



# The BEACON

## *News from The Coalition for Excellence in Science and Math Education*

Volume XX, No. 4 *Queries? email Rebecca Reiss (next page)*

Copyright © Oct 2018

**In this issue: President's Message: What are we Looking for in our Next Secretary of Education? pgs. 1-2. Data Show New Mexico's School Grading and Teacher Evaluation System Needs Systemic Change. pgs. 2-7. Editor's notes, pg. 7.**

### **PRESIDENT'S MESSAGE: Ken Whiton**

#### **What are we Looking for in our Next Secretary of Education?**

Before I go any further, let me assure you that The Coalition For Excellence In Science And Math Education (CESE) is a 501c(3) entity and is therefore forbidden from taking any partisan political position. In this writer's estimation, both parties bear responsibility for our failures in education. Both parties have done us wrong.

It is important to point out that, in spite of all the turmoil we've been through for almost eight years, New Mexico's education system still ranks at, or near, the bottom of all the states in educational achievement. Any honest assessment of the leadership of our Public Education Department would result in a letter grade of "F".

We can do better – we must do better.

But rather than focus on what we don't want, let's "build" an ideal Secretary of Education, someone who has, as my Mom used to say, "their feet on the floor." This person must:

- 1) Be a true leader, someone who will inspire, validate, encourage and listen to all our educators – teachers, school counselors, school staff, principals, administrators and district leadership - with a vision of what we can be.
- 2) Defend public education. According to several sources including the Council for American Private Education, approximately 90% of students in America attend public schools. Those students, parents and educators deserve strong support and advocacy.
- 3) Lead by example. Be a graduate of an accredited College of Education with a minimum of a Masters Degree and a 3.0 or better GPA. Most teachers have a Masters Degree and many have PhDs. Anyone expecting to lead an educational system must have those minimal credentials as a scholar and be a lifelong learner.
- 4) Have a minimum of two graduate-level classes in statistics receiving a grade of B or better and at least one in a science lab that shows appropriate usage of statistics. In today's data-driven world this knowledge is essential for fully understanding how to evaluate what works and doesn't work in any teacher evaluation and school grading system. These tasks cannot be understood without a basic knowledge of statistics and how to use them. It's not enough to delegate this responsibility or to just claim to be data-driven. One must understand what constitutes real data.
- 5) Foster respect for science. Our students need the strongest background in all the sciences that New Mexico, a constant leader in scientific endeavors around the world, can provide. This means there is no room for pseudoscience, watered-down science, or science tainted by anyone's personal belief system.
- 6) Have a minimum of five years experience teaching in a public school classroom. This is the only way to gain the skills necessary to successfully fill any administrative position. This writer and his family have 65+ years of experience teaching in public schools, mostly working with some of the least advantaged students. One of the many things we have learned from our personal experience is that the best teachers usually make the best administrators. On the most basic level we have discovered that successfully managing a classroom and successfully managing adults both require the same people skills, and one of the most basic of those is treating everyone with respect.

*(Continued page 2)*

<http://www.cese.org>

The Beacon is published by the Coalition for Excellence in Science and Math Education (CESE). A 501(c)(3) nonprofit corporation, we are incorporated in the State of New Mexico. Visit our website at <http://www.cese.org>.

**BOARD OF DIRECTORS**  
**PRESIDENT**  
 Ken Whiton  
[kwhiton43@icloud.com](mailto:kwhiton43@icloud.com)

**VICE PRESIDENT/PRES. ELECT**  
 Lisa Durkin  
[durkinteacher@gmail.com](mailto:durkinteacher@gmail.com)

**SECRETARY AND WEBMASTER**  
 Dave Thomas  
[nmsrdave@swcp.com](mailto:nmsrdave@swcp.com)

**TREASURER**  
 Steve Brugge  
[s.brugge@yahoo.com](mailto:s.brugge@yahoo.com)

**PAST PRESIDENT**  
 Jesse Johnson  
[garand555@comcast.net](mailto:garand555@comcast.net)

**MEMBERS AT LARGE**

Dr. Jessica Apgar  
[dr.jessica.apgar@gmail.com](mailto:dr.jessica.apgar@gmail.com)

