



REVIEW OF DATA QUALITY AND METHODS FOR SCHOOL GRADES AND TEACHER EVALUATIONS

Presented to the Legislative Education Study Committee and the
Legislative Finance Committee
Las Vegas, NM

By
The Coalition for Excellence in Science and Math Education
(CESE) – <http://www.cese.org>
27 August 2014

M. Kim Johnson
kimber@comcast.net

HISTORY

“...things which apply elsewhere often do not apply in New Mexico.”

Lew Wallace, circa 1880

“And then he quit trying to effect change and wrote Ben-Hur.”

M. Kim Johnson, Circa Many Years Ago



CESE BACKGROUND

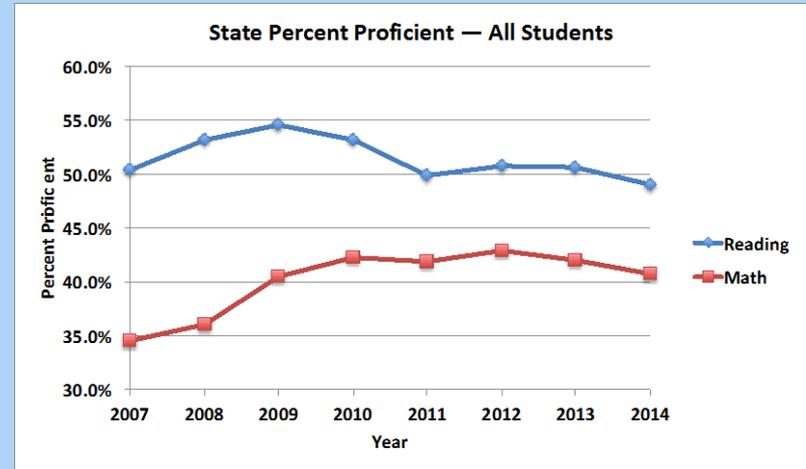
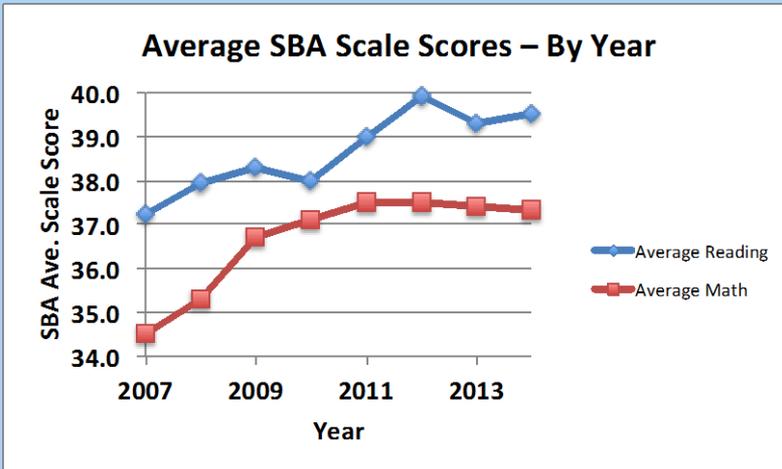
- **CESE, is a non-profit, non-partisan 501(c)(3)charitable corporation**
- **Members include National Laboratory personnel and retirees, industrial scientists, educators, parents, college professors, etc.**
- **We have analyzed New Mexico public education data and policy issues for over 15 years**
- **Our primary focus is to help improve New Mexico schools using New Mexico unique data**

CONTENTS

- NMSBA Test based performance results in the last 7 years – School and student performance results and short analysis
- ABCDF Act, as implemented – Some background, good news, and critique
- How to Supplement the ABCDF Act – We must show schools HOW to improve and close the Achievement Gap
- Teacher Evaluations – The 50% based on Student Growth – This is a potentially serious problem that needs addressing

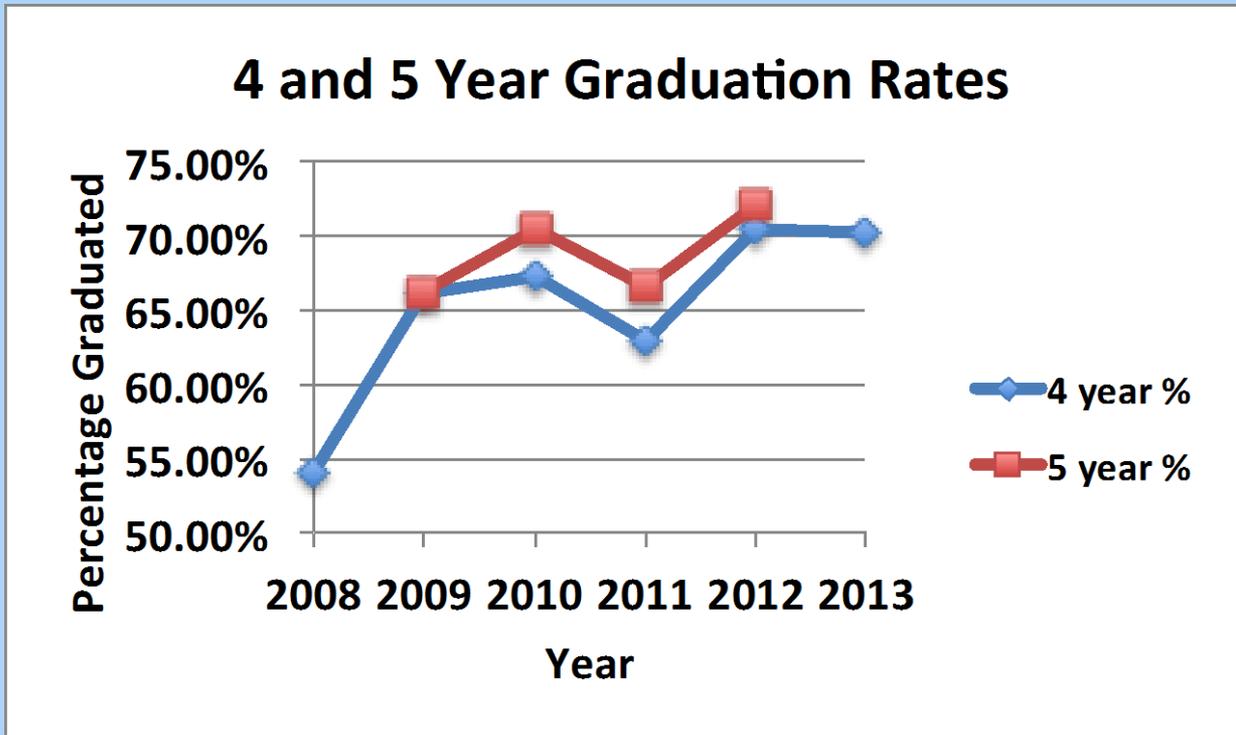
Past Performance from 2007 Using New Mexico Standards Based Assessment Tests

