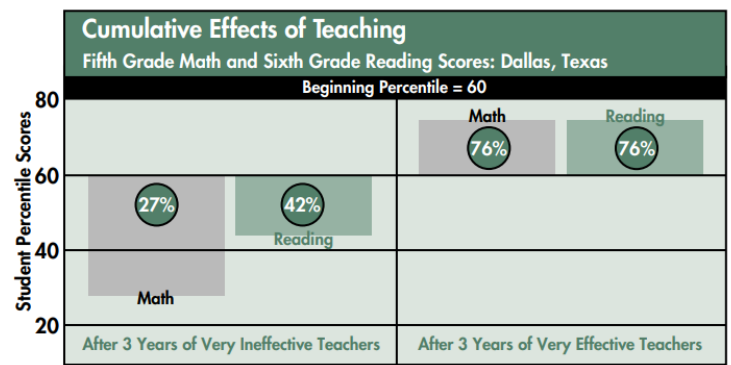


However, New Mexico students' growth, or positive test score changes from one year to the next, is higher than other states (Figure 1). Determining how to help New Mexico students catch up to those in other states is vital for New Mexico students' academic success.

New Mexico currently uses a number of different performance measurement systems to examine student, teacher, and school achievement. The state annually tracks the percentage of students passing exams for each grade disaggregated by income, language status, gender, and ethnicity/race. In addition, PED uses value-added models, which include student growth as a part of the model, to assess teacher effectiveness and school grades. These models incorporate information at a

student, teacher, and school level to determine the specific effects of the teacher or school. When examining teacher quality, research from Washington state found for all measures of teacher quality, high quality teachers are more likely to be in schools with higher student income and historical student academic performance, these effects were shown for both schools and school districts. LFC evaluations have shown lower performing schools tend to have a higher number of beginning teachers and teachers with lower licensing exam scores. Having teachers with less experience may influence student test scores as well as student growth. Academic research by Hanushek finds a high performing teacher, one at the 84th percentile of all teachers, when compared with just an average teacher, produces students who should move up more than seven percentile rankings within a year. If at-risk students have highly effective teachers, then they may be able to catch up to their peers who have average teachers. However, it is unknown if at-risk students are receiving effective teachers at similar rates as lower risk students, and it is unknown whether New Mexico can expect the same effects shown in previous research.

Figure 2. Effects of Three Years of Highly Effective versus Highly Ineffective Teaching



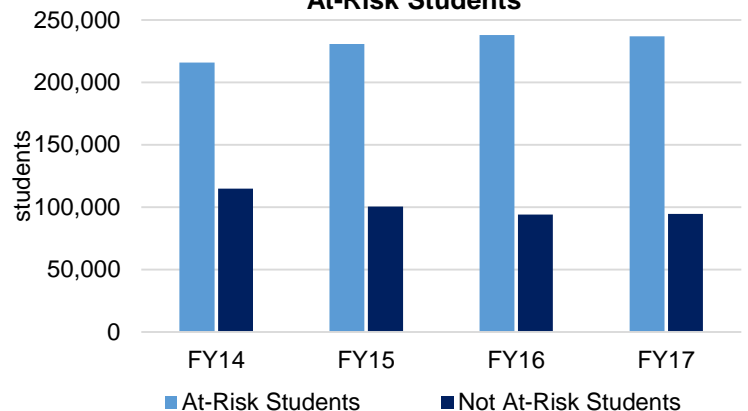
Source: Southwest Education Development Laboratory, 2001; Jordan et al., 1997

Long established academic studies find school quality is a large factor in student academic achievement. If school quality is not examined, student and teacher achievement effects may be overestimated. Therefore, it is important to determine what distinguishes a high and low quality school. One measure readily available in New Mexico is the school letter grade the Public Education Department (PED) assigns to schools. School grades in New Mexico attempt to summarize the overall performance of schools, including measuring student growth and snapshot test performance. The complexity of the system, constant changes to measures, and strong correlations of school performance and student poverty levels call into question the grade’s validity. Nonetheless, no other superior system exists and some schools demonstrating best practices “beat the odds” and achieve high performance.

Targeting Finances to Address the Achievement Gap

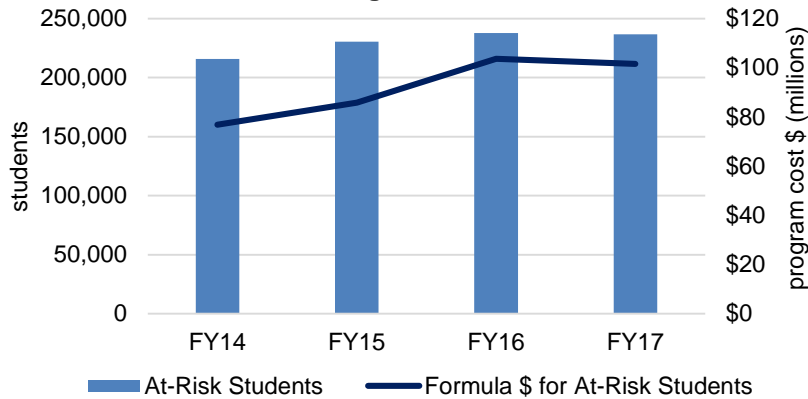
In an effort to close the achievement gap for at-risk students, New Mexico provides additional funding to school districts and charter schools. In FY17, the New Mexico public education funding formula allocated \$2.5 billion for school district and charter school operations based on student educational needs and other factors. Four percent of this funding, or \$102 million, was allocated to school districts and charter schools for serving at-risk students as defined by students’ low-income, English learner, or mobile status. The state has long recognized these students tend to lag behind peers and require additional resources to help them catch up. Around two-thirds of students statewide are identified each year as being at-risk in the funding formula (Chart 1).

Chart 1. Public Education At-Risk and Not At-Risk Students



Source: LFC Files

Chart 2. Public Education At-Risk Students and Formula Funding for At-Risk Students



Source: LFC Files

The Legislature passed legislation in 2014 to increase formula funding for at-risk students. At-risk funding has increased 32 percent, or \$25 million, from FY14 through FY17 (Chart 2). Current state statute requires, but PED does not use, the budget process to account for how school district resources, such as at-risk index funding, may affect student performance. Given the Legislature’s financial commitment to supporting at-risk student success, it is important to evaluate student risk factors, interventions, and outcomes over multiple years to determine how best to improve at-risk student outcomes.

In addition to funding school districts based on their number of at-risk students, PED has a variety of initiatives, mainly implemented during early childhood, focused on improving at-risk student outcomes.

If students enroll in targeted interventions at an early age, the achievement gap between at-risk and non-at-risk students may shrink or disappear. Prekindergarten (PreK) and K-3 Plus programs, which were funded in FY17 at \$52 million and \$24 million respectively, have been expanded since 2007 and both show positive effects on student test scores. In a meta-analysis focused on PreK nationally, most studies examined PreK effects between grades kindergarten through four, and found a decline in significant effects across those years. Of the studies that examined PreK effects on kindergarten performance, 77 percent found significant positive effects, while 35 percent of the studies examining PreK effects on fourth grade performance reported significant positive effects. Furthermore, most states have not examined the effects of PreK after fifth grade; however, those that did (New York and Maryland) reported positive effects up to 10th grade. In the 2017 LFC Early Childhood Accountability Report, New Mexico PreK was shown to affect student proficiency rates through fifth grade. However, it is unknown whether there are effects for PreK beyond elementary school. The landmark Perry preschool study, which followed children from PreK to age 40, found children who were in a high quality preschool had higher graduation rates, were less likely to be arrested, and had higher earnings than those who were not enrolled in a PreK program.

Additional academic research finds students that have a slight standard deviation increase in first grade test scores, have almost a \$5,000 increase in lifetime earnings. New Mexico has positive effects of PreK on children in third and fifth grades, but longer term effects have not been examined. Given the research highlighting how low-income children are more positively affected by PreK programs, and the fact that New Mexico has a high percentage of children in low-income households, New Mexico may benefit more than other states from programs such as PreK, and these effects may be longer lasting. K-3 Plus, a program designed to add 100 days of learning over four years for students in low-income or low performing schools, has been scientifically studied showing some effects through third grade. However, additional examination of K-3 Plus program outcome effects as well as current program implementation are needed.

Purpose of Report

New Mexico has persistently low test scores when examining student performance on annual exams. While important, these indicators do not show whether students demonstrate adequate academic growth over time. This study seeks to assess groups of students over a long period of time to learn more about their academic progress and factors that may impede or improve success.

Limitations

This program evaluation does not examine how academic interventions or student risk factors influence future high school graduation. LFC staff requested but did not receive student-level graduation data. This program evaluation also does not assess the fidelity of K-3 Plus program implementation since LFC did not receive data linking teachers to students.

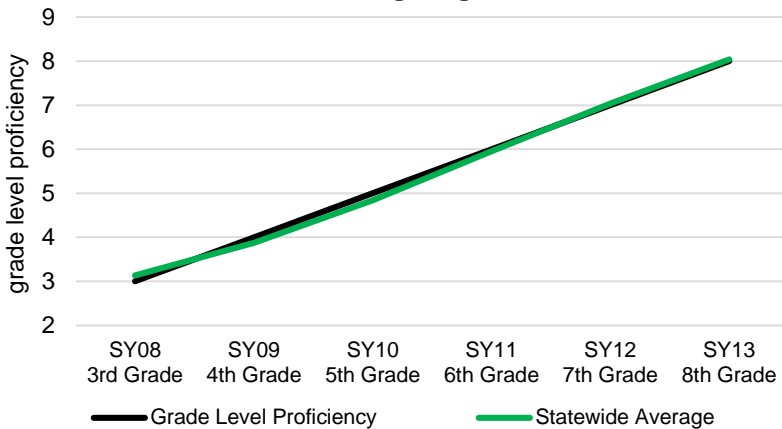
In general, New Mexico schools produce a year of growth for each grade

This evaluation generally studied two main groups of students over a number of years. Analyses were focused on risk factors within both cohorts, such as family income, English learner status, and mobility. The first group of 20,210 students was studied from third through eighth grade from SY08-SY13. The

second group of 21,948 students were studied from kindergarten through third grade from SY13-SY16. In both groups, the majority of students were from low-income families; 57 percent for the cohort of third through eighth graders and 68 percent for the cohort of kindergarten through third graders. About 22 percent of both cohorts participated in English learner programs.

For both cohorts, student performance was examined using test scores from the Standards-Based Assessment (SBA) or the Partnership for Assessment of Readiness for College and Careers (PARCC). However, for the first cohort, student growth, or how student test scores changed over time, was also analyzed. Proficiency rates do not by themselves adequately convey student learning growth for a specified timeframe. To examine student growth, difference scores between third and eighth grade test scores were calculated after the data was standardized. These scores show the average growth of students from third to eighth grades relative to their peers. A score of zero signifies students learned the amount they were expected to learn (one year of academic growth), while a score above zero shows students learned more than would be expected over the five years data was collected. (For additional methodological information, see Appendix B).

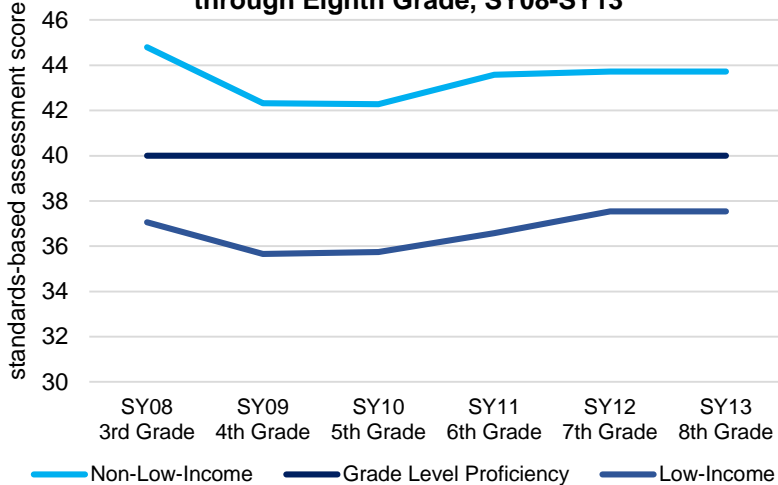
Chart 3. Grade Level Proficiency in SBA Reading Exam from Third through Eighth Grade, SY08-SY13



Source: LFC analysis of PED data;

Note: Average SBA scores for each year were divided by 40 (the SBA proficiency threshold score) and then multiplied by the grade level. An SBA score of 40 in third grade would be a value of three.

