

STATE OF NEW MEXICO
LEGISLATIVE EDUCATION STUDY COMMITTEE

REPRESENTATIVES

Rick Miera, Vice Chair
Nora Espinoza
Jimmie C. Hall
Dennis J. Roch
Sheryl M. Williams Stapleton
Mimi Stewart

State Capitol North, 325 Don Gaspar, Suite 200
Santa Fe, New Mexico 87501
Phone: (505) 986-4591 Fax: (505) 986-4338
<http://www.nmlegis.gov/lcs/lesc/lescdefault.aspx>



SENATORS

John M. Sapien, Chair
Craig W. Brandt
Gay G. Kernan
Howie C. Morales

ADVISORY

Alonzo Baldonado
Nathan "Nate" Cote
George Dodge, Jr.
David M. Gallegos
Stephanie Garcia Richard
Timothy D. Lewis
Tomás E. Salazar
James E. Smith
Christine Trujillo
Bob Wooley

ADVISORY

Jacob R. Candelaria
Lee S. Cotter
Daniel A. Ivey-Soto
Linda M. Lopez
John Pinto
William P. Soules
Pat Woods

Frances Ramírez-Maestas, Director

November 13, 2013

MEMORANDUM

TO: Legislative Education Study Committee

FR: Ian Kleats

RE: STAFF BRIEF: HJM 30, *STUDY USES OF STANDARDIZED TEST SCORES*

INTRODUCTION

During the June 2013 interim meeting of the Legislative Education Study Committee (LESC), committee members discussed the passage of House Joint Memorial (HJM) 30, *Study Uses of Standardized Test Scores*, which requests that the LESC convene a work group to study the validity of using standards-based assessments for other purposes, namely teacher and school administrator effectiveness, and school grading; and that the work group report to the LESC by October 1, 2013. Rather than convene a work group, committee members requested that a report be provided to the committee from outside experts.

To introduce the topic of studying the uses of standardized test scores, this staff brief includes:

- provisions of HJM 30, *Study Uses of Standardized Test Scores*;
- related discussion; and
- background.

LESC staff have arranged presentations from:

- Ms. Mariann Lemke, Principal Research Analyst, Education Program, American Institutes for Research, "Using Assessments to Determine Teacher Performance"; and

- Dr. Pete Goldschmidt, Director, Assessment & Accountability, Public Education Department (PED), “Use of VAM in Teacher/School Leader Evaluations”.

PROVISIONS OF HJM 30

Included as an **Attachment**, the body of HJM 30, before requesting its resolved actions, makes several assertions about:

- using standardized test scores for purposes beyond their originally intended use;
- application of value-added modeling for evaluation purposes; and
- unintended consequences of evaluations tied to standardized test scores.

Using Standardized Test Scores for Purposes

With respect to standardized test scores, HJM 30 suggests that:

- standards-based assessments are designed to measure the proficiency of individual students against content standards in academic disciplines;
- the scores from these assessments can be used to provide information to teachers regarding how their students are performing on identified standards, and help parents understand the academic proficiency of their students.

However, HJM 30 further qualifies that:

- a number of factors, aside from the teachers to whom the students’ test scores would be attached, have been found to have strong influences on student learning gains, including:
 - previous teachers;
 - tutors;
 - school conditions;
 - quality of curriculum materials;
 - content-area specialists;
 - class sizes; and
 - class schedules; and
- there is broad agreement among statisticians, psychometricians, and economists that student test scores alone are not reliable and valid indicators of teacher effectiveness; and
- when assessments are used for purposes beyond their originally intended use, the results are often invalid.

Application of Value-Added Modeling for Evaluation Purposes

With respect to value-added models, HJM 30 states that:

- analyses of value-added modeling have led researchers to doubt whether the methodology can accurately identify more or less effective teachers; and

- value-added modeling estimates have proven to be unstable across statistical models, years, and classes that teachers teach.

Unintended Consequences of Evaluations Tied to Standardized Test Scores

HJM 30 suggests that using standardized test scores for evaluating teacher effectiveness might:

- lead to excessive focus on reading and mathematics, and a subsequent narrowing and over-simplification of the curriculum; and
- discourage teachers from working in public schools with the neediest students.

