



Analysis of the New Mexico School Grading Technical Guide

Presented to the Legislative Education Study Committee

Coalition for Excellence in Science and Math Education

August 2012

M. Kim Johnson
kimber@comcast.net

Lisa Durkin
earthnskynlight@msn.com

Walt Murfin
murf2345@aol.com

Background

- CESE, a non-profit, non-partisan 501(c)3 corporation, performed this study
 - Members include Sandia National Laboratory retirees, an industry physicist, and a high school science teachers
- CESE has analyzed New Mexico public education data for over a decade
 - We have categorized school performance, in terms of comparison to standards and with respect to demographically neutral expectations
- CESE's primary focus is to help schools improve

Approach

- CESE was requested to use its expertise in the field of statistics and related numerical analysis to Provide feedback on:
 - The ability of schools to effectively use the Public Education Department (PED) publication: “New Mexico, School Grading Technical Guide, Calculation and Business Rules”
 - Replication of school grade calculations by experts at math/statistical modeling.
- We have also estimated how well a non-expert in data analysis could use the publication.
 - We simply asked our science teacher to act in this role as a proxy for a typical district superintendent, principal, etc.
- Our conclusions are based upon our direct experience in working with a number of schools and districts in New Mexico

Caveats and General Comments

- CESE did not have all of the data that PED uses
 - Exact replication of results was therefore not possible and not attempted
- If the grade calculations were formulaic in nature, we simply assessed our ability to perform them
- We have provided comments where improvement in the technical manual appears to be needed

Comments/Questions

- Grades need to be vetted with a knowledgeable individual from the school district– i.e. a “sanity check” before being publicized
 - One error can propagate through the entire calculation process for all the growth areas
- Manual Section VI.E.2 through VI.I lacks clear and/or consistent definitions
 - The mathematical knowledge and software required to perform the calculations are too complex for most school districts
- The calculation for “growth” requires an expert – and most probably, custom software not available to school districts (even power users of Excel may not be able to perform these calculations).
- Will the entire database used for Sections VI.E.2 through VI.I be made available to each district?
- Will the software used to perform the grade calculation be provided to schools districts with training?
- Some specific issues:
 - Section I.I School Growth. “Cohorts – Mixed effects is inconsistent with the SAS Manual Section 3.2
 - Section I.J
 - ✓ What is “Panel?”
 - ✓ The definition of “growth” is inconsistent with the ABCDF Act NMSA (1978) 22.2E-2. Definitions, and 22-2E-4. Annual Ratings
 - There is no way to tell a before/after school by code number for reorganized schools

Comments/Questions (continued)

- Some specific issues (continued):
 - Data Validation – FAY must be consistent with data in School Accountability Reports
 - Where are the formulae in E.2 and F? Where are the definitions?
 - VI.F: multiple questions
 - ✓ It appears there is a notation change from equation (1) to equation (2) – perhaps
 - ✓ Mixed models for school evaluation seem inconsistent
 - In general, there is inadequate explanation in the PED manual about the growth model
 - ✓ It is not clear it is consistent with the ABCDF Act
 - The Table of Point Boundaries is not at all explained
 - ✓ How were they derived?
 - ✓ Were the lowest and highest performers accounted for?
 - In General, once all the details are derived, they are put together according to another set of rules to arrive at a grade
 - How can one de-convolve the effects of demographics (VI.F, G, H, and I) and then justify putting them back into the mix to form a part of the grade ascribed to a given school?
 - Expanding on this, adding the effects of demographically neutral growth to the performance against standards is something like adding oranges and cows to derive pickup trucks. The result is not obviously meaningful.
 - ✓ Each serves a separate purpose, and cannot be simply added together.
- For ALL the factors in the manual, we have no evidence that they are weighted according to any set of criteria other than “this seems OK”

Comments/Questions (concluded)

- The PED uses a VAM model to derive a “Conditional score (aka “Residual” as per the CESE method of calculating actual score minus expected score.
 - The reason to look at residuals is to tell which school (or teachers) are performing above expectations based on removing demographic effects that are uncontrolled by the school or a teacher
 - Schools that outperform calculated expectations should be studied for best practices, for a given demographic classification, to apply to other schools with similar demographics
- Factors involving performance compared to a standard tell a school where it stands regarding its overall test performance, which is useful in a different way
- The addition, in any manner of a conditional score (demographics removed) with an unconditioned score makes no mathematical sense as far as we have been able to determine.

Results

Manual Calculations Section	Expert Ability to Replicate	Estimated Non-Expert Ability to Replicate
VI. B. Participation	Yes	Yes
VI. C. Attendance	Yes	Yes
VI. D. Graduation	Yes	Yes
VI. E Current Standing	Difficult, but Probable with Qualifications	No
VI. F. School Growth	Difficult, but Probable with Qualifications	No
VI. G. Student Growth	Difficult, but Probable with Qualifications	No
VI. H. Highest Quartile Student Growth	Yes, if definition is clarified	No
VI. I. Lowest Quartile Student Growth	Yes, if definition is clarified	No
VI. J. Opportunity to Learn	Yes	Yes
VI. K. College and Career Readiness	Yes	Yes

Specific Recommendations

- The existing PED manual needs substantive improvement
 - It should be peer reviewed by knowledgeable, independent educational statisticians
 - The methodology should be clearly and completely defined, and all data and software should be provided to districts provided to districts prior to grade distributions
- It would be easier for districts to use a simpler model for the current VAM used:
 - ☞ Canonical correlation model that automatically weights different values to optimize removal of demographic effects
 - ☞ Others
- The PED should not try to put everything together in one grade, e.g., demographically neutral growth and proficiency residuals do not combine with non-demographically neutral proficiency scores in a meaningful way. These are two different measures with two different outcomes.

Specific Recommendations (Concluded)

- Any factor may or may not pertain to performance outcome. Even if they do, it is not clear if they are redundant or actually add information.
 - If they do not add information, or do not pertain to outcome, then they are not appropriate to use.
- Even if the weighted factors may be appropriate, there is no indication of how the weighting was decided. There May be other options (perhaps mathematical) rather than using “best guesses” as it now appears.
- Clarification is necessary for:
 - How the ABCDF Act has been followed using the manual
 - The method for selecting and combining performance factors.

Characteristics of a Valid Grading System Not Apparent from the Manual

- The PED grading system should include the following traits:
 - A defensible, clearly defined, and more easily replicable mathematical process that districts can also use
 - A transparent, defensible process (possibly mathematical) to determine the optimum grading factors' weighting
 - The weighting factors are necessary for combining the various similar, meaningful performance output factors to derive a grade
- Without these characteristics the grading system loses meaning and doesn't necessarily help education reform