Chart 4. Average Reading SBA Scores from Third through Eighth Grade, SY08-SY13



Source: LFC analysis of PED data

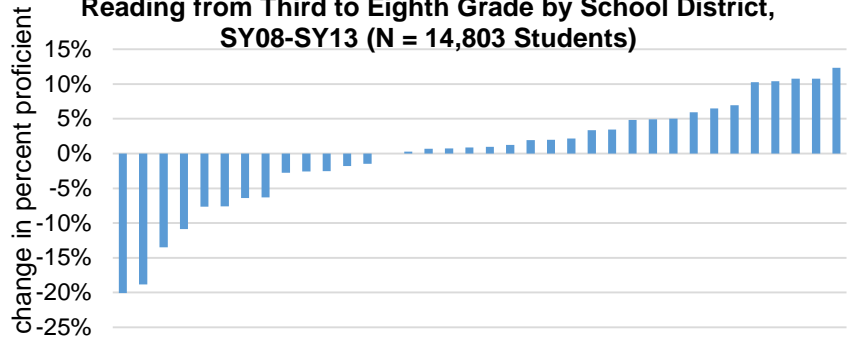
In general, students show consistent academic growth over time but not enough to overcome gaps in performance.

Student performance growth on standardized tests from third to eighth grade stay relatively constant, however some years' scores are lower than others. For

New Mexico students from third through eighth grade, student standardized scores tend to dip in transition years between elementary and middle school (Chart 3), potentially indicating increased support may be needed during those years or there may be a misalignment of content standards and tests between elementary and middle schools. Low-income students continue to score below non-low-income students from third through eighth grades (Chart 4). Test scores are highest statewide for non-low-income students in third grade, with only low-income students scoring slightly higher in eighth grade. These data highlight low-income students need to learn at higher academic growth rates than their peers to close achievement gaps.

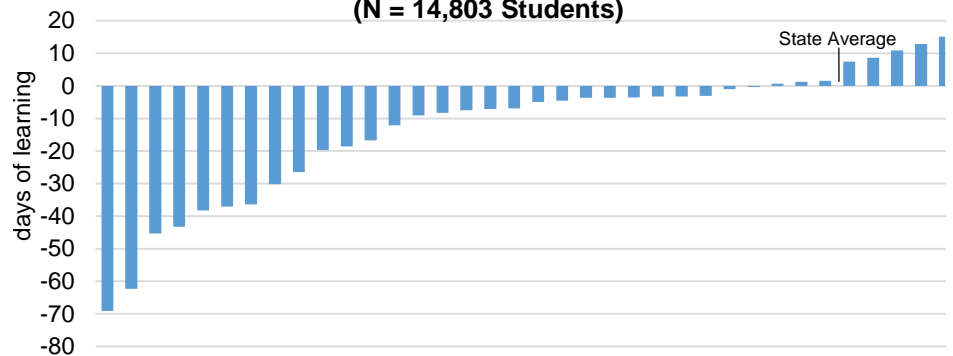
Some school districts are closing significant learning gaps and boosting overall performance, but district performance varies significantly. Specifically, some school districts increased their percentage of students proficient in reading by ten percentage points between third and eighth grade (Chart 5). This difference can be translated into days of learning (Chart 6). In general, for one cohort from third to eighth grade, on average students in New Mexico gained approximately 3.5 additional days of learning, or a standardized difference of .006, over the five years than expected. However, school districts vary widely, with some school districts losing a large number of days of learning, while others gain additional days of learning.

Chart 5. Change in Percent of Students Proficient in SBA Reading from Third to Eighth Grade by School District, SY08-SY13 (N = 14,803 Students)



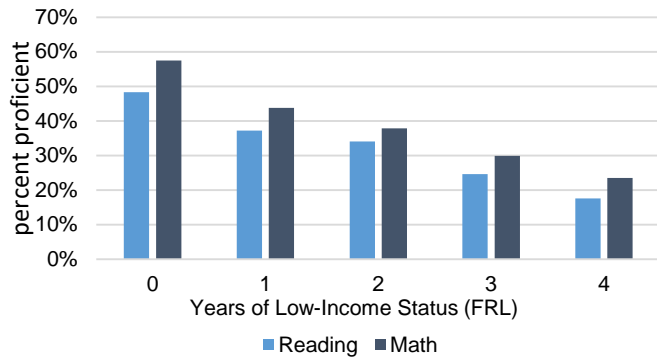
Source: LFC analysis of PED data; Note: 36 Districts shown. Districts with fewer than 30 students with both SY08 and SY13 SBA data omitted from chart.

Chart 6. Average Annual Student Growth, in Days of Learning, from Third through Eighth Grade by School District, SY08-SY13 (N = 14,803 Students)



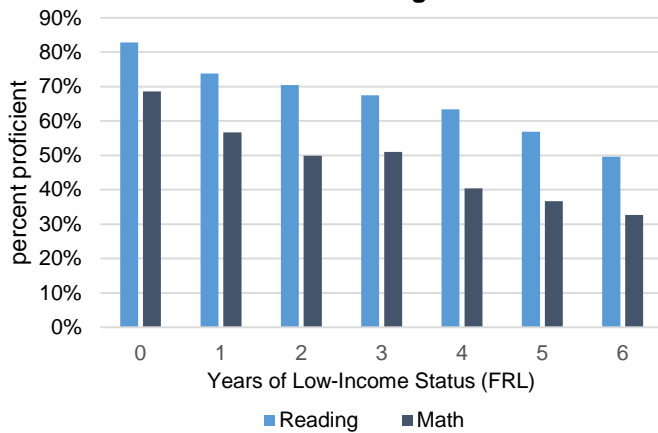
Source: LFC analysis of PED data; Note: Average growth score is the average difference of standardized SY13 SBA reading and math scores minus SY08 standardized reading and math scores. Thirty-six districts are shown on chart. School districts with fewer than 30 students with both SY08 and SY13 SBA data were omitted.

Chart 7. Third Grade PARCC 2016 Proficiency by Cumulative Low-Income Status from Kindergarten to Third Grade



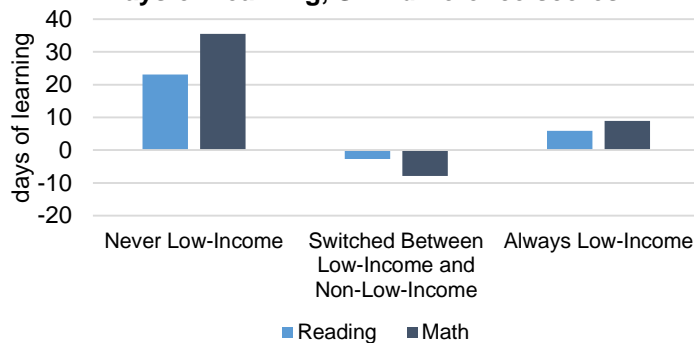
Source: LFC analysis of PED data

Chart 8. Eighth Grade SBA 2013 Proficiency by Cumulative Low-Income Status from Third to Eighth Grade



Source: LFC analysis of PED data

Chart 9. Student Growth from Third to Eighth Grade by Income Status, Shown as Days of Learning, SBA difference scores



Source: LFC analysis of PED data; Note: growth score is the average difference of SY13 standardized reading and math SBA scores minus SY08 standardized reading and math.

Most students consistently participate in free or reduced school lunch and show a cumulative negative effect on performance and growth. Students who qualify for the federal free and reduced lunch (FRL) program, a measure of low-income status, for all four years from kindergarten through third grade have a significantly lower rate of reading and math proficiency compared to those students who were enrolled in the program for fewer years, or never classified as low-income. These effects are consistent for students from third to eighth grades as well, with students less likely to be proficient in reading or math as the number of years of low-income status increase (Chart 7 and Chart 8).

Importantly, most students in New Mexico are low-income, with 67 percent in the group with the highest cumulative number of years enrolled in free and reduced lunch for the cohort from kindergarten to third grade (4 years) and 56 percent of students for the cohort from third to eighth grade (6 years).

Low-income students do not grow academically as quickly as non-low-income students from third to eighth grade. Low-income students from third to eighth grade had lower growth scores than students who were never low-income. Specifically, “always” low-income students had an average math growth score of almost 10 additional days of learning, while non-low-income students had a growth score of approximately 35 additional days of learning (Chart 9). Those students who went into or out of low-income status between third and eighth grades had the lowest student growth scores.

The timing of when students are no longer classified as English learners (EL) appears to affect future student proficiency in both reading and math. Students who are no longer classified as EL in or before third grade have higher student test scores. Not only closing the achievement gap, but also scoring better than students who were never EL students. However, students who remain classified as EL in third grade, and those students who were no longer classified as EL after third grade show negative effects of EL status (Chart 10 and Chart 11). For the kindergarten through third grade cohort, 4,169 students were classified as EL students. About 85 percent of those who were classified as EL students continued as EL through third grade and of those 85 percent, less than 10 percent read at grade level. About 15 percent of EL participants exited and demonstrated proficiency levels at or above students who were never EL. Students who are classified as EL after third grade have negative test score effects. Specifically, of EL students from third through eighth grade, only 20 percent were proficient in reading and 13 percent were proficient in math.

Students who are or were EL have higher growth scores than students who were never EL. If a student was EL sometime between third and eighth grades, they gained an additional 28 instructional days compared to the average student (Chart 12). For math, the difference is more pronounced with students currently EL status learning at a rate equivalent to an additional 64 instructional days over the five-year span. Students who are EL have large improvements in test scores as they are receiving the appropriate level of instruction. It is expected students who exited EL status would grow less than those who are currently EL, as they have already made the gains needed, however they continue to grow at a faster rate than those who have never been in classified as EL. These data highlight how low-income and EL students need to learn and grow academically at higher rates than their peers in order to close achievement gaps.

Chart 10. Third Grade PARCC 2016 Proficiency by English Learner Status from Kindergarten through Third Grade

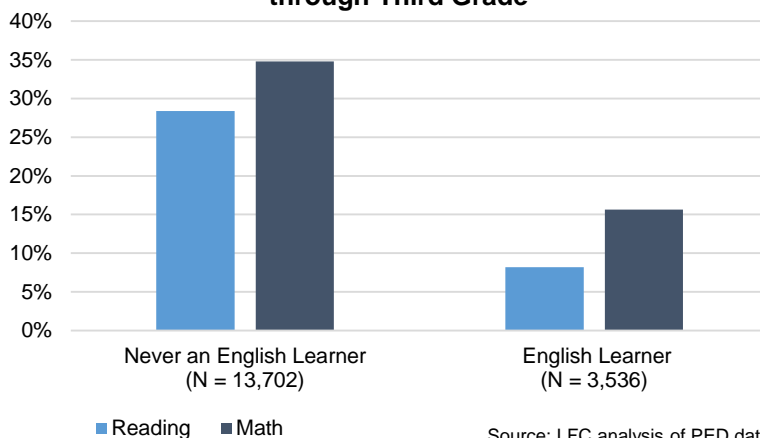


Chart 11. Eighth Grade SBA 2013 Proficiency by English Learner Status from Third through Eighth Grade

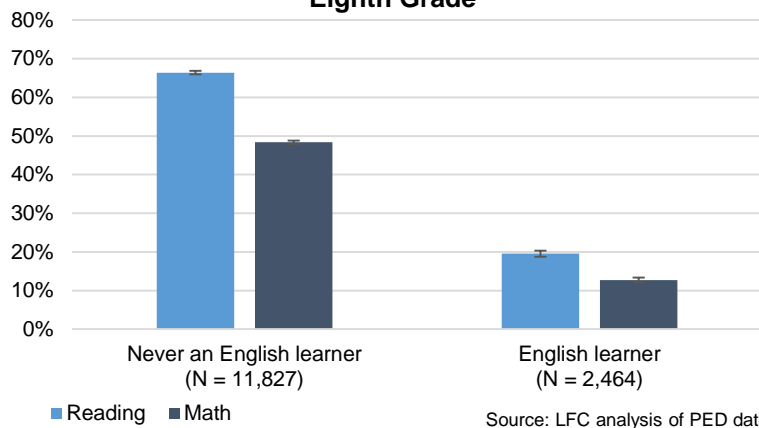
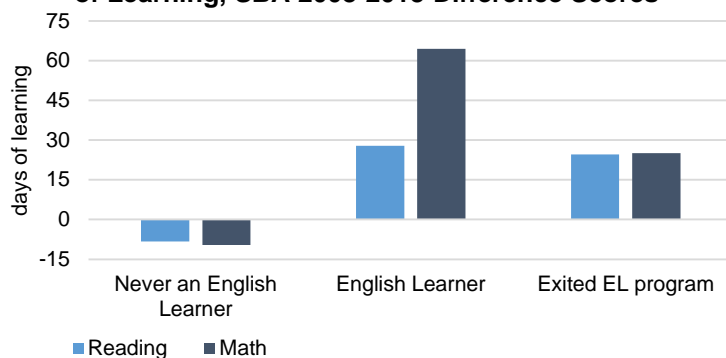


Chart 12. Student Growth from Third to Eighth Grade by English Learner Status, Shown as Days of Learning, SBA 2008-2013 Difference Scores



Source: LFC analysis of PED data. Note: growth score is the average difference of SY13 standardized reading and math SBA scores minus SY08 standardized reading and math SBA scores.