Cindy Chapman  
[HARRISB609@aol.com](mailto:HARRISB609@aol.com)

Dr. Terry Dunbar  
[dunbarabq@gmail.com](mailto:dunbarabq@gmail.com)

Jack Jekowski  
[JPJekowski@aol.com](mailto:JPJekowski@aol.com)

M. Kim Johnson  
[kimber@comcast.net](mailto:kimber@comcast.net)

Dr. Rebecca Reiss  
[beetle@zianet.com](mailto:beetle@zianet.com)

Marilyn Savitt-Kring  
[marilynsavitt-kring@comcast.net](mailto:marilynsavitt-kring@comcast.net)

**BOARD MEMBER EMERITUS**  
 Jerry Shelton

CESE annual dues are \$25 for individual, \$35 for family, and \$10 for students. Please see page 7 for membership form.

E-mail Beacon submissions to Editor, Rebecca Reiss, [beetle@zianet.com](mailto:beetle@zianet.com)

7) Believe that an important part of “treating everyone with respect,” involves transparency. Secrecy is out. Transparency is in. One of the principles of 12-step programs is, “You are as sick as your secrets.”

8) Understand why teachers and principals are leaving our state or leaving education entirely and work to find remedies.

9) Be recognized by students, parents, fellow teachers and administrators, as being a superior teacher and leader. This recognition goes far beyond student test scores and teacher evaluation schemes. It goes to the heart of the matter: is this person a great teacher? Can this person lead others?

On January 1, 2019 New Mexico will have a new governor. Whichever candidate wins, we hope this person will be ready, equipped and willing to give New Mexico’s children, parents and educators a fresh start with a new Secretary of Education. We hope our new governor will diligently search for a qualified candidate who will, beginning on Day One, listen to teachers, school counselors, principals and school administrators and begin a meaningful dialogue that will help all of our educators feel valued, respected and encouraged.

These actions would go a long way toward improving the morale in public schools that has been steadily slipping over the last several years and give our students the skills they need to survive and succeed in today’s rapidly changing world.

---

## **Data Show New Mexico's School Grading and Teacher Evaluation System Needs Systemic Change**

### **Introduction**

Sometimes it is best to let data speak for itself. In this article, we present graphs of data concerning the New Mexico public schools that indicate serious problems with the current methods of teacher and school evaluation. These metrics impact the lives of teachers, principals, students, parents, school superintendents, and other education personnel.

The complexities of evaluating schools and teachers and the efficacy of the outcomes become apparent if one looks at the associated data. These data are either inputs or outputs of assessments using multiple, dissimilar input variables that are often controversial, especially regarding teacher and school assessments.

For New Mexico’s teacher evaluation model and school grading model (ABCDF), different elements are used with different weightings to derive a single grade or evaluation score. But how does one combine attendance with a test score to derive a single, useful evaluation score? Can this be done so that it is useful in the sense that the single output, e.g., a letter grade, has real meaning that can be used to classify or provide useful feedback

for improvement?

Our goal is to demonstrate that the data show that there needs be systemic change in the way school grades and teacher evaluations are determined. In the first example, below, we show the problems with combining two very different test types, as has been done by the New Mexico Public Education Department (PED), that cause erroneous data inputs to assessment calculations. Next, we consider what the standardized test data do show us. In the third example, we expose flaws in the combined teacher evaluation and school grading models that lead to serious inconsistencies.

**Example 1: Comparing Results from NMSBA and PARCC Testing**

Figure 1 shows the effect of the shift from the New Mexico Standards Based Assessment (NMSBA) standardized test to the PARCC (Partnership for Assessment of Readiness for College and Careers) test in the 2014-2015 schools year (SY). The percent proficient decreased for both math (greater than a factor of two) and for Reading/English Language Arts (ELA) (about a factor of 1.5) when the changeover occurred. This means that “proficient” for the NMSBA is not the same as “proficient” for PARCC. How is that possible?

What were the effects on those items that depended on standardized test proficiency scores?