AVERAGE NEW MEXICO STANDARDS BASED ASSEMENT (NMSBA or SBA) SCALE SCORES AND PROFICIENCIES FOR THE STATE BY YEAR SINCE 2007



- Average scale scores have been trending generally upwards for math until 2011.
- Reading scale scores have trended upward with an anomaly in 2010 and have slightly decreased in 2013 and 2014 from its high in 2012.
- Proficiency percentages follow a similar but smoother pattern as scale scores.
- The adoption of Common Core Standards probably does not explain recent trends (NMSBA is based on different standards).

STATE GRADUATION RATES – ALL HIGH SCHOOLS



- The rates have been trending generally upward except for 2010 to 2011, and 2012 to 2013 is level (4 year rate).



ABCDF Grading System

Analysis

WHY DO WE USE A SCHOOL “GRADING” SYSTEM RATHER THAN MORE SIMPLE AND DIRECT TEST SCORES?

- State law requirement (the ABCDF Act).
- The Federal Department of Education instituted “alternate” state evaluation methods to NCLB or “waivers,” and this current form has been approved.
- Without the waivers, essentially all schools in all states would have failed NCLB requirements by 2014.
- We have no choice but to accept either monetary loss and punitive measures if we don’t get the waiver.

LET'S LOOK AT AN EXAMPLE PED GRADE SHEET

NM PED School Grade Report Card 2013
New Mexico Public Education Department

Final Grade
B

Certified

District: [REDACTED]
Grade Range: 07-08 [REDACTED]

This School [REDACTED]
Statewide C Benchmark [REDACTED]

		Grade	School Points	Possible Points
Current Standing How did students perform in the most recent school year? Students are tested on how well they met targets for their grade level.	21.3	A	33.64	40
School Growth In the past 3 years did the school increase grade level performance? For example did this year's 3rd graders improve over last year's 3rd graders?	5.8	B	8.30	10
Student Growth of Highest Performing Students How well did the school help individual students improve? The highest performing students are those whose prior scores placed them in the top three quarters (75%) of their school. Individual student growth over the past 3 years is compared to the state benchmark.	7.2	B	12.88	20
Student Growth of Lowest Performing Students How well did the school help individual students improve? The lowest performing students are those whose prior scores placed them in the bottom quarter (25%) of their school. Individual student growth over the past 3 years is compared to the state benchmark.	15.3	F	6.38	20
Opportunity to Learn Does the school foster an environment that facilitates learning? Are teachers using recognized instructional methods, and do students want to come to school?	7.5	B	8.88	10
Bonus Points Does the school show exceptional aptitude for involving students and parents in education, reducing truancy, and promoting extracurricular activities?	1.6		1.00	5
			Total Points	71.07

3 Year Average

Overall Points: 58.6
C

Final School Grade

75.0 to 100.0	A
60.0 to 74.9	B
50.0 to 59.9	C
37.5 to 49.9	D
0.0 to 37.4	F

School Grading 2013

This is the % proficient combined for math and reading scale scores.

Proficiency is a federal requirement. This is Value Added Model (VAM) adjusted, which is a questionable practice for this application.

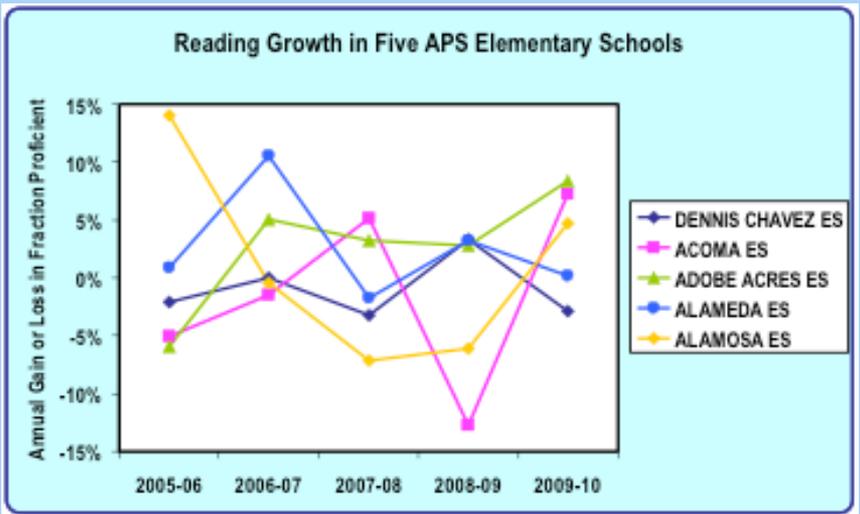
VAM adjusted School "Growth" is used even though growth is chaotic in the short term and favors the more disadvantaged demographic schools while dis-favoring the more advantaged.