State accountability systems attempt to recognize short-term growth using complex calculations; however, incorporating long-term performance may prove useful.

Current LFC report cards use snapshot data to assess how students are performing, and most PED accountability measures available to the public also use data for one year. While PED's teacher evaluation and school grading systems include both student performance and growth, these accountability systems are hard to understand, and are not well received by the school community at large. Examining long-term student growth and performance may be another way to determine how students are performing while including enough information to get a comprehensive measure of student growth.

Recommendations

The Public Education Department should:

- Monitor statewide school district growth rates to determine which districts may need additional supports or attention.
- Continue to identify, implement, and monitor specific interventions that may be helpful in improving low-income and English learner student outcomes.

Student mobility may undermine student performance over time and undercut accountability systems

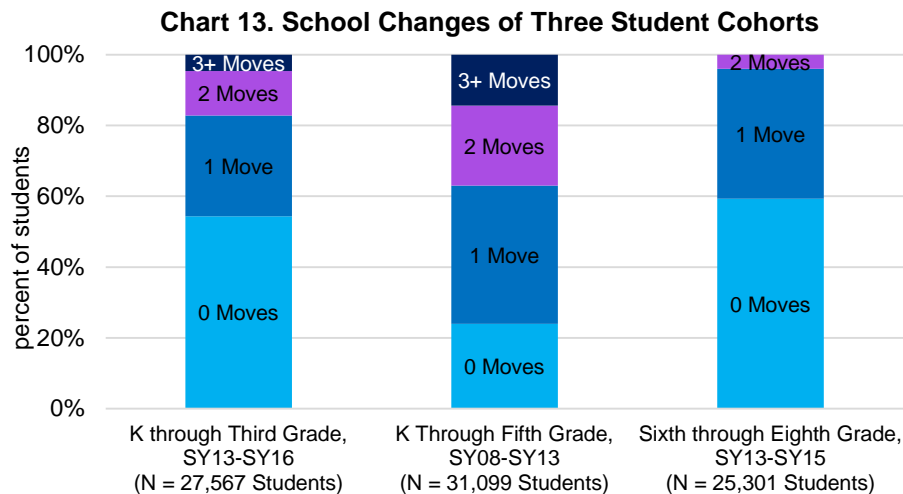
By traditional design, students attend three different schools before graduating high school, but many students change schools more often.

Students who move frequently are included as at-risk students in the funding formula. Students may move for a variety of reasons such as parental employment or family events. Many students change schools more frequently than would be expected. LFC staff calculated mobility rates as the percentage of students in a grade level cohort who changed schools between academic years. Forty-six percent of students who started kindergarten in SY13 changed schools at least once by third grade and 17 percent changed schools two or more times (Chart 13). Of the students who started kindergarten in SY08, only 24 percent of students remain at the same school for kindergarten through fifth grade. For middle schools, forty percent of SY13 sixth graders changed schools by eighth grade. These data indicate many students are changing schools more frequently than the scheduled transitions from elementary school to high school.

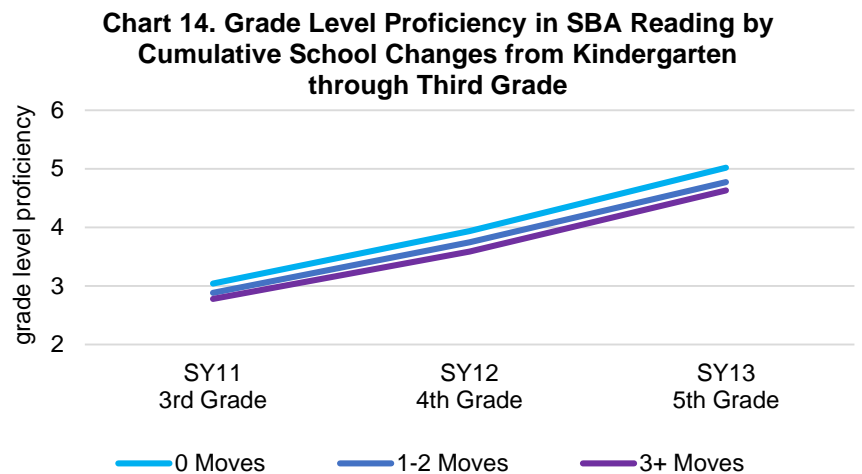
Students moving more often have lower test scores and are more likely to be from low-income families.

School changes from kindergarten through third grade have lasting effects on academic performance in third through fifth grade. Students who changed schools more often from kindergarten in SY08 through third grade in SY11 had lower average SBA test scores in third grade through fifth grade than students who moved less often (see Chart 14).

Students in the third, sixth, and tenth grade in SY16 who changed schools more often had lower reading and math PARCC scores than their peers. These negative effects of student mobility on test scores were still statistically significant after controlling for low-income and English proficiency. Students who changed schools more often on



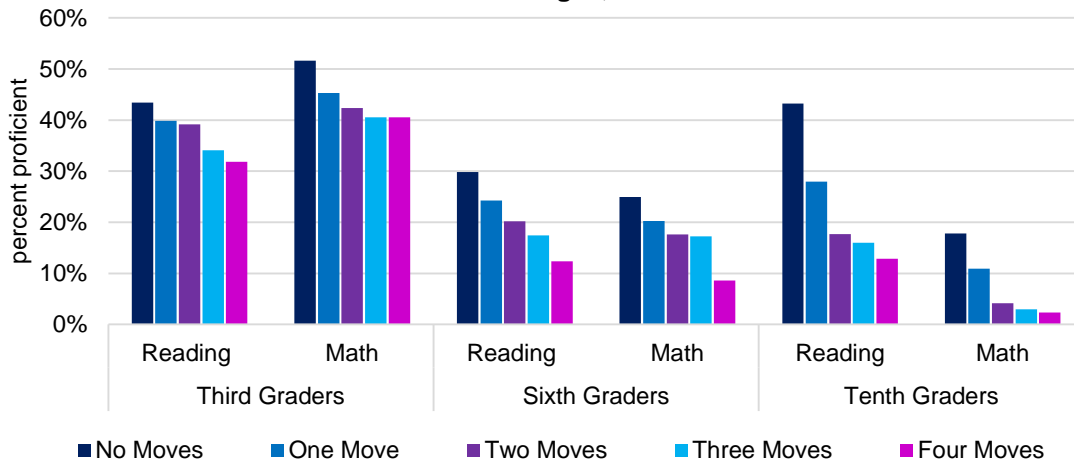
Source: LFC analysis of PED data



Source: LFC analysis of PED data;
 Note: LFC staff examined the number of times students moved from kindergarten through third grade for SY11 third graders. Average SBA scores for each year were divided by 40 (the SBA proficiency threshold score) and then multiplied by the grade level. An SBA score of 40 in third grade would be a value of three.

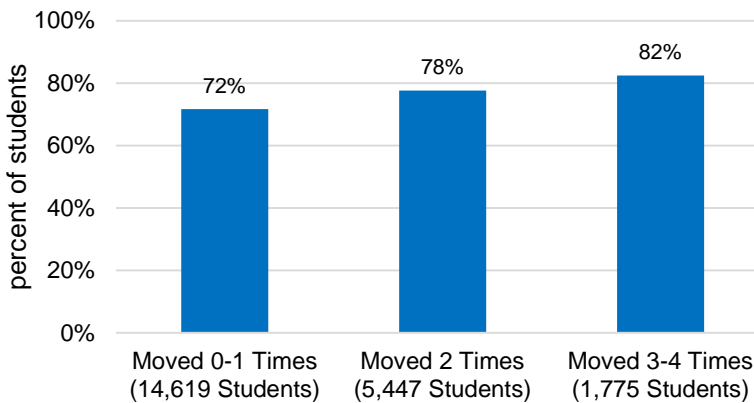
average had lower proficiency levels than their peers. Chart 15 shows the percent of students who were proficient on PARCC in SY16 by the number of times they moved since SY13.

Chart 15. Percent of Students Proficient on SY16 PARCC by Number of School Changes, SY13-SY16



Source: LFC analysis of PED data

Chart 16. Percent of SY16 Sixth Graders with Low-Income by School Changes from SY13 to SY16 (N = 21,841)



Source: LFC analysis of PED data

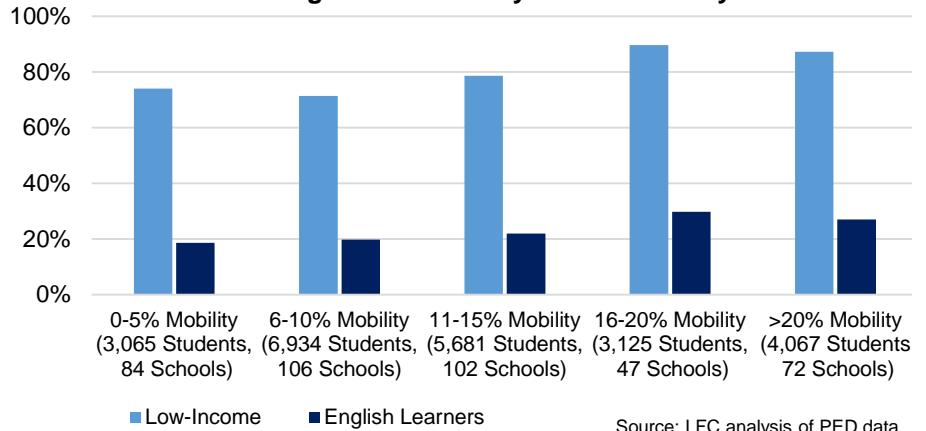
Higher percentages of students who changed schools frequently were from low-income families. Specifically, 7,222 sixth grade students moved schools more often than expected, and a disproportionate number of these students were from low-income families. Chart 16 shows the percentage of SY16 sixth grade students with low family income by the number of times students changed schools from SY13 through SY16. Although PED reports school proficiency percentages disaggregated by low family income, English learner status, and special education status, PED does not report school proficiency percentages disaggregated by student mobility. PED should consider reporting PARCC results disaggregated by a measure of student mobility, given the impact of student mobility on test scores.

School mobility rates vary widely and undermine academic interventions.

When looking at mobility at the school level, average mobility rates vary from 84 schools with less than 5 percent mobility to 72 schools with mobility rates above 20 percent. LFC staff calculated school mobility rates for grade levels as the proportion of students at a school who changed schools between academic years. Schools with high rates of student mobility have higher proportions of low-income and EL students (Chart 17 and Chart 18).