The two tests are based on different standards, and because of this, the tests are very different in terms of specific materials, learning methods, and emphases. They can be mathematically “mapped” to one another, as can any two linearly scaled tests with an anchor point (proficiency) at the midpoint of the scale scores. But when one test is based on one set of standards and the other is based on totally different standards, they cannot be linked. The general material covered may be the same, but the specifics of what is tested and how it is tested are very different. Additionally, the PARCC test is primarily aimed at college-tracked students unlike the NMSBA test. These tests use different formats for multiple choice, open ended, and multiple choice asking for specific evidence. PARCC includes the display of writing skills. Different standards, different test types, and different modes of taking the test—all lacking commonality. It is not at all reasonable to think that they could ever be related in any rational, mathematically sound mapping. This sheds significant doubt on the whole process that requires combining the two for the value-added models (VAMs), which require continuity in test output form.

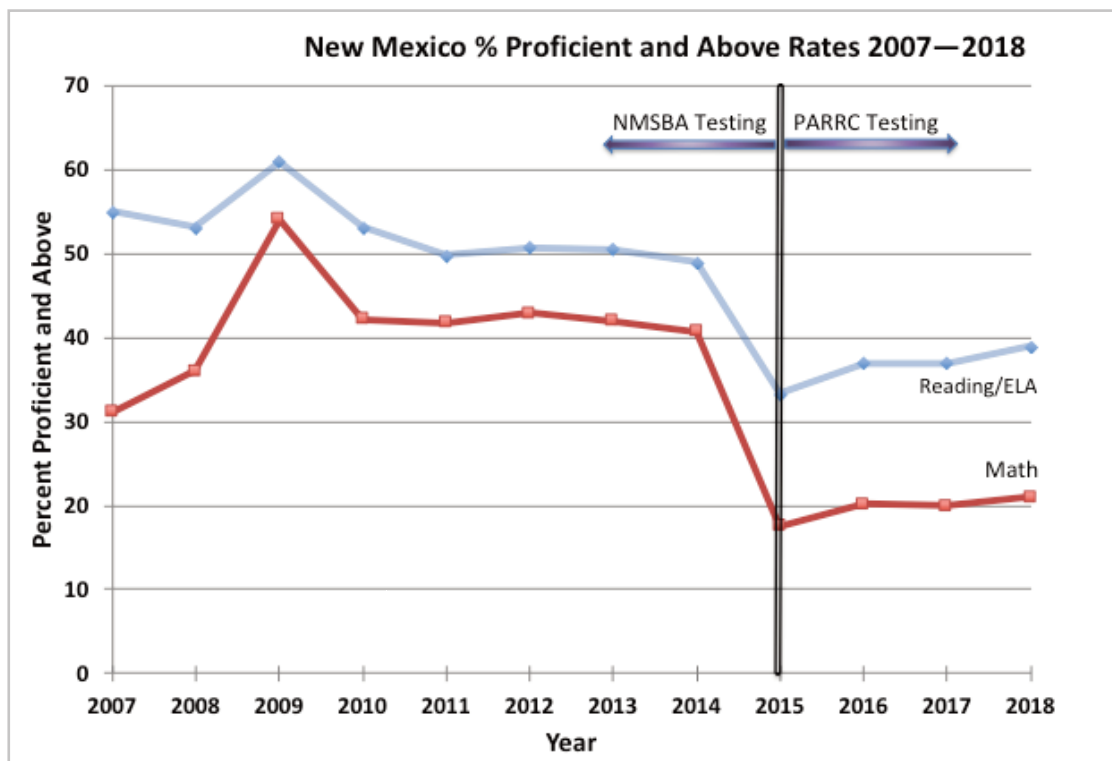


Figure 1. A comparison of NMSBA and PARCC percent proficient and above show the significant differences between Math and ELA scores..

<http://www.cese.org>

The New Mexico Public Education Department (PED) did perform a mathematical mapping that makes the NMSBA and PARCC look the same, but that is nothing more than a straight line on a piece of paper. Using an analogy, it is straightforward to map a 0 to 80-point science test to a 0 to 200-point social studies test, but there is no meaning to the mapping. Figure 1 indicates there may be a possible problem with making different “proficiencies” for different tests appear to be the same, if for no other reason than the scoring for proficient percentage students passing is different by about a factor of two. We know that the two tests measure different aspects of the same subjects, and that they have been mapped. But the outcomes are very different for the one defined anchor point; percent proficient. What is the impact on students, school grades, and teacher evaluations based on the results shown in Figure 1?