Why is this "40" points (divisible by 10) instead of perhaps 28.3 or 42.8?

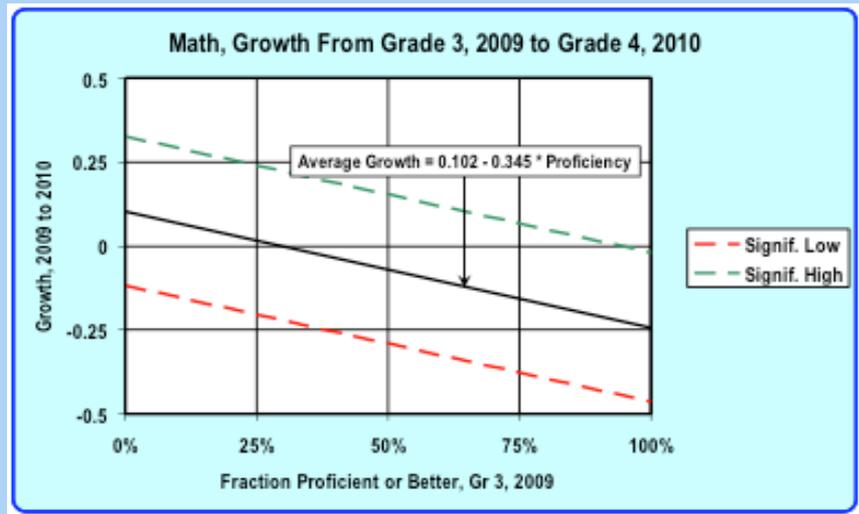
Why is this "10" points?

THESE WEIGHTINGS ARE VERY IMPORTANT BUT ARE NEVER JUSTIFIED

LET'S LOOK AT GROWTH



Short-term growth is somewhat random and NOT a good measure of how a school is performing overall.

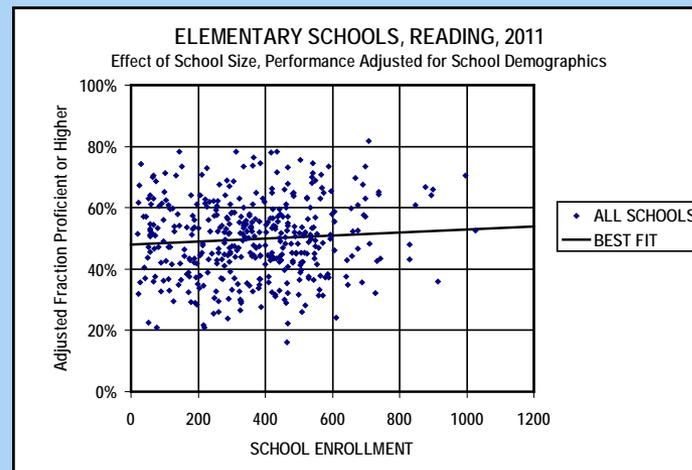


Direct growth measures favor the more disadvantaged demographic schools while disfavoring advantaged demographic sectors.

LET'S LOOK AT VAM AS IMPLEMENTED

The NM PED VAM adjusts for*:

- Proportion of student body that is FAY**
- School size (total enrollment)
- Students' prior scaled scores aggregated by school



- School size does not significantly correlate with NMSBA.
- Prior performance correlates with demographics AND everything else. (According to W.L. Sanders, prior performance contains all demographic information, but does it do so in a useable manner?)
- FAY provides relatively low correlation to performance.

*From the PED "New Mexico "School Grading Technical Guide Calculation and Business Rules" (Date not shown, but posted in 2012)

** Full Academic Year

HOW DO PED ABCDF SCHOOL GRADES COMPARE TO NMSBA SCALE SCORES?