Increased mobility at a school can reduce the effectiveness of interventions, as potentially, a school is only consistently intervening with about a quarter of the students, if looking at a kindergarten through fifth grade elementary school. Given the relationship between low-income and student mobility, PED and school districts need to consider mobility when implementing academic interventions for both low-income and English learner students.

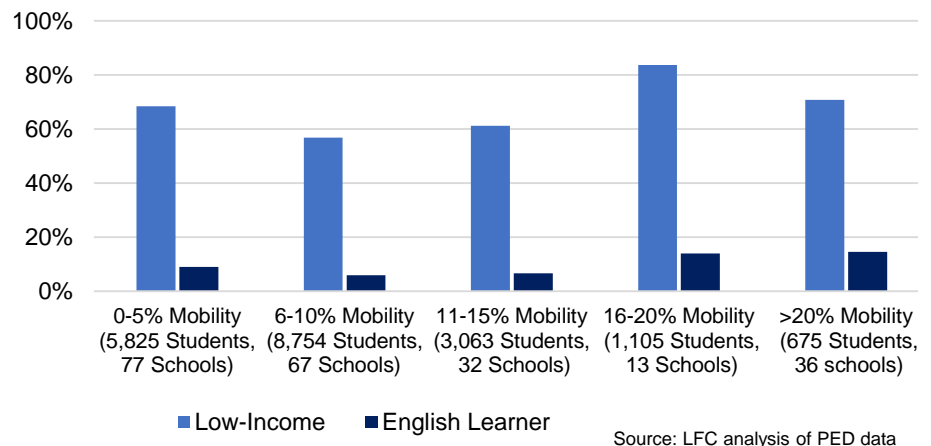
Chart 17. Percent of SY16 Third Graders with Low Income or who are English Learners by School Mobility Rates



School and teacher accountability systems may not sufficiently take into account the academic impact of student mobility.

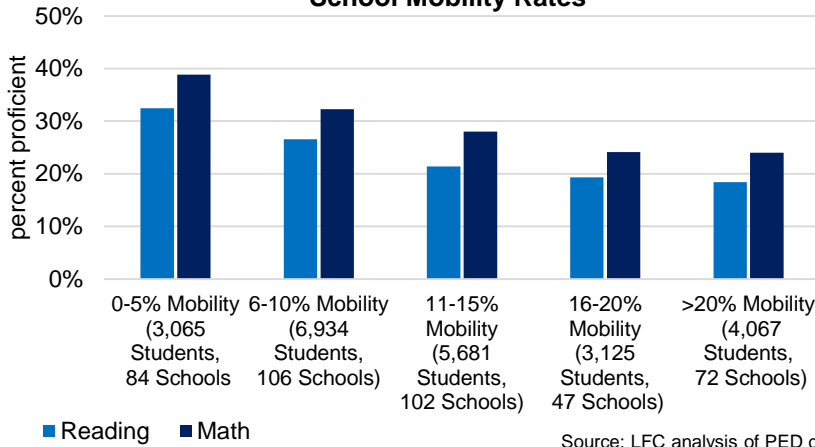
PED uses student mobility data to adjust expected scores in its school A-F grading and teacher rating value-added models. The school A-F grading system takes into account the proportion of students at a school enrolled for a full academic year, while the teacher rating system considers the proportion of an academic year students were enrolled in a teacher’s class. Specifically, accountability systems ensure annual snapshot performance data only captures kids at that school or classroom for the year measured, but does not mean the student was in the school the year before, or will be there the year after. However, these accountability systems may not account for the cumulative academic effects of multiple school changes over time.

Chart 18. Percent of SY15 Tenth Graders with Low-Income or who are English Learners by School Mobility Rate



State policy allows for additional student mobility, potentially affecting student performance.

Chart 19. SY16 Third Grade PARCC Proficiency by School Mobility Rates



Source: LFC analysis of PED data

Schools with high mobility rates had lower third grade PARCC proficiency levels than other schools. Chart 19 shows the percent of students who achieved a proficient score on the PARCC math and reading exams in SY16 by their school mobility rate.

State policy allows students to transfer from a school rated a D or F for two consecutive years. This study analyzed a handful of F schools’ enrollment patterns in Albuquerque. Most students move from the F school to another school rated F or D.

However, when they left for higher graded schools their performance improved relative to peers staying behind. This analysis did not control for selection bias or prior performance, but does demonstrate the need to closely examine policies encouraging changing schools.

The expansion of charter schools and other district magnet schools and special programs gives parents more choices to change their kid’s school. Students in charter schools move more than students in public schools. Absent purposeful education planning at both the state and local level, policies encouraging changing schools may undermine performance, making it necessary for a deeper understanding of the root causes of “failing” schools. PED already uses school district mobility rates in the public education funding formula; however, PED should consider publishing school mobility rates in accountability reports to allow examination of outcomes for this at-risk group of students.

Recommendations

The Public Education Department and school districts should:

- Examine mobility patterns and their impact on school and student performance.

The Public Education Department should:

- Publish student mobility rates at the district and school level.
- Report PARCC scores for school districts and schools disaggregated by a measure of student mobility so PED, school districts, and the Legislature can identify schools or school districts with high mobility rates that are successfully improving student outcomes as well as those who are not.

A more coordinated use of state funded interventions would help close achievement gaps

The state will spend \$88.2 million in FY18 on a variety of public school professional development and intervention programs, including Prekindergarten and K-3 Plus.

New Mexico should better align and coordinate its educational initiatives across schools and school districts. In addition to the funding provided to school districts and charter schools through the public education funding formula, the state provides earmarked appropriations, referred to as “below-the-line” funding, to PED for a variety of professional development and intervention programs. In FY18, the state appropriated \$88.2 million from the general fund for below-the-line programs (Table 1). New Mexico does not prioritize funding for schools or school districts that coordinate implementation of these programs. The 2016 National Conference of State Legislatures education report, *No Time to Lose*, found that top performing countries used aligned reforms to build their world-class education systems. Rather than adopting only one or two “silver bullet” policies, these countries reimagined and re-engineered their entire systems.

Table 1. FY18 General Fund "Below-the-Line" Education Appropriations

K-3 Plus Fund	\$23,700,000
PED Pre-Kindergarten	\$21,000,000
Early Reading Initiative	\$12,500,000
Breakfast for Elementary Students	\$1,600,000
After School & Summer Enrichment	\$350,000
NMTEACH Evaluation System	\$4,000,000
Science, Technology, Engineering, Math, and Health Teachers	\$1,900,000
Next Generation School Teacher and School Leader Preparation Programs	\$2,100,000
College Preparation, Career Readiness, and Dropout Prevention	\$2,200,000
Advanced Placement	\$825,000
Interventions and Supports for Students, Struggling Schools, Teachers, and Parents	\$15,000,000
GRADS - Teen Pregnancy Prevention	\$200,000
Teachers Pursuing Excellence	\$900,000
Regional Education Cooperatives	\$935,000
Stipends for Teachers in Hard-to-Staff Areas	\$1,000,000
Total	\$88,210,000

Source: LFC Files

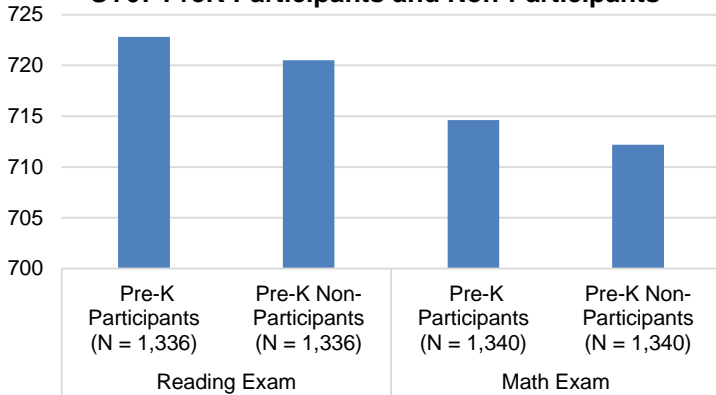
New Mexico prekindergarten shows lasting academic gains through eighth grade as well as other benefits.

New Mexico prekindergarten (PreK) is an early education half or extended day program for three and four year olds implemented by the Children Youth and Families (CYFD) and the Public Education Departments (PED). In FY16, about 10,000 children received PreK services. PreK was only available to four year olds until FY16 when policymakers expanded services for three year olds and authorized extended day services. Over 450 children participated in early PreK programs in FY16. In FY16, 20 percent of PED’s PreK students were served in full days programs.

Research indicates early childhood education has significant impacts on later academic achievement and life outcomes. A 2010 paper from the National Bureau of Economic Research found strong correlations between high kindergarten test scores, future college attendance, higher earnings, and home ownership. Therefore, programs such as PreK, which are designed to improve student test scores for kindergarten and beyond, have consistent positive effects. Furthermore, previous LFC reports found lasting effects for students

participating in PreK through the third grade and found PreK reduces special education identification rates and the need to hold students back.

Chart 20. Average FY16 PARCC Scores of SY07 PreK Participants and Non-Participants



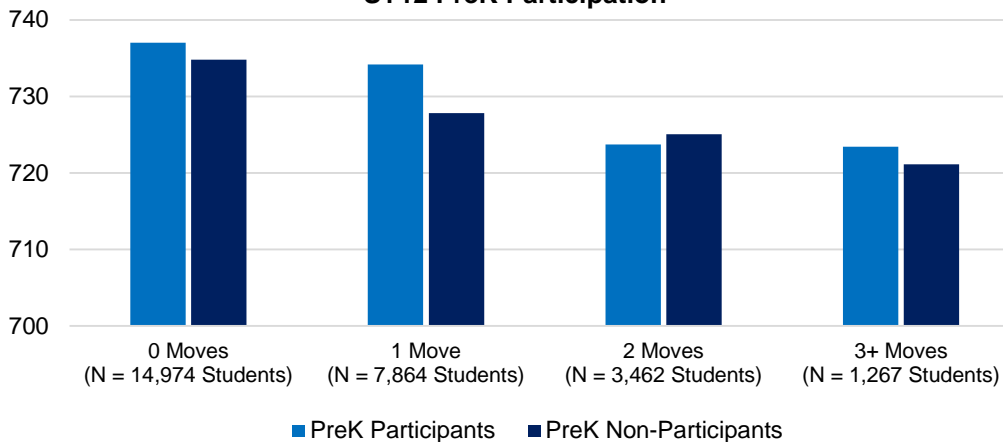
Source: LFC analysis of PED data

New Mexico PreK, has positive long-term impacts on reading and math test scores through eighth grade. LFC staff compared SY16 eighth grade PARCC scores of students who participated in PreK in SY07 with the PARCC scores of students who did not participate in PreK in SY07 by matching comparable groups of PreK participants and non-participants by a number of factors. PreK participation positively impacted eighth grade test scores in both reading and math at statistically significant levels (Chart 20). On the PARCC math exam, 20 percent of sampled PreK participants and 17 percent of the sampled non-participants were proficient.

PreK participation was also associated with increased school attendance, which likely contributes to PreK’s positive effects. These data indicate PreK has effects lasting at least nine years after the completion of the program.

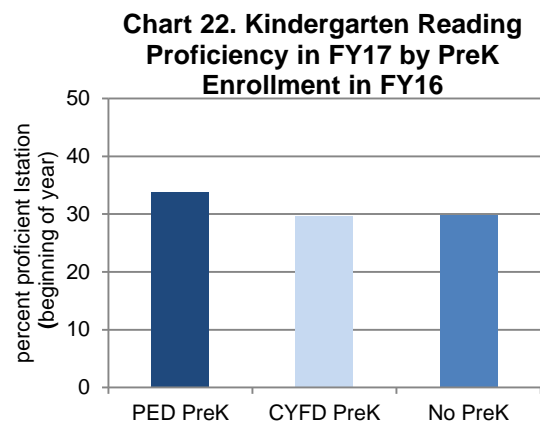
PreK participation in SY12 lessened the negative effect of mobility on third grade PARCC test scores in SY16. PreK participation moderated the negative effects of student mobility on math scores at a statistically significant level (Chart 21 below); however, PreK participation did not significantly counter the negative effects of student mobility on reading scores. These mixed results suggest PreK participation can counter the negative effects of student mobility for some student outcomes. Further analysis should be conducted to determine whether PreK is a protective factor for students who move or if there are other factors such as parental involvement that may be causing these effects.