During the changeover year to PARCC and for the next two years, the NMSBA scores were used for some of the VAMs to predict student performance for teacher evaluations and for schools’ report cards grades. Since the NMSBA and PARCC scores could not be properly linked (or aligned) and since there was a large disparity between the NMSBA and PARCC proficiency results from 2014 to 2015 (Figure 1), there was an impact. Proper linking means that two different tests will provide almost the same results when given to the same or very similar students. In this case, the proficiency percentages should have approximately matched from one test to the other. They did not do so. The specific quantitative impact from the changeover over the three years is not known precisely, but the impacts are non-trivial. There are impacts that will extend into the future. For example, the NMSBA is still used for Spanish speakers in grades 3 through 11. Since the NMSBA and the PARCC tests are different, the results for Spanish speakers cannot be aligned with other students.

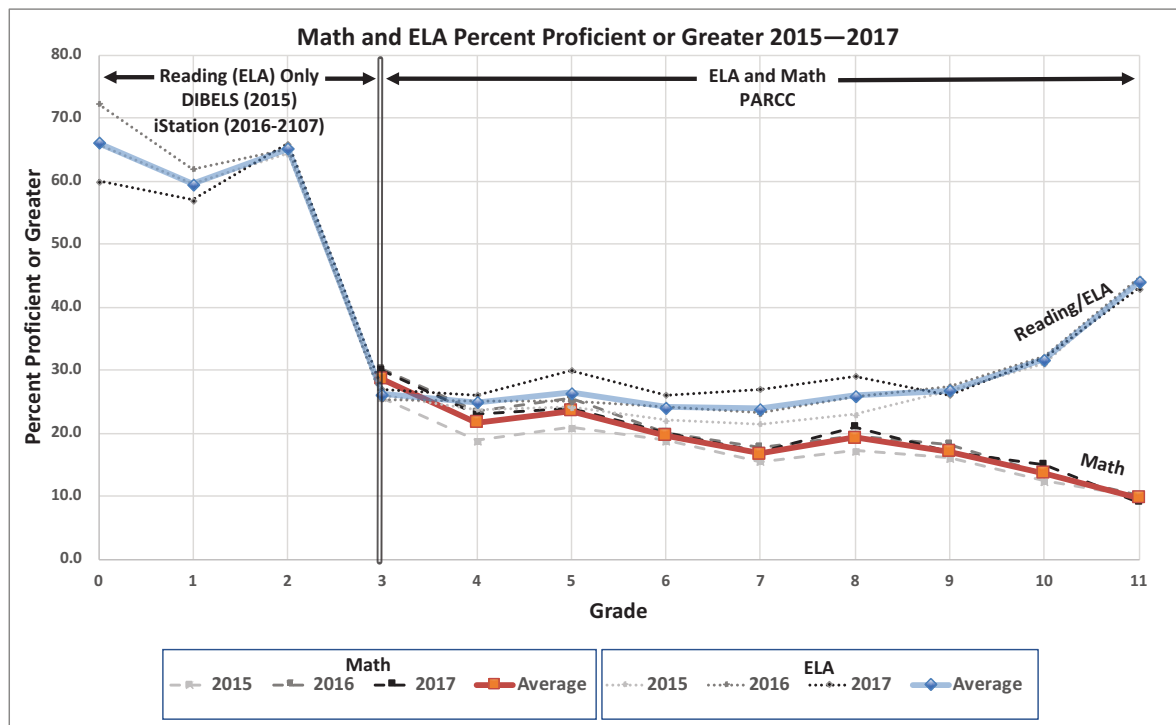
Additionally, 11th grade students’ major hurdle for graduating is to pass the PARCC test at a proficient level. However, the PED lowered the PARCC proficient level for graduation for 11th graders to passing only the 3rd of 5 scoring levels of PARCC, which is less than the proficient level, level 4. This changes in 2020 to the actual designated proficient level 4. This was apparently done to keep the graduation levels up at to at least the

same vicinity they were at the year before. One would logically think that the 2019-2020, graduation rates would decrease. However, there are alternate paths to graduation other than passing PARCC in the 11th grade, so we shall see.