Based on those assertions, HJM 30 resolved that the LESC be requested to study the validity of using standards-based assessments for other purposes, principally teacher and school administrator effectiveness and school grading, for presentation during the 2013 interim; however, as noted in the introduction section of the staff brief, committee members requested that a report be provided to the committee from outside experts in lieu of a formal work group.

RELATED DISCUSSION

PED, in its bill analysis of HJM 30, suggested that student assessment scores “are a fundamental component of effectiveness evaluation systems, and research has found that student gains on standardized assessments are meaningfully related to more challenging achievement assessments, student perception surveys, expert observations of instructional practice, and assessments of teachers’ content knowledge.”

However, other research identifies some cautions in the use of standardized student assessments in evaluating teachers.

- In 2010, the Economic Policy Institute (EPI) published *Problems with the Use of Student Test Scores to Evaluate Teachers*, a briefing paper which concluded that:
 - student test scores “should be only one element among many considered in teacher profiles. Some states are now considering plans that would give as much as 50 percent of the weight in teacher evaluation and compensation decisions to scores on existing poor-quality tests of basic skills in math and reading. Based on the evidence we have reviewed above, we consider this unwise. If the quality, coverage, and design of standardized tests were to improve, some concerns would be addressed, but the serious problems of attribution and nonrandom assignment of students, as well as the practical problems described above, would still argue for serious limits on the use of test scores for teacher evaluation”; and
 - standards-based evaluations of teaching practice have been implemented in some districts and have provided more useful evidence about teaching practice. Furthermore, research indicates associations of standards-based evaluations for teachers with student achievement gains.
- In January 2013, the Bill and Melinda Gates Foundation released the results of their three-year-long study on teacher effectiveness; the Measures of Effective Teaching

(MET) Project. For the final report, *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*, the MET Project:

- used the data collected during school year 2009-2010 to build a composite measure of teaching effectiveness, combining the following three measures to predict a teacher's impact on another group of students:
 - student surveys;
 - classroom observations; and
 - a teacher's track record of student achievement gains on state tests;
- randomly assigned a classroom of students to each teacher and tracked his or her students' achievement during school year 2010-2011; and
- compared the predicted student outcomes to the actual differences that emerged by the end of school year 2010-2011.

Findings from the MET Project's final report include:

- the measures of effectiveness from school year 2009-2010 identified teachers who produced higher average student achievement following random assignment;
- as a group, the teachers identified as more effective produced greater student achievement growth than other teachers in the same school, grade, and subject;
- even though the three measures used to evaluate teacher effectiveness were collected before random assignment, these measures generated predictions of teachers' impact on students after random assignment; and
- reliable measures to identify effective teachers can be developed and that "a more balanced approach – which incorporates the student survey data and classroom observations – has two important advantages: ratings are less likely to fluctuate from year to year, and the combination is more likely to identify teachers with better outcomes on assessments other than the state tests."

Finally, the MET Foundation cautions:

- a prediction can be correct on average but still be subject to prediction error;
- anyone using these measures for high-stakes decisions should be cognizant of the possibility of error for individual teachers; and
- that they did not randomly assign students or teachers to a different school; therefore, the findings should not be used for gauging differences across schools because the process of student sorting across schools could be different than sorting between classrooms in the same school.

BACKGROUND

In 2011, the Legislature considered, but did not pass, legislation that would have implemented a new system for evaluating teachers and principals. Through executive order in the 2011 interim, the Governor created the New Mexico Effective Teaching Task Force, whose charge was to provide recommendations to the Governor regarding how best to measure the effectiveness of

teachers and school leaders based on specific parameters. Those recommendations led to legislation introduced in the 2012 session, which the Legislature considered, but did not pass.

In April 2012, the Governor issued a press release directing PED to formulate a new teacher and principal evaluation system. According to the press release, the development of a framework for a new evaluation system was one of the conditions for the *Elementary and Secondary Act* (ESEA) Flexibility Waiver from the federal *No Child Left Behind Act*, which PED had recently obtained; and the new evaluation system will incorporate many of the measures that were part of the 2012 legislation. In addition, this press release prescribed components of the system; assigned values, or weights, to those components; and presented a timeline for the development and implementation of the new evaluation system.