Chart 21. Average SY16 Third Grade PARCC Math Scores by SY12 PreK Participation



Source: LFC analysis of PED data

PreK gains are higher in the PED system, but both PED and CYFD have high and low performing PreK programs. As discussed above, prekindergarten (PreK) consistently improves student test scores throughout elementary school, however the PED program improves test scores more than the CYFD program, with 34 percent of children in PED PreK proficient in reading at the beginning of kindergarten, while 30 percent of the other students were proficient (Chart 22). CYFD PreK did not significantly affect test scores. However, since PreK programs are located in low-income areas, CYFD PreK participants having test scores similar to the general student population is indicative of effectiveness. These current findings between PreK programs are consistent with previous LFC analyses.



Source: LFC analysis of PED and CYFD data

Identifying the effectiveness of providers is crucial to continuing quality improvement efforts of PreK programs. The providers who were low performers had students who performed worse than students who did not attend PreK. However, this negative effect may be due to factors that were not examined in this analysis. Most of the lowest performing PreK providers were administered by CYFD. PED had five of the top ten performing PreK providers, while CYFD had nine of the ten lowest performing PreK providers. However, some of the low performing CYFD PreK providers serve high proportions of at-risk student populations. The high variability of student test scores by PreK provider indicates the need for performance monitoring to ensure consistent positive effects of PreK.

Better alignment of K-3 Plus with other interventions and closing funding gaps may help improve student performance.

A small proportion of students participate in both PreK and K-3 Plus. When students participate in both, good things happen. For a cohort of students from kindergarten through third grade, only four percent participated in both PreK in SY12 and K-3 Plus. For those students who participate in both PreK and K-3 Plus, the achievement gap was closed in kindergarten. However, only 42 percent of schools with K-3 Plus programs also have PreK programs, and this number drops to 34 percent if only districts with multiple schools are considered (Table 2). These data indicate that most schools with PreK and K-3 Plus are not implementing the programs schoolwide. However, this estimate only includes PED PreK programs, and does not include CYFD PreK programs potentially available in the school zone, which could still lead to stacking of services, although this may be less likely to occur. To encourage students to enroll in K-3 Plus before kindergarten, PreK programs should have information about K-3 Plus programs. While student performance improves when both PreK and K-3 Plus are combined, for those students who are only enrolled in K-3 Plus, student achievement gains are minimal, likely due to fidelity issues, including no students studied participated in the full 100 day program. When the program was scientifically studied, students who stayed with the same teacher through K-3 Plus and the school year showed positive gains.

Table 2. Percentage of Schools with both PED PreK and K-3 Plus programs, SY17

District	Percent of K-3 Plus Schools with PreK
Chama Valley	100%
Central	86%
Cobre	75%
Santa Fe	71%
Bernalillo	66%
Deming	66%
Hatch	66%
Taos	66%
West Las Vegas	66%
Grants-Cibola	60%
Albuquerque (including Charters)	45%
Roswell	42%
Los Lunas	40%
Gallup-McKinley	25%
Gadsden	20%
Espanola	18%
Artesia	0%
Belen	0%
Carlsbad	0%
Clovis	0%
Hobbs	0%
Jemez Mountain	0%
Las Cruces	0%
Las Vegas	0%
Lovington	0%
Mora	0%
Rio Rancho	0%

Source: LFC analysis of PED data; Note: Districts with 2+ schools with programs.

Table 3. Districts with Highest and Lowest School Average Percentage K-3 Plus Enrollment Summer 2017

District	Average K-3 Plus Enrollment Capacity
Vaughn	100%
Mesa Vista	93%
Las Vegas	82%
Wagon Mound	71%
Carrizozo	61%
Hagerman	58%
Jemez Mountain	55%
Alamogordo	54%
Bernalillo	52%
Zuni	50%
Gadsden	29%
Dexter	28%
Loving	28%
Espanola	27%
Albuquerque (with Charters)	26%
Artesia	26%
Gallup-McKinley	26%
Rio Rancho	26%
Belen	25%
Ruidoso	25%
Socorro	24%
Chama Valley	23%
Grants-Cibola	20%
Carlsbad	18%
Lovington	16%
Hobbs	15%

Source: LFC analysis of PED data;
Note: For a full list see Appendix D.

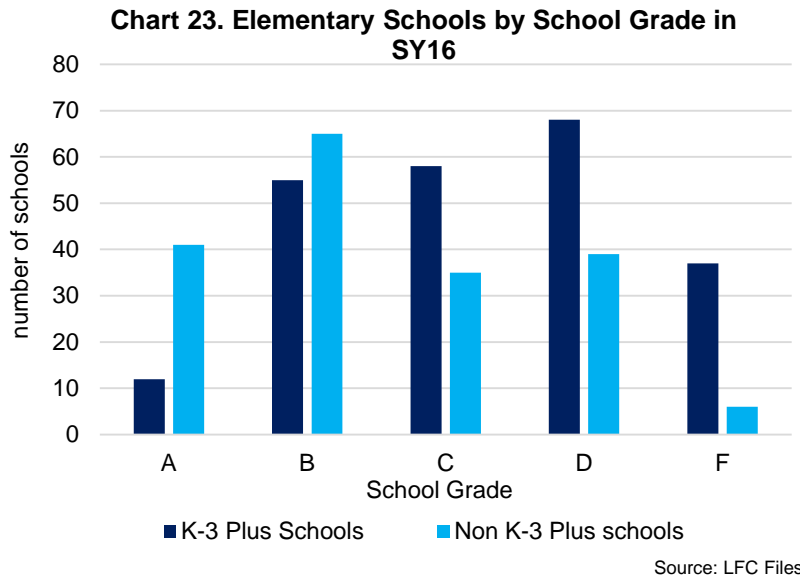
Not operating K-3 Plus on a schoolwide basis undermines correct implementation of the program. Currently, the majority of students at K-3 Plus schools do not participate in the program. For a cohort of 27 thousand students who entered kindergarten in SY2013, only 6.4 thousand participated in K-3 Plus at some point from kindergarten through third grade. Of almost 13 thousand students who took the third grade PARCC test in K-3 Plus schools, only about 35 percent of students were ever enrolled in K-3 Plus (Table 3). Sixty-five percent of those students who enrolled in K-3 Plus (and were in a K-3 Plus school for whom LFC staff had DIBELS scores) enrolled for one year, and most, 70 percent, started the summer before second grade. If students do not receive the full K-3 Plus treatment of 100 extra days of school over four years, they will not gain the full expected benefits of the program. Currently, the state has not set clear performance expectations for K-3 Plus, as implementation of the program does not allow for a majority of students to enroll for all four years, but the state expects students' outcomes to improve as if they were enrolled for four years.

In the summer of 2017, K-3 Plus served an average of 33 percent of students at schools where the program was implemented. As the goal of K-3 Plus is to provide an additional 100 instructional days from kindergarten through third grade, it is important for the program to be implemented schoolwide and to have districts implement the program at all eligible schools. PED and school districts should work together to prioritize funding to schools who commit to at least 50-75 percent K-3 Plus enrollment schoolwide.

If K-3 Plus was implemented schoolwide, it may be easier to place students with their same teacher, likely improving the impact of the program. There is evidence the program's impact is reliant on being placed with the same teacher for K-3 Plus and the regular school year, as well as having the program end within two weeks of the next school year. Fifteen percent of students enrolled in K-3 Plus were proficient in reading, compared to 25 percent of those who were not enrolled. However, as there was no reliable baseline measure available for the majority of students who were in kindergarten in SY13, and as districts frequently refer students who are behind to K-3 Plus programs, selection bias was not fully controlled for. More comprehensive examination of K-3 Plus programs is needed once a sufficient sample of students participate in the program for multiple years and when the programs are implemented to fidelity, including maintaining the student with their same teacher.

State law targets K-3 Plus funding to the most challenged schools, however the state does not prioritize funding to schools with other support interventions. K-3 Plus programs are targeted to schools with high proportions of low-income populations or low performing schools. As shown in Chart 23, most K-3 Plus schools had a D school grade in SY16, with C school grades the next most common. However, non-K-3 Plus schools were more likely to have better school grades, with most schools scoring a B. If schools are only implementing K-3 Plus, without other interventions to improve teacher quality (such as Teachers Pursuing Excellence), principal quality (such as Principals Pursuing Excellence), or additional early interventions (such as PreK), the effect of K-3 Plus alone may not be enough

to significantly improve student outcomes, especially if the program is not run to fidelity.



Recommendations

The Legislature should consider:

- Continued expansion of prekindergarten programs, including extended day services.
- Closing funding gaps for K-3 Plus and requiring PED to prioritize funds for schoolwide or districtwide implementation.
- Monitoring the implementation of K-3 Plus and specifying more statutory requirements absent administrative changes by PED.

The Public Education Department and school districts should:

- Align and coordinate below-the-line funding for school districts and schools.
- Enhance performance and fidelity monitoring of K-3 Plus programs and ensure programs provide 25 additional instructional days and teachers stay with the same students.
- Prioritize funding to schools willing to implement K-3 Plus schoolwide and who maintain fidelity to the program by keeping students with their same teacher and who end K-3 Plus programs within two weeks of the start of the school year.

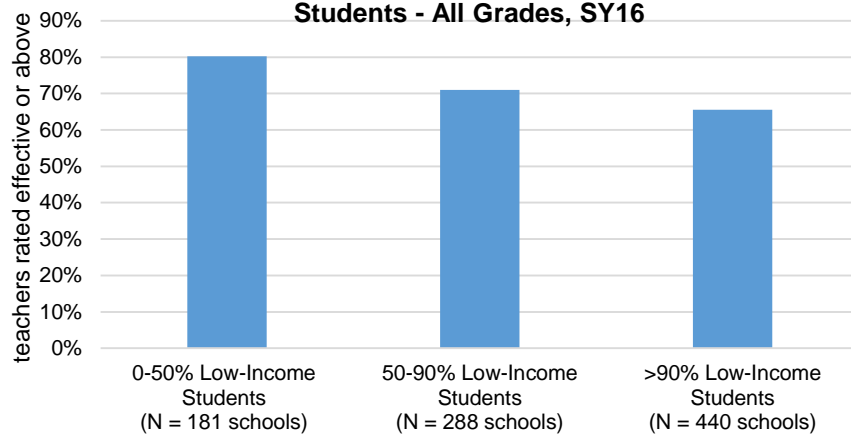
Teacher effectiveness and school grades vary by schools' rates of at-risk students

Teacher effectiveness varies by school risk factors.

New Mexico's teacher effectiveness rating system was revised in 2014, leading to fewer teachers being rated as effective, and New Mexico having the fewest teachers rated as effective or above in the country. Most states have less than 1 percent of teachers rated as less than effective, while New Mexico has almost 29 percent. High quality teachers are needed to improve student outcomes, specifically in schools with a large proportion of high risk students, as teachers are a main factor in student achievement and growth. Research has shown an effective teacher can increase each student's lifetime earnings by almost \$11 thousand, and if the class has 20 students, the teacher could raise the class aggregate earnings by just over \$200 thousand.

Schools with a high proportion of low-income or English learner (EL) students have a lower percentage of teachers rated as effective, highly effective, or exemplary. In SY16, 80 percent of teachers were rated as effective or above in schools with zero percent to 50 percent low-income students while only 65 percent of teachers were rated as effective or above in schools with over 90 percent low-income students (Chart 24). Some of this variation may be due to a higher proportion of new teachers at higher poverty schools, as found in a previous LFC evaluation. This previous finding examining the distribution of teachers along with the analyses shown in this report highlight that school districts do not systematically match their best teachers with their high need schools.