### ***Example 2: What Do the Current PED Mandated Standardized Tests Show?***

Figure 2 shows what has happened in the last three years as far as iStation and PARCC trending for kindergarten (‘0’ on the figure) through 11th grades. The plot shows the average of the three years, which makes the plot easier to read and accurately shows the trending we discuss in the following. The maximum variation between scores for each grade is less than 10%, except for the incoming kindergarteners, where the maximum variation is about 12%, but decreases to 5% by the 1st grade. The iStation test is given through the 2nd grade and is for reading only. It is used in school grading and teacher evaluations as both a math and reading score. Scores are used diagnostically and for predicting both math and reading into the 3rd and 4th grades. And, yes, there are almost certainly potential problems with that simplistic approach. If 2nd grade reading from a test loosely correlated to PARCC is used to predict the next year’s math score, error will be introduced. A student’s reading ability may or may not correlate with math ability, though as we can see from Figure 2, the average reading and math proficiencies according to PARCC scores are within about 20 to 25% of each other in the third grade. The assumption that reading scores can predict math scores introduces error directly into the school grade and teacher evaluations for math teachers in particular. There are other considerations, but this one point is sufficient to show there are almost certainly errors introduced when using the iStation test score to predict math scores. Note that this says nothing about the fundamental problems with grading and teacher evaluation models that these tests drive.

The Dibels test actually provides a fairly close match to the iStation test. That is a reasonable goal, although the two tests are different. Both generally measure the same attributes for the lower grades. However, in the 3rd grade, the proficiency or above percentages drop by a factor greater than two. This is significant and, unlike the NMSBA to PARCC mapping, which had little to no impact on grades or teacher evaluations after



**Figure 2. iStation (2016 and forward), Dibels (2015), and PARCC test results show the mismatch between grades 2 and 3 (drop in proficient percentage) and the downward trend in math.**

2017, the Dibels to iStation to PARCC change is an ongoing source of potential predictive error. That is, usually, the PED maps different tests to each other using a 200-point scale, where the mapping is linear with the score of proficient being at the midpoint of the range for the test being mapped. However, when the tests are not written to the same standards, there will not be a viable transition from one test to another. This is exemplified by the drop in Figure 1 going from NMSBA to PARCC and in Figure 2 going from iStation to PARCC. However, iStation (and Dibels) testing is not used in quite the same way as was the case in going from the NMSBA to PARCC. Rather it is used as a VAM predictor for both math and reading. This is where error is introduced into any model using a VAM to predict and compare next year's scores to the predictions.

The state average is fairly level through the 8th grade for the PARCC test by itself; for grade 3 through 11, and the ELA scores are only slightly higher than math. At this point the ELA diverges from the math. After the 8th grade the divergence is a factor of four with math proficiency at 10% and ELA proficiency at 40%. There is extensive and reliable anecdotal evidence that students try harder to pass the PARCC tests in the 11th

grade because they understand it is mandated by law that they pass a graduation competency test. Passing tests at the PED specified performance level (currently 'Approaching Proficient' as opposed to the 'Proficient' level) in the 11th grade is the major hurdle for graduation. It is clear from Figure 2 that this extra effort may help for ELA, but not for math. The decline in math scores from the 8th grade through to the 11th grade indicates that motivation (pressure?) to pass PARCC math does not help students. There are a number of possible reasons that the math results are so much lower than the ELA results. In those same informal surveys, many students say the math is just too hard. Are NM students resigned to the myth that they can't do math? This is an important question for which testable hypotheses can be developed. If PARCC scoring really reflects student understanding, we need to know why this low math performance exists.