In May 2012, PED requested nominations for 18 people to serve two-year terms on the New Mexico Teacher Evaluation Advisory Council (NMTEACH) in order to develop the details of a new teacher and school leader evaluation system based on student achievement. In June 2012, NMTEACH held its first meeting. In July 2012, PED held a public hearing to solicit public comment on draft provisions of the new “Teacher and School Leader Effectiveness” rule. In August 2012, PED published the final draft of the “Teacher and School Leader Effectiveness” rule in the *New Mexico Register* (6.69.8 NMAC). The final rules contained several changes from the original version.

Throughout the 2012 interim, the LESC heard testimony about the evaluation of teachers and principals. This testimony raised questions about the alignment between the proposed rule and evaluation provisions already in law and about the use of student assessments, including standardized assessments, in the evaluation of teachers and principals.

At the September 2013 meeting of the LESC, the committee heard testimony from some superintendents suggesting that the timeline for implementation of the evaluation system might be occurring too quickly, noting that other initiatives were requiring district resources, including:

- implementing new curricula based on the Common Core State Standards (CCSS);
- preparing district infrastructure and personnel for a transition from the current standards-based assessment to the Partnership for Assessment of Readiness for College and Career assessments based on the CCSS; and
- understanding, explaining, and effectively utilizing data from the A-F school grading system.

According to the provisions of the ESEA Flexibility Waiver from the federal *No Child Left Behind Act*, the state must implement the new evaluation system by school year 2014-2015. However, the PED timeline implements the system during school year 2013-2014, a full year ahead of the waiver’s requirement.

A JOINT MEMORIAL

REQUESTING THE LEGISLATIVE EDUCATION STUDY COMMITTEE TO
CONVENE A WORK GROUP TO STUDY THE USE OF STANDARDIZED TEST
SCORES FOR PURPOSES OTHER THAN THOSE FOR WHICH THE TEST WAS
DESIGNED.

WHEREAS, standards-based assessments are designed to
measure the proficiency of individual students against content
standards in academic disciplines; and

WHEREAS, the use of standards-based assessment scores is
intended to provide information to teachers regarding how
their students are performing on identified standards; and

WHEREAS, individual student assessment scores are also
useful in helping parents understand the academic proficiency
of their children; and

WHEREAS, recently, New Mexico has begun implementing a
school grading system and a teacher evaluation system that
rely on an approach to measure growth using value-added
modeling; and

WHEREAS, value-added modeling uses student test scores
over several years to predict where the student will score in
a subsequent year and then attributes those gains to the
current teacher; and

WHEREAS, a number of factors have been found to have
strong influences on student learning gains, aside from the

teachers to whom their scores would be attached, and those factors include such things as previous teachers, tutors, school conditions, quality of curriculum materials, specialists, class size and schedules; and

WHEREAS, there is broad agreement among statisticians, psychometricians and economists that student test scores alone are not reliable and valid indicators of teacher effectiveness and should not be used in high-stakes personnel decisions; and

WHEREAS, analyses of value-added modeling results have led researchers to doubt whether the methodology can accurately identify more or less effective teachers; and

WHEREAS, value-added modeling estimates have proven to be unstable across statistical models, years and classes that teachers teach; and

WHEREAS, the use of student test scores to evaluate teachers leads to an excessive focus on reading and mathematics and the subsequent narrowing and oversimplification of the curriculum; and

WHEREAS, one possible unintended consequence of tying teacher evaluations and sanctions to test score results may be to discourage teachers from working in public schools with the neediest students; and

WHEREAS, when an assessment instrument, such as the standards-based assessment, is used for a purpose other than that for which it was designed, for example for measuring

teacher effectiveness, the results are often invalid;

NOW, THEREFORE, BE IT RESOLVED BY THE LEGISLATURE OF THE STATE OF NEW MEXICO that the legislative education study committee be requested to convene a work group to study the validity of using standards-based assessments for other purposes, principally teacher and school administrator effectiveness and school grading; and

BE IT FURTHER RESOLVED that the work group report to the legislative education study committee by October 1, 2013; and

BE IT FURTHER RESOLVED that a copy of this memorial be transmitted to the legislative education study committee.