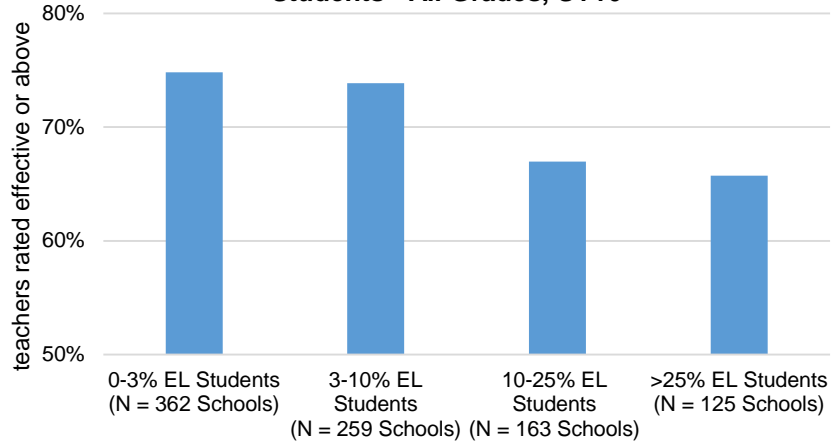
Chart 24. Percent of Teachers Rated Effective or Above in Schools by Proportion of Low-Income Students - All Grades, SY16



Source: LFC analysis of PED data; Note: Low-Income defined as free or reduced lunch eligibility.

When examining schools with different proportions of EL students, similarly, 75 percent of teachers were rated effective or above at schools with zero to three percent EL students while 66 percent of teachers were effective at schools with over 25 percent EL students (Chart 25). These data indicate schools with more teachers rated as effective or above have fewer at-risk students, as defined by low-income and EL status. The percentage of effective teachers at schools did not significantly differ by school mobility rate. In developing the current PED value-added model, PED looked at including demographic indicators, but did not see significant differences when so doing. If possible, adjusting the current value-added model used for teacher evaluations to control for low-income and EL status in a similar way as it controls for mobility, may be useful in decreasing variation caused by school level factors. Furthermore, school districts and PED should determine where the most effective teachers are located, and determine how to increase the number of effective teachers at high-risk, low performing schools.

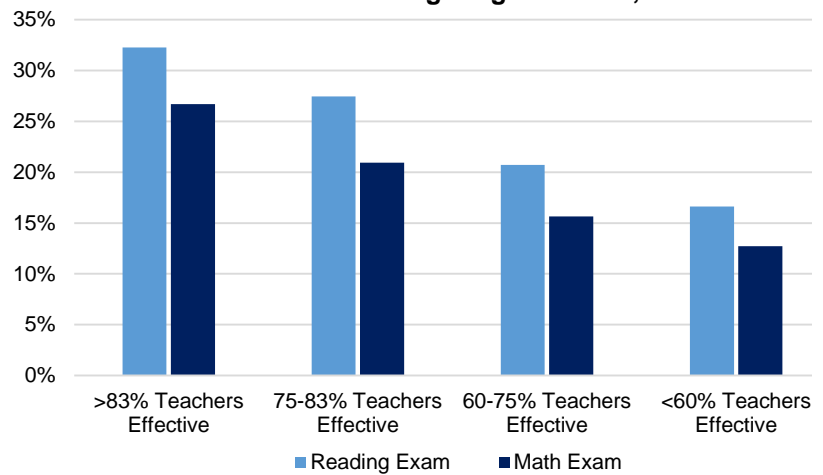
Chart 25. Percent of Teachers Rated Effective or Above in Schools by Proportion of English Learner Students - All Grades, SY16



Source: LFC analysis of PED data; Note: English Learner status defined as anyone who is currently enrolled or has since exited an English Learner program.

Schools with a higher percentage of teachers rated as effective or above also have a higher percentage of students achieving proficient scores on PARCC exams. However, even schools with a high proportion of highly effective teachers have less than 35 percent of students meeting proficiency standards since teacher evaluations examine student growth rather than student proficiency. After reviewing teacher effectiveness and PARCC proficiency data for students in sixth through eighth grade in SY16, LFC staff observed a positive relationship between a school's percent of effective or above teachers and the school's percent of students proficient on PARCC reading and math exams (Chart 26). As the teacher effectiveness value-added model is based on student achievement, these results help to validate the model. Additional information on teacher effectiveness and student achievement should be examined at the student or teacher level, and highly effective teachers should be incentivized to teach at low performing or at-risk schools.

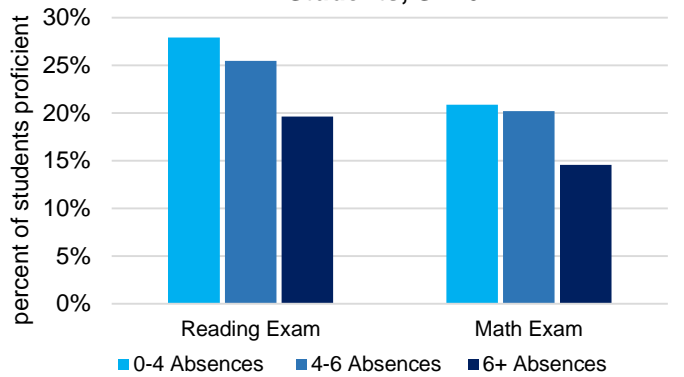
Chart 26. Percent of Students Proficient on PARCC by Proportion of Teachers Effective or Above in School - Sixth through Eighth Grade, SY16



Source: LFC analysis of PED data

Teacher absences are associated with decreased student proficiency rates in both reading and math. LFC calculated each school's average number of teacher absences in SY16 and compared the school student

Chart 27. School Percent of Students Proficient by School Average Teacher Absences - All Sixth through Eighth Grade Students, SY16



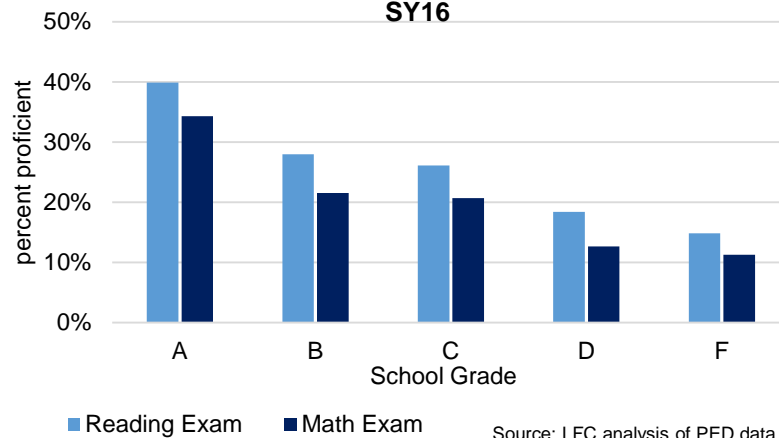
Source: LFC analysis of PED data

proficiency rate by the school’s average number of teacher absences. Schools with a higher number of average teacher absences had lower levels of middle school students scoring proficient on the PARCC reading and math exams (Chart 27). Therefore, teacher attendance is related to student performance, however, the largest drops occur after teachers miss at least six days of school. The current analysis considered days absent at school, but not absent from the classroom. However, if teachers are consistently absent from the classroom for meetings or other reasons, it is expected student test scores would also be negatively affected. Currently, teachers are allowed to miss six days before those absences are calculated into their effectiveness rating.

School grades are impacted by school risk factors as well as teacher effectiveness.

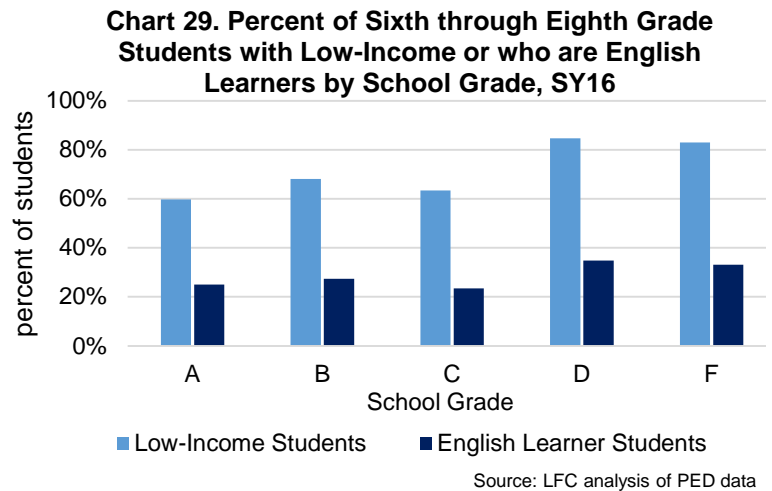
Better school grades are associated with higher proficiency rates, fewer at-risk students, and more teachers rated as effective or above. The Legislature enacted an A-F school grading system during the 2011 regular legislative session (Laws 2011, Chapter 10; SB427). Consequently, PED assigns an A to F letter grade to schools each year based on measures of academic proficiency, academic growth, and school practices. LFC found middle schools with better A-F school grades in SY16 had higher average PARCC reading and math proficiency rates than other schools (Chart 28). Schools with high or low PARCC proficiency rates can receive an A grade because the A-F school grading system uses multiple measures in addition to proficiency rates. This explains how A schools can still have low proficiency rates. Consequently, A-F school grades are an indicator of a school’s performance relative to other schools rather than an indicator of a school’s absolute academic performance.

Chart 28. Percent of Sixth through Eighth Grade Students Proficient on PARCC by School Grade, SY16



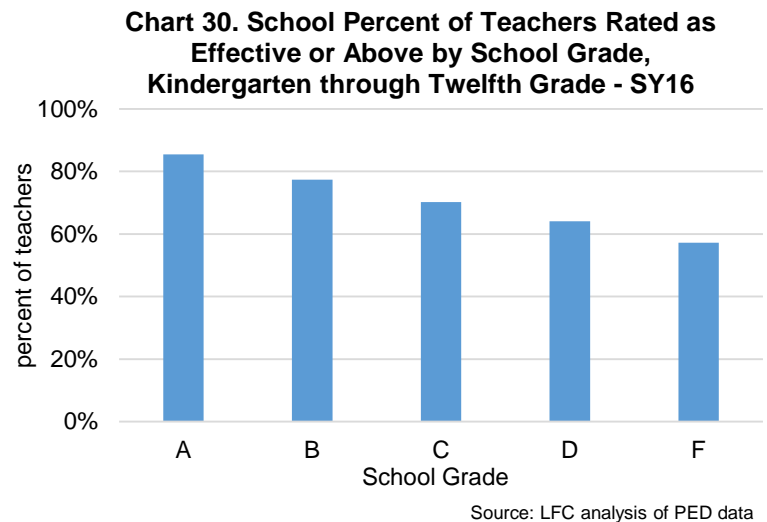
Source: LFC analysis of PED data

Middle schools with an A grade in SY16 had 23 percent fewer low-income students and 26 percent more teachers rated as effective than middle schools with an F grade (Charts 29 and 30). These data suggest school grades directly measure levels of academic success and indirectly reflect distributions of at-risk students and effective teachers. Although the A-F school grading system takes school mobility rates and academic growth into account, PED should consider controlling for the effects of student risk factors in its school grading value-added models, as the current school grading technical guide does not include the rate of low-income or English learners in school grading. Furthermore, as expected, more effective teachers are at A schools than F schools. To improve student performance at low performing schools, PED and school districts should focus on how to attract and retain effective teachers at these schools.



Regardless of high proportions of at-risk students, some schools can still achieve high levels of academic proficiency and good A-F school grades. Previous LFC analyses have found that some at-risk schools implementing best practices can “beat the odds” and achieve high performance. For example, an elementary school in 2016 with 100 percent low-income and 67 percent EL students achieved 59 percent reading proficiency, 46 percent math proficiency, and an A grade. These schools indicate that school practices, as well as student at-risk factors, influence the A-F letter grade a school receives.