PED
 LETTER
 GRADES

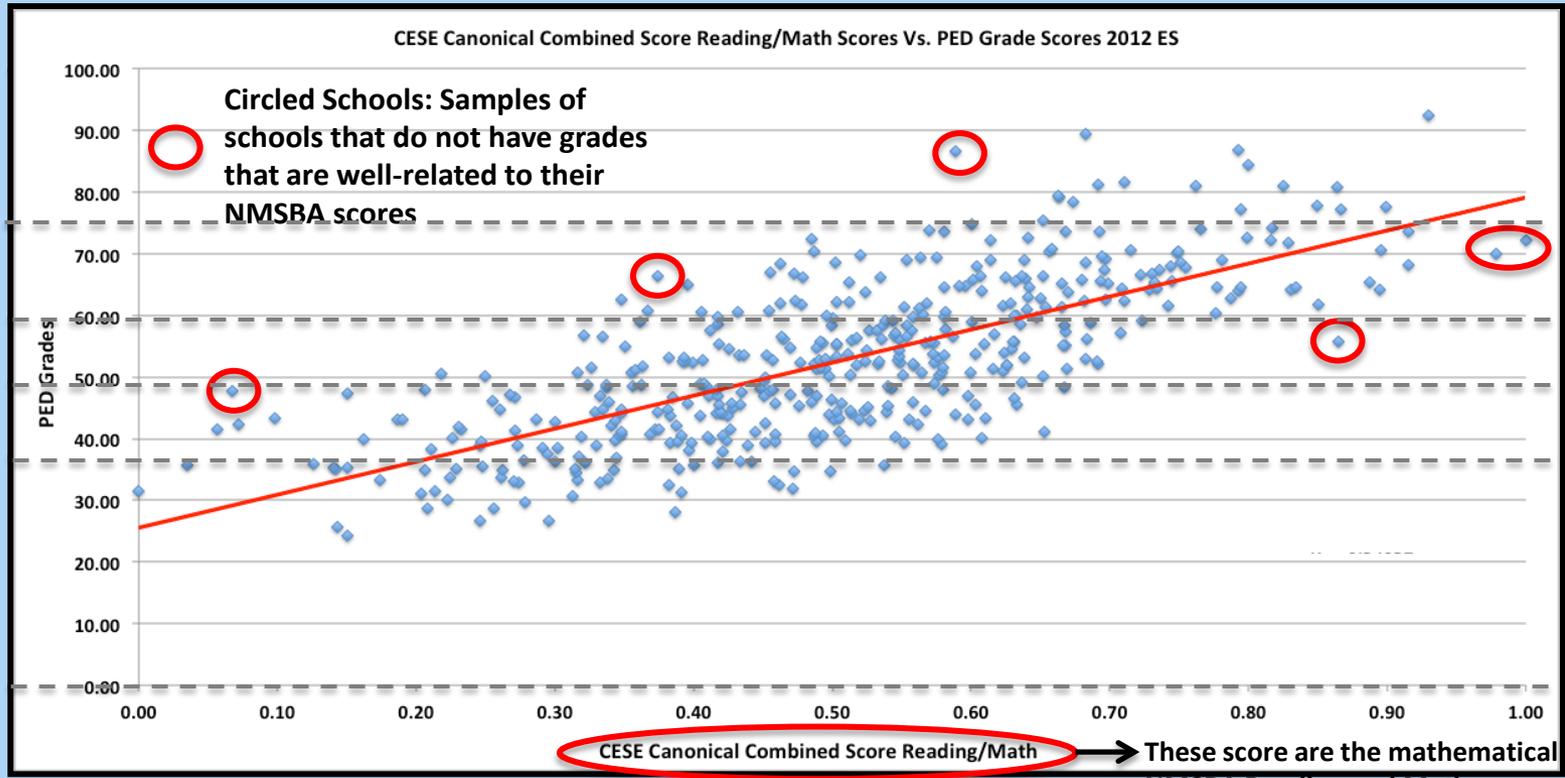
A

B

C

D

F



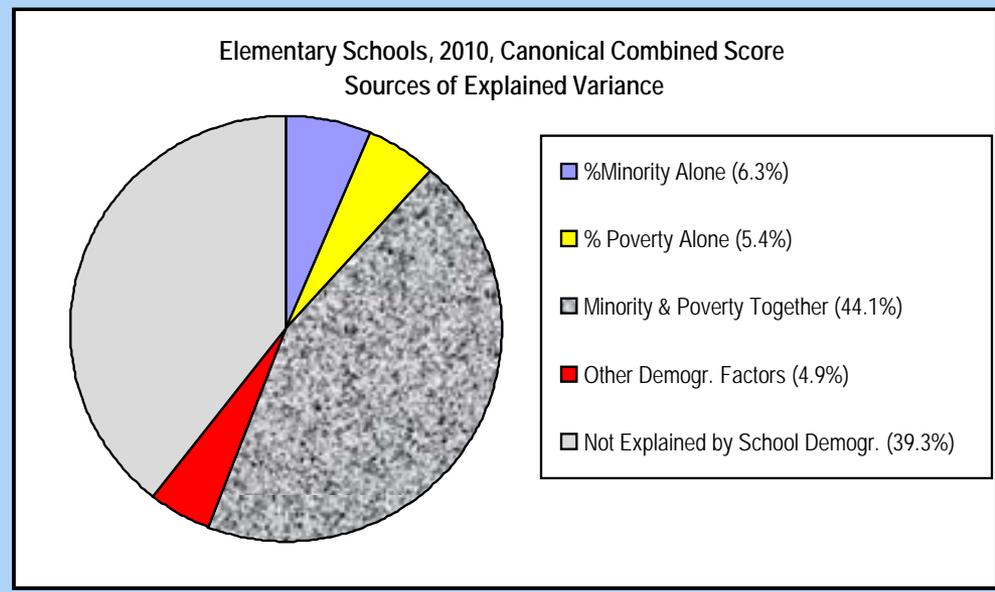
ABCDF CONCLUSIONS

- **What the ABCDF Act does to help NM Schools:**
 - ✓ It provides immediate relief to the NCLB requirements that all students be proficient by 2014
 - ✓ It sets new goals for improvement (AMO's – Annual Measurement Objectives or SGT's for the NM waiver –Student Growth Targets)
- **The ABCDF data:**
 - ✓ “Kind of” reflect actual school/student performance, but with some significant anomalies
 - ✓ Appear to be too complex to show a path to improvement
- **Without a “Path to Improvement” New Mexico will not get better at turning out well-educated K-12 students and meet the Student Growth Targets.**