PARCC is a harder test than the NMSBA according to testing results, but some of this may be the result of teachers' lack of proper training for the Common Core State Standards. At least that is one hypothesis based on comments from many educators. Another potential contributor to low math scores may be the lack of qualified math teachers. Or we may not have addressed

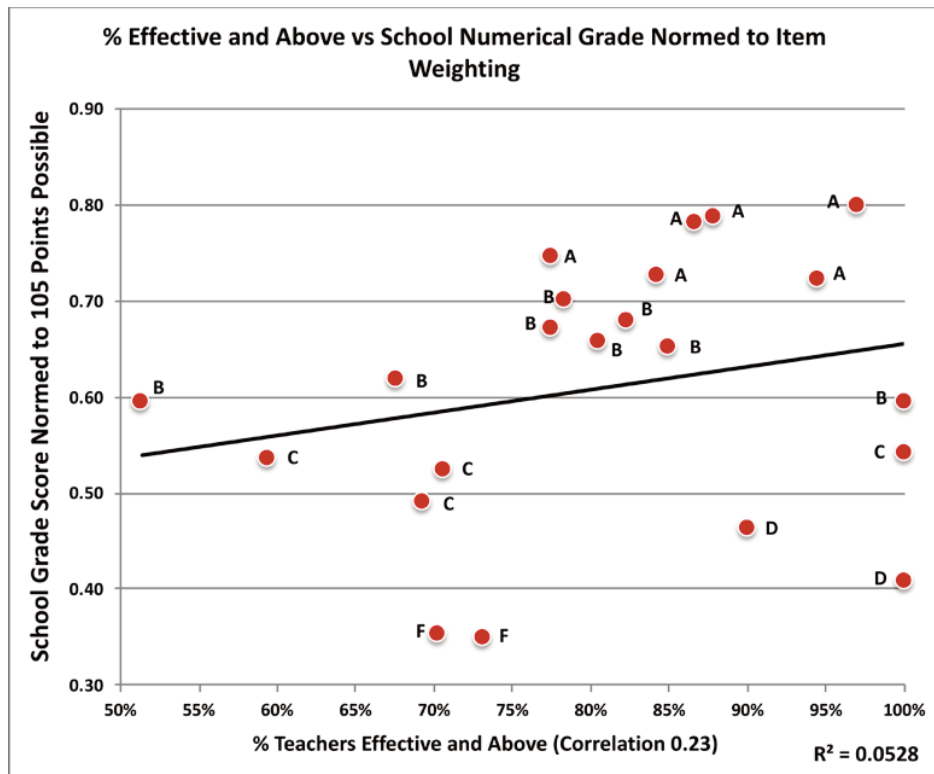
significant learning problems associated with the large numbers of New Mexico students living in low socio-economic conditions. Regardless, considering the motivation in high school for students to pass the PARCC test in math because of the graduation requirement, the state simply is not getting the job done.

**Example 3: How Well Do PED Teacher Evaluations of Effective or Above Rankings Align with PED ABCDF Grades?**

Figure 3 is a plot of grades for 2017 versus the percentage of teachers in each selected school who received an effective or above ranking by the PED within the school. These data are particularly telling regarding the efficacy of either or both the PED ABCDF grading models and the PED teacher evaluation models. Each marker represents one school; elementary, middle, and high schools. The letter by the school is the grade assigned, and the scores on the y-axis are the points given to each school on the PED report cards. The grade scales for elementary and middle schools are the same, but high schools use a slightly different scale.

Looking at the schools receiving a C, the effective or above percentage for teachers ranges from 58% to 100%. The schools receiving a B show teacher effectiveness or higher ranging from 53% to 100%. The A schools range from 76% to 97%. Looking at the data vertically, we note that A schools have effective or higher teachers ranging from 78% compared to 100% for B and C schools.