School districts and PED should identify strong interventions to attract and retain effective teachers at at-risk schools. As the most effective teachers according to PED’s evaluation system are not located at low performing schools, school districts and PED should determine how to attract highly effective teachers to at-risk schools (schools with a high proportion of English learner, low-income, and mobile students). If students have three years of highly effective teachers, their math and reading scores can increase by 16 percent, however if students have ineffective teachers, their scores can drop by as much as 33 percent. As students in high risk schools are inherently higher need and as these students have more dollars allocated to them in the funding formula, it is important to determine how to improve these student outcomes and how to most effectively use this additional allocation. Improving the effectiveness of teachers in these at-risk schools as well as attracting more effective teachers to these schools would likely improve student outcomes. There are a number of potential strategies to improve teacher effectiveness including incentive payments for highly effective teachers who move to low performing schools. However, school



districts and PED should determine what interventions have been shown to work elsewhere that would be most appropriate for each region or district in the state.

Recommendations

The Public Education Department should:

- Include teacher effectiveness information in PED accountability reports in order to determine whether teachers are improving their effectiveness over time, especially at high risk or low performing schools.
- Determine if there are ways to better control for the effects of student risk factors, such as low-income and English learner status, in its school grading value-added models.

The Public Education Department and school districts should:

- Continue to identify and implement interventions to attract and retain highly effective teachers in high risk schools.



Insert Here



Appendix A: Evaluation Scope and Methodology

Evaluation Objectives.

- Examine the long-term outcomes associated with risk factors such as low-income and special education status.
- Assess how student mobility and other factors affect student and school performance.
- Analyze how teacher, programming, and school characteristics affect student achievement.

Scope and Methodology.

- Interviewed PED and school district staff.
- Interviewed national education researchers.
- Reviewed state and federal laws, regulations, and policies.
- Reviewed relevant performance measures, administrative data, and related documents.
- Reviewed existing research on student risk factors and academic achievement.
- Reviewed national best practices.
- Reviewed and analyzed state public education fiscal data.
- Reviewed and analyzed teacher effectiveness data from PED.
- Reviewed and analyzed longitudinal student demographic and outcome data from PED.

Evaluation Team.

Dr. Sarah Dinces, Program Evaluation Project Lead
Clayton Lobaugh, Program Evaluator

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 Public Education Department Deputy Secretaries and their staff on November 13, 2017.

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

Appendix B: Detailed Research Methodology

Multiple cohorts of students were analyzed when examining different findings. The cohorts used in each finding will be discussed along with the analyses used for each finding. For all analyses with significance testing, our significance level was $p=0.05$.

Student Achievement and Growth

Student Achievement: Multiple cohorts were examined for student achievement and growth. The two most commonly used cohorts included all student demographic and test information from SY13-SY16 or SY08-SY13. Within these cohorts, cumulative enrollment in services for at-risk groups such as Free and Reduced Lunch (FRL; a measure of low-income), English Learner (EL) status, special education enrollment, and mobility were calculated. For the EL cumulative measure, as student status may change from current enrollment, to exited for 1-3 years, group enrollment was determined based on their highest level achieved. Special education enrollment included students who were twice exceptional (i.e. requiring both special education and gifted services), and was examined from K-3 as well as from 5th-8th because there were no measures of gifted education for SY08 and SY09, leading to uncertainty regarding which children were in special education for gifted. To determine student test scores from SY08-SY13, descriptive statistics were run to return the mean for each year. To determine the effects of cumulative FRL, EL or special education enrollment, analyses of variance (ANOVAs) were run.

For mobility analyses, student mobility was first calculated by determining if the student stayed in the same school from one year to the next. Then these moves were summed across the years of interest in the cohort. To determine the number of kindergarteners who moved between kindergarten and third grade, frequency analyses were run on the number of moves variable, selecting only for students who were enrolled in kindergarten in SY13. Similar analyses were completed for third graders in SY13. Next, frequency analyses were run to determine the percent of students who move year by year, again selecting for students who were in kindergarten or third grade in SY13. Next, cross tabs were run to determine whether students who have a higher number of moves have a higher rate of risk factors (FRL, EL, special education enrollment). Additional crosstabs analyses were done at the school level to determine if schools with higher rates of mobility also have higher rates of student risk factors. ANOVAs were run to determine whether the number of moves a student has affects student test scores. These were run for students who were in SY13 in third, sixth, and tenth grades. Finally, ANOVAs were also run examining school level effects of mobility on student proficiency rates, however these analyses were not used to make any significance claims as no MLM was run but were used to collect descriptive statistics on school level effects of mobility.

Student Growth: Growth scores were determined by standardizing the two years of interest, most often SY08 and SY13 and taking the difference between the later and earlier year. To determine the number of days of learning measurement, methodology was borrowed from the 2017 Credo Texas Charter school study. Using their methodology, for every .01 standard deviation change, 5.7 days of learning should be added or subtracted. The days of learning variable should be used to help generally estimate the effects of student growth in a tangible way, and is an imprecise measurement.

To determine the high and low growth districts, student individual growth was averaged across district for both reading and math. These scores were then averaged together to create a composite district score. Districts were then ranked by this composite score. ANOVAs were run to determine the effects of student risk factors on student growth, using the same cumulative measures used for student achievement analyses. For the mobility analyses with student growth, data was examined between SY11 and SY14, rather than SY08 to SY13, but however the growth scores were calculated in similar ways.

Early Childhood Program Effects:

To determine the effects of PreK and K-3 Plus on PARCC test scores, propensity score matching was used to select matched control groups. For PreK, the treatment and control groups were matched on FRL (a measure of low-income), English Learner status (EL), level of special education integration, and district attended. Once the specific groups were created, one for math scores and one reading, analyses of variance and covariance were run. Within

these analyses, FRL and EL were controlled for as well as days present. Additionally days present was also examined as a dependent variable. The potential protective effect of mobility was also examined in further exploratory analyses.

In addition to examining the effectiveness of PreK on eighth grade test scores, a list of high and low performing PreKs was created by examining 3 years of PreK (2014-2016) and Kindergarten reading proficiency data (2015-2017). Data was standardized, and averaged by provider across these three years. Providers in the top and bottom 10 out of the 84 total providers were specified as high or low performing.

For K-3 Plus, the treatment and control groups were examined using a dataset that combined student attendance in K-3 Plus from Kindergarten through third grade (SY13-16). Additionally, students were only selected if they: 1) had PARCC third grade test scores, 2) were enrolled in a school with a K-3 Plus program, and 3) had first grade (SY14) DIBELS test scores. Therefore, only reading test scores were examined. Analyses of covariance (ANCOVA) were run, with K-3 Plus enrollment, and DIBELS scores as the main effects, as well as the interaction between these two variables included in the model. The dependent variable was 3rd grade PARCC reading scaled scores. Additional analyses were completed looking at the percentage of students enrolled in K-3 Plus compared to the total enrollment in grades K-3. These analyses were completed by using the 40 day demographics reported online and the K-3 Plus roster provided by PED. Finally, when determining schools which offer both PreK and K-3 Plus, rosters for both programs were compared by district.

Furthermore, due to the large variance in effects of K-3 Plus by district, high and low performing schools by district were examined. This was determined by examining the difference between those students who attended K-3 Plus programs and those who did not for all schools that had at least 10 students in total included in the sample. A weighted average for K-3 plus years was used to determine the average test scores for the K-3 plus group and the average score for those students who did not attend K-3 Plus was subtracted from the K-3 Plus group average score. Then schools were ranked and divided into deciles. The schools in the highest difference score deciles were the high performing schools, while the schools in the lowest difference score decile were defined as the low performing schools. In order to determine the number of students per district in high or low performing schools, two new datasets were created, one for the students in the low and one for students in the high performing schools and this was used to determine the average change in student performance by K-3 Plus enrollment.

Teacher and School Grade Effects:

For teacher effectiveness analyses, first, percent effective was calculated by dividing the total number of teachers in a school by the number of exemplary, highly effective and effective teachers for SY16. Then, this information was merged with the SY11-SY16 cohort dataset to examine effects on PARCC scores and the SY16 PARCC and demographic files to examine how demographic factors affect teacher effectiveness. Then, descriptive statistics were examined to determine how the percent of effective teachers may change based upon school percent low-income, percent EL (including all students who exited a program as well as those currently enrolled), school mobility. Teacher attendance was also calculated by school, and descriptive statistics were examined for teacher attendance by risk factor, as well as examining average percent proficiency rates for each school by teacher effectiveness and teacher attendance. Importantly, all teachers who were assigned to a district office were excluded from these analyses as they were not assigned to a specific school. Furthermore, for test score analyses only 6-8 grades were examined. In all analyses of teacher effectiveness, only teachers assigned to a school were included for percent of teacher effective.

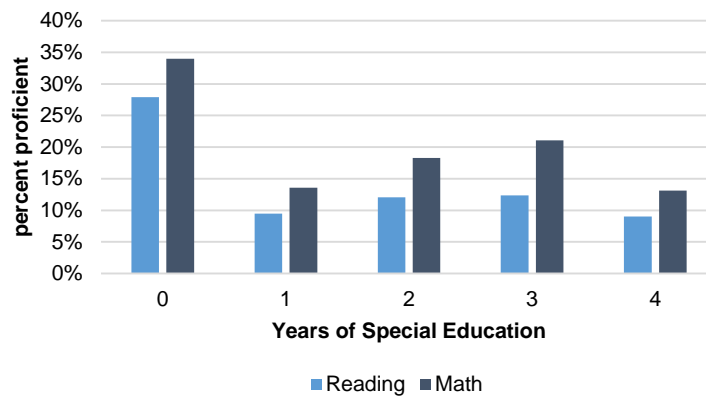
School grade analyses were determined by merging SY16 school grade information with the SY11-16 cohort dataset. Then descriptive analyses were run to compare school grade by average risk factors. Furthermore, the breakdown between school grade and student proficiency rates by school were examined by running descriptive statistics.

None of the analyses conducted for teacher effectiveness, teacher attendance, or school grade used significance testing as not enough information was available to run reliable Hierarchical Linear Models.

Appendix C: Effects of Student Proficiency and Growth Due to Special Education Status.

Students enrolled in special education from kindergarten through third grade from SY13-SY16 and from fifth through eighth grade from SY10-SY13 had lower test scores. However, students enrolled in two or three years of special education from kindergarten through third grade show improved test scores relative to those students enrolled for one or four years. This improvement may indicate the positive effects of early intervention. Thirty-four percent of students never enrolled in special education from kindergarten through third grade are proficient in math according to the 2016 PARCC while only 14 percent of students enrolled in one year of special education are proficient. The number increases to 21 percent if students are enrolled for three years (Chart 31). This improvement is likely due to students receiving needed early intervention services.

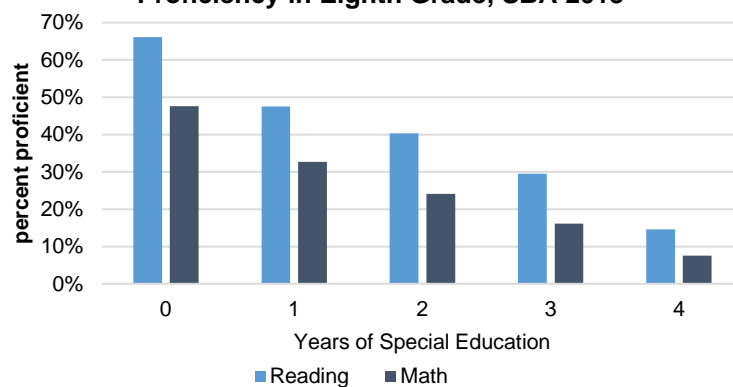
Chart 31. Effect of Special Education Status Kindergarten through Third on Student Test Proficiency in Third Grade, PARCC 2016



Source: LFC analysis of PED data

For students enrolled in special education from fifth through eighth grade, as expected, test scores decline as the number of years in special education increase, indicating while special education early in a student's educational career may improve test scores to some extent, these same benefits are not shown for later enrollment (Chart 32).