What Do We Suggest? –THE CESE METHOD–

How we might get better
Without Silver Bullets

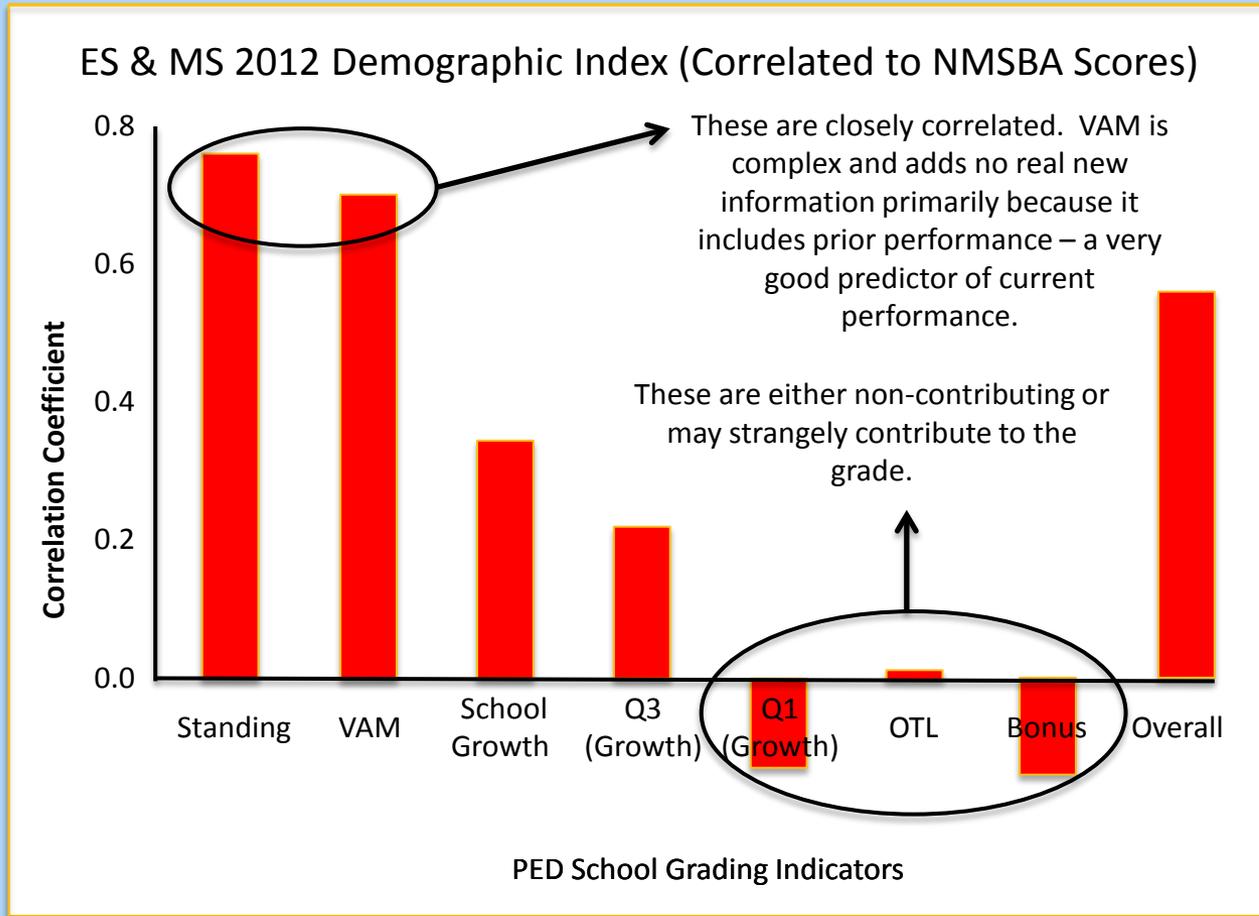
EFFECTS OF SCHOOL DEMOGRAPHICS ON PERFORMANCE – ONE EXAMPLE



- **Poverty by itself is not the major factor**
- **Minority status by itself is not a major factor**
- **But, the combination of minority status and poverty overwhelms all other factors**
 - ✓ **Minority students tend to be economically disadvantaged**
 - ✓ **Economically disadvantaged students tend to be minorities**

Between 60% - 80% of school performance is explained by school demographics

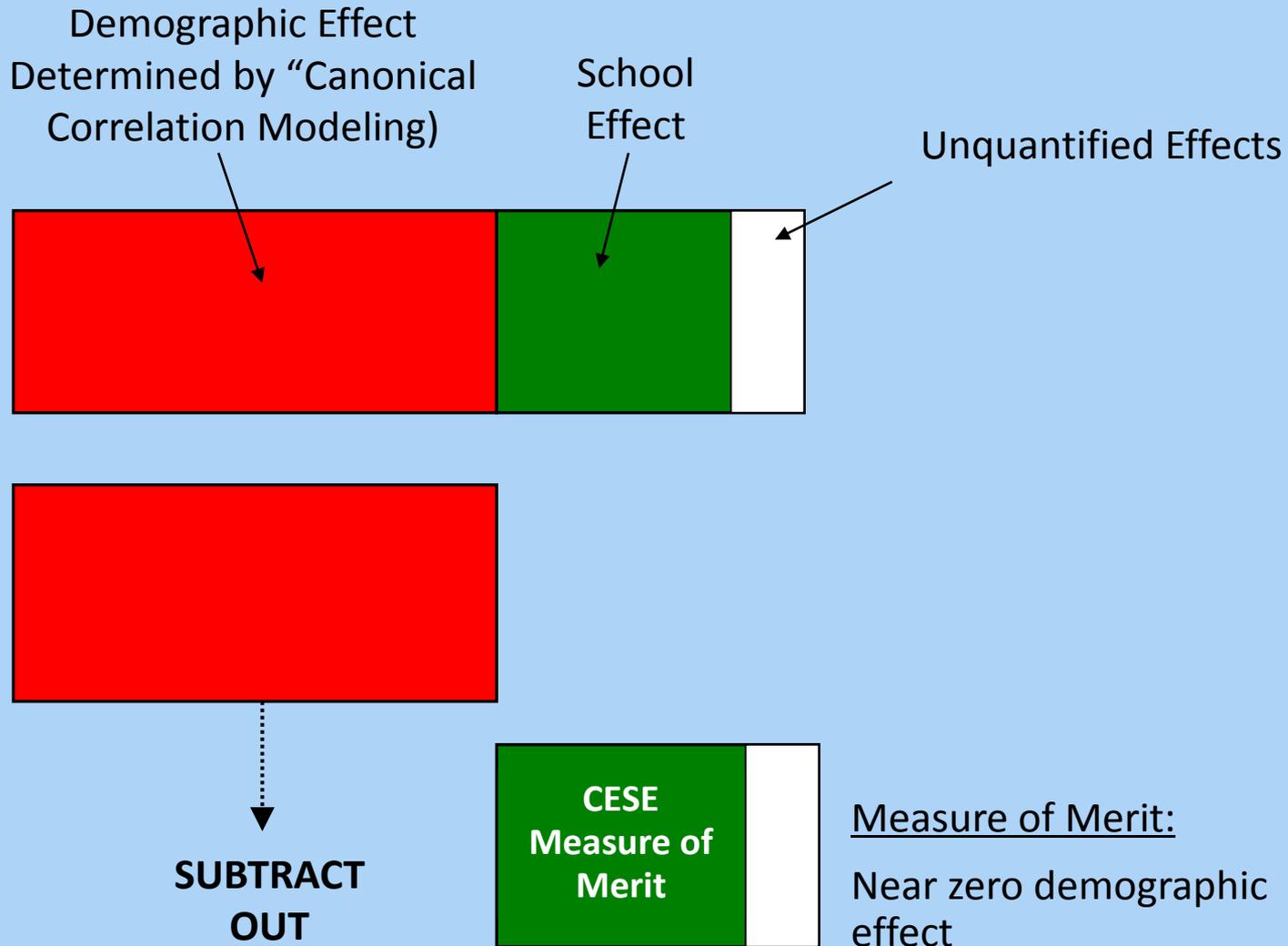
HOW DO THE VARIOUS GRADING ELEMENTS USED BY THE PED CORRELATE TO DEMOGRAPHICS AND THUS NMSBA SCORING?