One of the most instructive things we can do is to analyze the predictive ability of the teacher effectiveness to predict a school's grade. Note the black, straight line on the graph. This is a "best" statistical fit of a line that predicts the y-axis score from the x-axis percentage. There are two values to note on the x-axis, the correlation and "R<sup>2</sup>," the coefficient of determination. The correlation is 0.23, which is considered to be very weak by statisticians. The coefficient of determination is more informative. It is a metric that indicates how much a value on the x-axis can be trusted to predict a corresponding value on the y-axis. R<sup>2</sup> varies between 0.00 and 1.00, and a value of 1.00 indicates the predictive value is 100%. This means that all the schools would fall on the regression line



**Figure 3. Comparison of grades of selected schools to the percentage of teachers with effective or higher ranking in each school shows that teacher effectiveness is a poor predictor of school grades.**

(solid black line on Figure 3). The  $R^2$  value for these data is 0.05, indicating that predictive value of whether or not a teacher is effective or higher compared to the school's grade is 5%. That is, there is no predictive value. Teachers' evaluation scores are not predicting schools' report card scores. Note that we can reverse the predication by placing the grades on the x-axis and the percent of teachers evaluated at effective or higher on the y-axis. That is, we could try to predict a school's average, teacher performance from its grade, but the results would be the same. A school's grade does not predict the PED's teacher performance evaluation rating, nor does a school's grade predict its teachers' evaluation ranking.

### Summary

Even though sampling of schools and teachers in Figure 3 does not include all schools in NM, it does show that there are sufficient contradictions between teacher rankings and school grades such that this system of grading and evaluating teachers does not work. In fact, the percentage of effective teachers or above does not predict a school's PED grade outcome. Some might argue that it is just the teacher ranking that is wrong. Others argue that the school grading is wrong. But, in fact, we have covered the reasons why neither school grades nor teacher evaluations are appropriate as implemented<sup>1,2</sup>. Even the architect of this model, Dr. Pete Goldschmidt<sup>3</sup> indicated that this model of teacher evaluation should not be used.

There are well-documented reasons to question the utility of the current school grading system and teacher

evaluation system. There has been no real progress in getting New Mexico out of the cellar when compared to other states since their inception. Is doing the same thing over and over again over the last 6 to 7 years while getting almost no difference in results indicative of a sane school grading and teacher evaluation system that help schools and teachers improve?

### References

1. The Beacon, Volume XXVII, No 1, January 2015—<http://www.cese.org/wp-content/uploads/2015/02/2015-01-Beacon>.
2. The Beacon Volume XX, No. 1, May 2017 <http://www.cese.org/wp-content/uploads/2017/05/2017-05-Beacon.pdf>, page 2 and page 5
3. Martinez, J. F., Schweig, J., and Goldschmidt, P. "Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy," Educational Evaluation and Policy Analysis December 2016, Vol. 38, No. 4, pp. 738–756 DOI: 10.3102/0162373716666166 © 2016 AERA. <http://eepa.aera.net>

**Editor's Notes:** We are saddened by the passing of Marvin Moss in August. Marvin was a founding member of CESE and a long-time board member. His extensive legacy in NM includes the establishment of the first Montessori school in Albuquerque and the Explora museum.

Please stay tuned to the Beacon in 2019. CESE is taking on the challenge of finding data-driven solutions to repair NM's public education system. Our next report will be published in January.

**Membership Dues/Donation Form**  
**Coalition for Excellence in Science and Math Education (CESE)**

**501(c)(3) non-profit, tax deductible**

**Dues and Donations cheerfully accepted year round**

**(Expiration date is found on address label)**

Member \$25, Family \$35, Student \$10. Lifetime: Individual \$500, Family \$750

You may contribute through United Way, PayPal, or USPS.

Mail checks to CESE, 803 Maverick Trail SE, Albuquerque NM 87123.

New Membership [ ]      Renewal [ ] (Please indicate any changes for renewing members)      Donation [ ]

Name \_\_\_\_\_ Date \_\_\_\_\_

Profession and/or affiliation(s) \_\_\_\_\_

Mailing Address \_\_\_\_\_

Phone \_\_\_\_\_ Cell \_\_\_\_\_ Fax \_\_\_\_\_

E-mail \_\_\_\_\_ (Most of our communication is by E-mail).

Please send change of address to Dave Thomas <nmsrdave@swcp.com>

<http://www.cese.org>

Coalition for Excellence in Science and Math Education  
803 Maverick Trail SE  
Albuquerque, NM 87123-4308

**Return Service Requested**



A Toon by Thomas



The Gathering Storm...