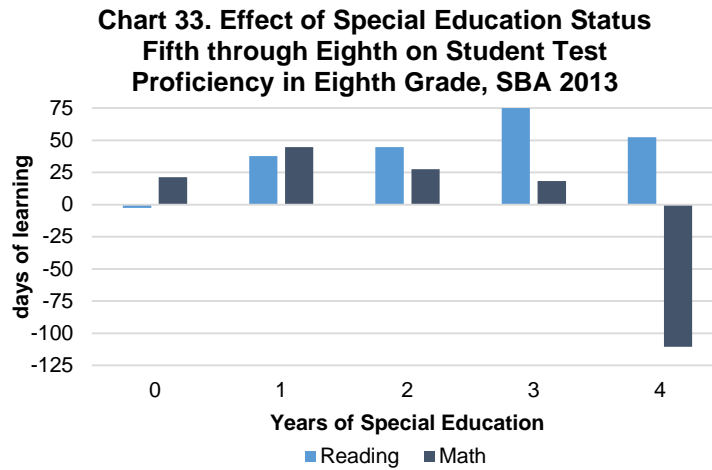
Chart 32. Effect of Special Education Status Fifth through Eighth on Student Test Proficiency in Eighth Grade, SBA 2013



Source: LFC analysis of PED data

For students enrolled in special education, the amount of growth a student achieves is dependent upon the number of years in special education. Specifically, students enrolled in special education for one to three years have positive growth scores in reading and math. While reading growth scores increase from one to three years of special education, at 38 to 83 respective additional instructional days, math growth scores decrease from one to three years

of special education at 44 to 18 respective additional instructional days (Chart 33). Those students enrolled in special education for four years have increased student growth in reading and highly negative student growth in math, learning at the equivalent of an additional 52 instructional days for reading and the equivalent of not receiving 110 instructional days in math. Further examination of student growth for special education students is needed to determine why students in special education for all years studied have such different patterns of growth in math and reading.



Source: LFC analysis of PED data. Note: growth score is the average difference of SY13 standardized reading and math SBA scores minus SY08 standardized reading and math scores.

Appendix D: School Average Percent of K-3 Plus Enrollment by District, Summer 2017

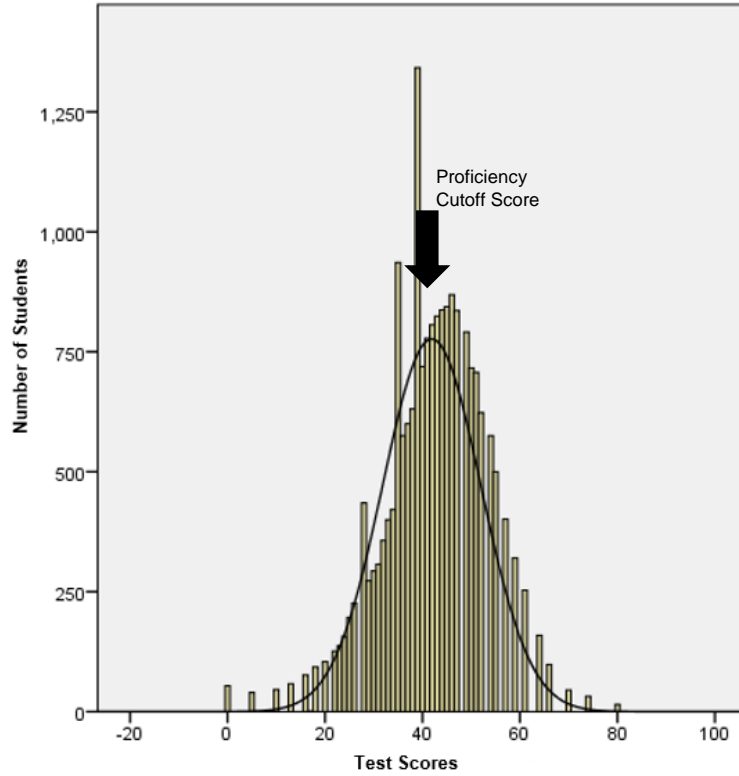
Table 4. School Average Percent of K-3 Plus Enrollment by District, Summer 2017

District	Average K-3 Plus Enrollment Capacity
Alamogordo	54%
Albuquerque (including Charters)	26%
Artesia	26%
Belen	25%
Bernalillo	52%
Carrizozo	61%
Carlsbad	18%
Central	44%
Chama Valley	23%
Clovis	41%
Cobre	41%
Deming	32%
Dexter	28%
Dulce	31%
Espanola	27%
Eunice	33%
Gadsden	29%
Gallup-McKinley	26%
Grants-Cibola	20%
Hagerman	58%
Hatch	37%
Hobbs	15%
Jemez Mountain	55%
Jemez Valley	40%
Las Cruces	35%
Las Vegas	82%
Lordsburg	42%
Los Lunas	30%
Loving	28%
Lovington	16%
Maxwell	43%
Mesa Vista	93%
Mora	39%
Pecos	32%
Questa	43%
Rio Rancho	26%
Roswell	38%
Ruidoso	25%
Santa Fe	33%
Socorro	24%
Truth or Consequence	30%
Taos	35%
Vaughn	100%
Wagon Mound	71%
West Las Vegas	36%
Zuni	50%

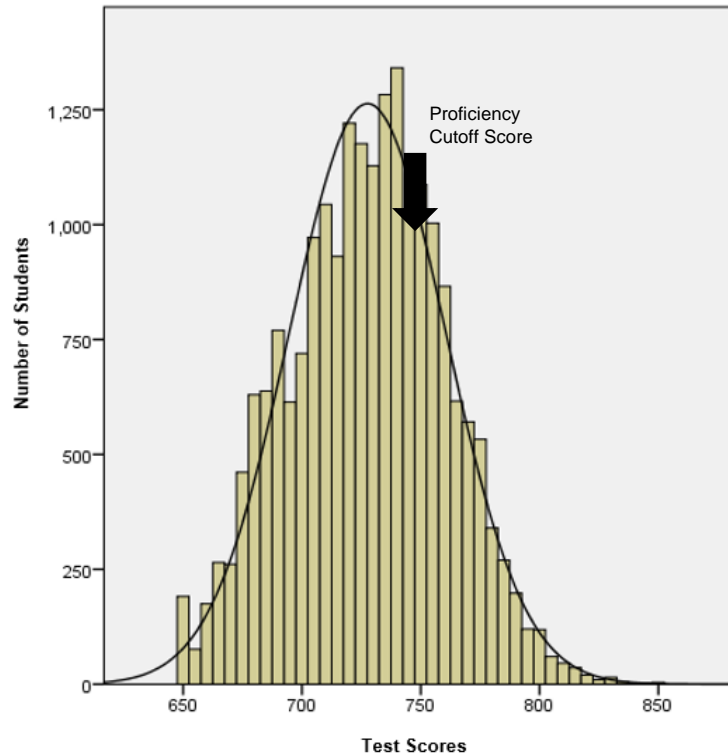
Source: LFC analysis of PED data

Appendix E. Eighth Grade SY13 SBA and SY16 PARCC Reading Score Distributions

SY13 Eighth Grade SBA Reading Scores
(40 = Proficient, N = 19,632 Students)



SY16 Eighth Grade PARCC Reading Scores
(750 = Proficient, N = 20,838 Students)



Appendix F. Selected Bibliography – Listed in Order Referenced

1. Center for Research on Educational Outcomes. (2013). National Charter School Study 2013. Stanford University. Available at <http://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf>
2. DePaoli, J.L., Balfanz, R., Bridgeland, J., Atwell, M., & Ingram, E.S. (2017). Building a Grad Nation: Progress and Challenges in Raising High School Graduation Rates. Civic Enterprises Everyone Graduates Center at the School of Education at John Hopkins University. Available at: <http://new.every1graduates.org/wp-content/uploads/2017/05/2017-BGN-Report-vFINAL.pdf>
3. Education Commission of the States. (2014). Rating States, Grading Schools: What Parents and Experts Say States Should Consider to Make School Accountability Systems Meaningful. Available at: http://www.ecs.org/docs/rating-states_grading-schools.pdf
4. Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44(5), p.293-307.
5. Legislative Finance Committee. (2014) Performance and Improvement Trends: A Case Study of Elementary Schools in New Mexico. LFC Program Evaluation Report #14-11. Available at https://www.nmlegis.gov/Entity/LFC/Program_Evaluations?Category=Public+Schools
6. Hanushek, E. A. (2002). Teacher Quality. In Izumi, L. T., & Evers, W. M. (Eds.), *Teacher Quality* (p.1-12). Hoover Press.
7. Jordan, H., Mendro, R., & Weerasinghe, D. (1997). Teacher effects on longitudinal student achievement. Dallas Independent School District.
8. Southwest Educational Development Laboratory. (2001). Are Our Teachers Good Enough? SEDL Newsletter, 13(2), p.1-23.
9. Heyneman, S. P., & Loxley, W. A. (1983). The Effect of Primary-School Quality on Academic Achievement across Twenty-nine High- and Low-Income Countries. *American Journal of Sociology*, 88(6), p.1162-94.
10. Gilliam, W. S., & Zigler, E. F. (2004). State Efforts to Evaluate the Effects of Prekindergarten: 1977 to 2003. New Haven, CT: Yale University Child Study Center.
11. Legislative Finance Committee (2017). 2017 Early Childhood Accountability Report. Available at https://www.nmlegis.gov/Entity/LFC/Documents/Program_Evaluation_Reports/Final%202017%20Accountability%20Report%20Early%20Childhood.pdf
12. Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. National Bureau of Economic Research Working Paper No. 16381
13. Philips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarski, M., Magnuson, K. A., & Weiland C. (2017). Puzzling It Out: The Current State of Scientific Knowledge on Prekindergarten Effects, A Consensus Statement. Duke University and the Brookings Institute. Available at https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf
14. Cann, D., Karakaplan, M., Lubke, M., & Rowland, C. (2015). New Mexico StartSmart K-3 Plus Validation Study: Evaluator’s Report. Available at: http://cepr.unm.edu/sites/creativeworks/Early_Childhood/EvaluatorReport.pdf
15. Education Week. (2017, January). New Mexico Earns a D on State Report Card, Ranks 49th in Nation. Education Week. Available at <https://www.edweek.org/ew/qc/2017/state-highlights/2017/01/04/new-mexico-state-highlights-report-page.html>
16. Center for Research on Educational Outcomes. (2013). National Charter School Study 2013. Stanford University. Available at <http://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf>
17. National Conference of State Legislatures. (2016). No Time to Lose: How to Build a World-Class Education System State by State. Available at http://www.ncsl.org/documents/educ/edu_international_finai_v2.pdf
18. Legislative Finance Committee. (2015). Early Childhood Services Accountability Report Card, Gaps Analysis, and Spending Plan. Available at: https://www.nmlegis.gov/Entity/LFC/Documents/Program_Evaluation_Reports/Early%20Childhood%20Services%20Accountability%20Report%20Card%20-%20Gap%20Analysis%20and%20Spending%20Plan.pdf

-
19. Hanushek, E. A. (2011). Valuing Teachers: How Much Is a Good Teacher Worth? *Education Week*, 11(3), p.41-45.
 20. Legislative Finance Committee. (2015). Public School Use of Additional Compensation on Recruitment, Retention, and Performance. LFC Program Evaluation Report #15-05. Available at https://www.nmlegis.gov/Entity/LFC/Program_Evaluations?Category=Public+Schools
 21. New Mexico Public Education Department. (2017). NMTEACH Technical Guide: Business Rules and Calculations. Available at <http://ped.state.nm.us/ped/NMTeachDocs/2017/NMTEACH%20Technical%20Guide%202016-2017.pdf>
 22. New Mexico Public Education Department. (2017). New Mexico A-F School Grading Technical Guide: Calculation and Business Rules for Schools and Districts. Available at <http://aae.ped.state.nm.us/SchoolGradingLinks/1617/Technical%20Assistance%20for%20Educators/Technical%20Guide%202017.pdf>
 23. Institute of Education Sciences. (2013). Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment. Available at <http://files.eric.ed.gov/fulltext/ED544269.pdf>