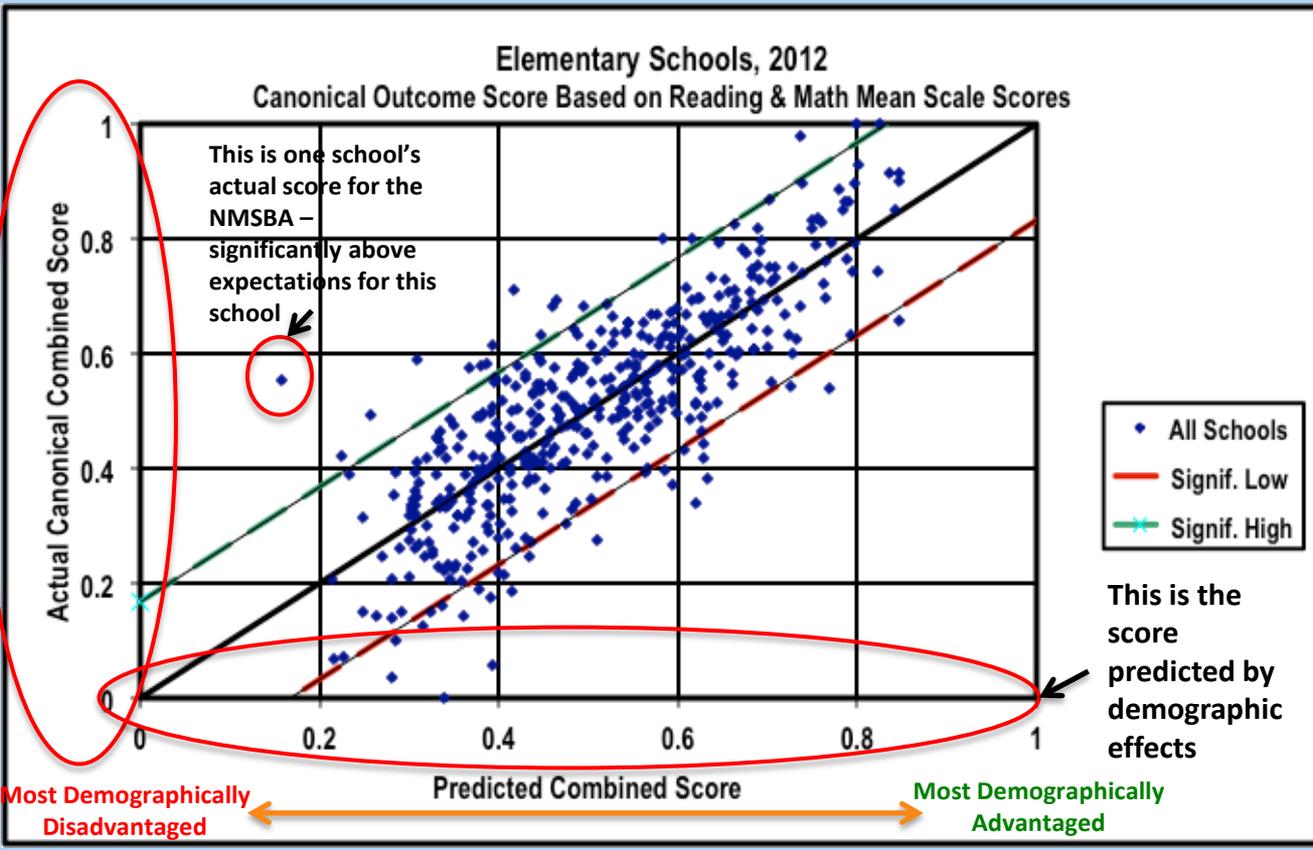
THE CESE APPROACH TO IMPROVING SCHOOLS

- **CESE developed an objective method that accounts for factors beyond schools' control.**
 - ✓ **Minority population**
 - ✓ **Students learning English**
 - ✓ **Students with disabilities**
 - ✓ **Poverty percentage**
 - ✓ **Student mobility**
- **The method also shows schools' comparison of performance to standards.**

GRAPHICALLY – WHAT IS A MEASURE OF MERIT?



COMPARISON OF ACTUAL TO PREDICTED SCORES

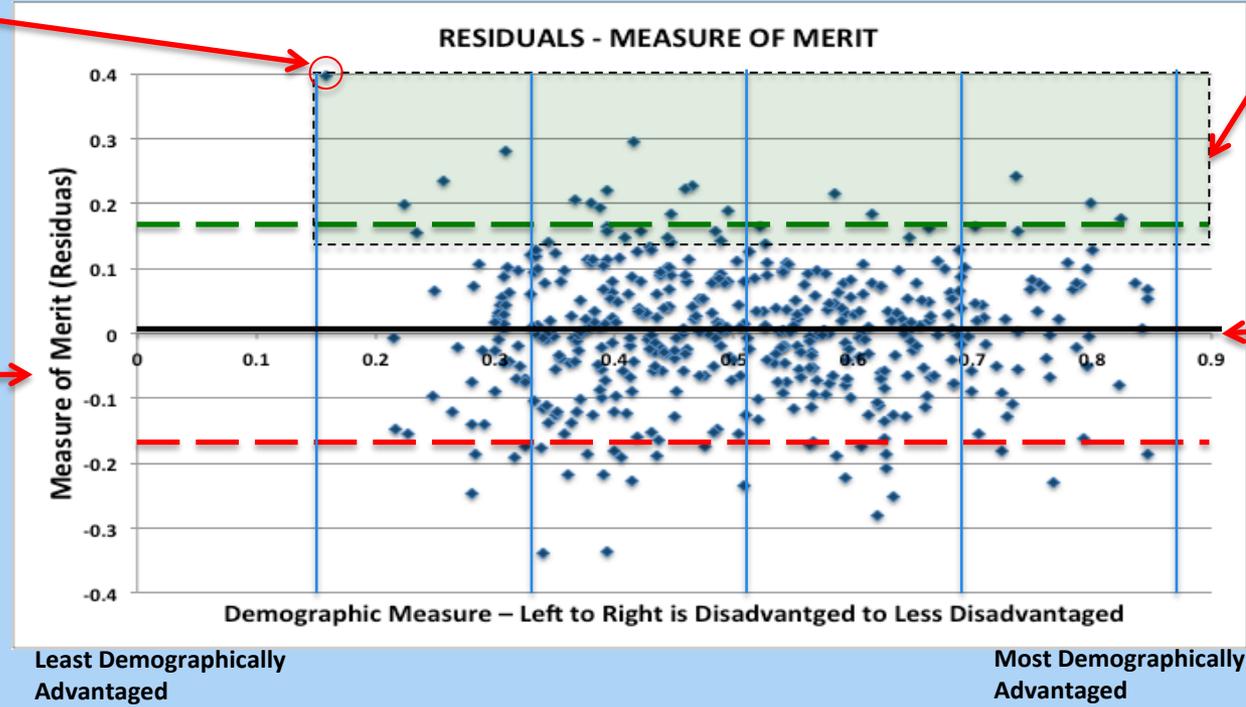


- Data shows NM schools that significantly outperform predictions and are candidate models for HOW to improve.
- This also shows how well schools perform with respect to the state NMSBA test results.

ROTATING THE PREVIOUS CHART TO SHOW RESULTS WITH REMOVED DEMOGRAPHIC EFFECTS AND PROVIDES METHOD TO IMPROVE PERFORMANCE

Same school circled on previous slide

Actual NMSBA scale score minus predicted from the last slide



Highest Performing Schools

Predicted scale score line from last slide

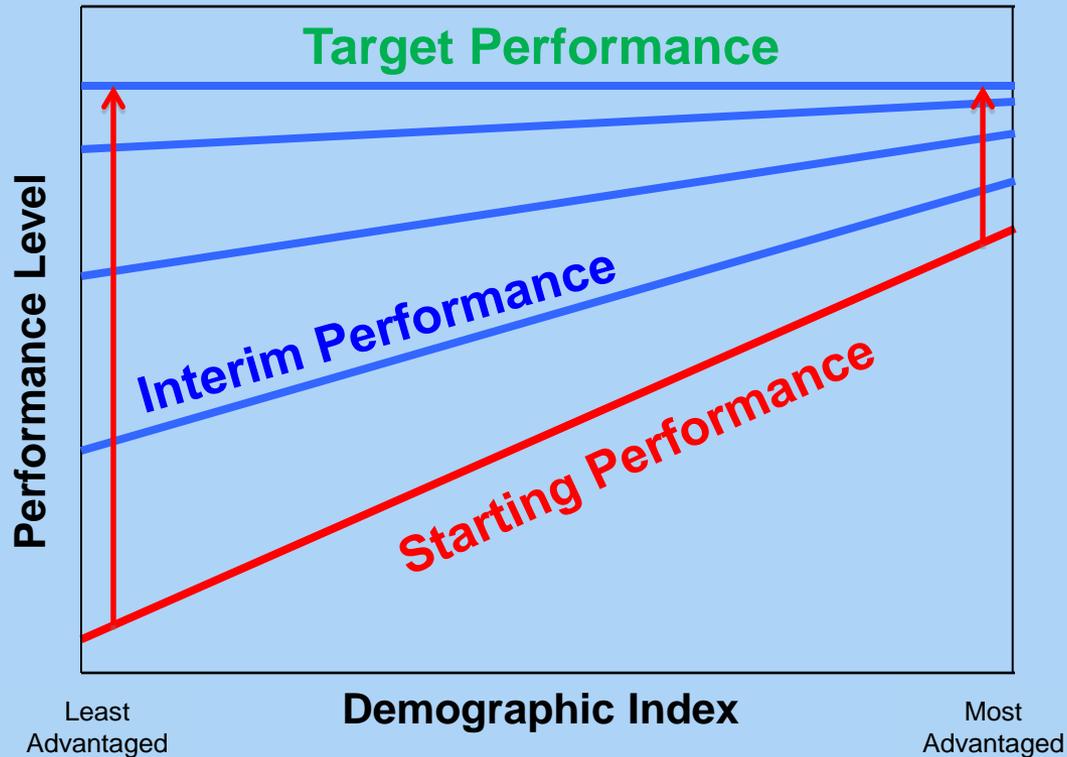
Blue lines are possible demographic "sectors" to be used for similar comparisons

See Next Slide for "Recipe" for Overall School Improvement:

Recipe for Overall School Improvement

1. Select a variety of higher than expected scoring schools (e.g., Navajo reservation schools, southern border schools, northern schools, far east plains schools, etc.) across a range of demographics (e.g., as divided by the blue, vertical lines on the previous slide) to study for best practices.
2. Select lower performing schools to study for comparison
3. Send in one or more teams: a teaching expert observer, administrative expert observer, and a systems analysis expert observer.
4. Take sufficient time to observe and document the schools' best practices.
5. Compare the differences between highest and lower performing schools in the same demographic sectors to derive a set of best practices for each demographic and similar group.
6. Apply the best practices and periodically re-observe as applicable.

WHAT ARE THE ULTIMATE GOALS?



- To lift the disadvantaged demographic end so that performance is minimally dependent on demographics and any other factors
 - ✓ We predict this provides a path to help close the Achievement Gap
- To raise total performance so that all students perform to their potential



TEACHER MERIT EVALUATIONS

The Portion Based Only on Student
Performance – 50% of the Evaluation

TEACHER MERIT EVALUATIONS

(The Portion Based Only on Student Performance – 50% of the Evaluation)

- Problem—under the previous NM evaluation system, 99% of teachers were rated effective, or above, and the public believes that poor performing teachers are difficult to remove from the classroom. (The general perception: **Do Something!**)
- **The U.S. Dept. of Education requires the basic provision to use teacher performance based on students' growth to receive a waiver from NCLB.**
- The NM PED developed the details of how performance is determined within the Federal Dept. of Education guidelines.

TEACHER MERIT EVALUATIONS

- 50% of a teacher's evaluation is based their students' performance **growth**.
- **The evaluation assumes that teachers are the ONLY cause for student performance variations** other than demographics. (It does remove students' demographic effects.)
- Many teachers are graded on End of Course (EoC) tests that are **not professionally created to use as a standardized test**.
- Some teachers are graded on the basis of what different teachers did in **different** subjects.

COMPARING TEACHERS TO TEACHERS

FEATURES AND CONCLUSIONS

The hidden assumption: for this VAM approach, only teachers control how well a student is performing. The inescapable conclusion: An average teacher's ranking is determined primarily by the previous two teachers' performance.

Teacher 1 (Two Years Ago) Student's Perf.	Teacher 2 (One Year Ago) Student's Perf.	Me (This Year) Student's Perf.	My PED Performance (I appear High relative to two Lows, etc.)
L	L	M	H
M	L	M	MH
H	L	M	M
L	M	M	MH
M	M	M	M
H	M	M	ML
L	H	M	M
M	H	M	ML
H	H	M	L

L = Low Perf., M = Median Perf., H = High Perf

This is simply not fair or reasonable!

**PREDICTION BASED ON THE PREVIOUS
SIMPLE MODEL
(BEFORE ANY GRADES WERE HANDED OUT)**

***This year, most teachers (3 quarters or so) will
still be scored as “Effective.”***

STATE NMSBA SCORES PER GRADE FOR THE LAST 3 YEARS



Given the rules for calculating teacher performance using growth:

- A 6th grade teacher in math will consistently score *low* on performance evaluation based on “growth.”
- An 8th grade reading and math teacher will consistently score *high* on performance evaluation based on “growth.”
- The 11th grade scores are consistently the lowest – except in 2014. Perhaps the students “cared” more this year because they must score above proficient to graduate the next year? (Just a guess)

A VERY IMPORTANT OUTSIDE CONCLUSION

A statement by the *American Statistical Association* dated April 8, 2014* concludes:

***“VAMs should be viewed within the context of quality improvements, which distinguishes aspects of quality that can be attributed to the system from those that can be attributed to individual teachers, teacher preparation programs, or schools. Most VAM studies find that teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions.*”**

Ranking teachers by their VAM scores can have unintended consequences that reduce quality.”

*ASA Statement on Using Value-Added-Models for Educational Assessment
https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

ASSERTIONS

- **There probably is no good way to measure any but the best and worst of teachers' performances using student performance growth.**
- **A 50% weighting of a bad measure may provide an even worse result.**

The best way to evaluate any professional is through good observation. This has been demonstrated in almost all professional organizations.

OVERALL CONCLUSIONS

- **Look more inward than outward for solutions to raise NM education results.**
- **To date, NM student performance has not improved significantly over the last 6 or 7 years. Actually, it has probably not improved significantly over the last 30 years.**
- **The ABCDF Act needs to be modified or recast to provide information that educators can use to help them improve.**
- **CESE has a method we believe will provide a way to improve performance.**
- **That portion of teacher evaluation based on student growth is almost certainly not going to cause improvement, help teachers improve, or provide accurate assessment of most teachers' performance. Until the requirement for this goes away, we must minimize the impact.**