

**State of New Mexico**  
**LEGISLATIVE EDUCATION STUDY COMMITTEE**

**REPRESENTATIVES**

Rick Miera, Chair  
Roberto "Bobby" J. Gonzales  
Jimmie C. Hall  
Mimi Stewart  
Thomas E. Swisstack  
W. C. "Dub" Williams

State Capitol North, 325 Don Gaspar, Suite 200  
Santa Fe, New Mexico 87501  
PH: (505) 986-4591 FAX: (505) 986-4338  
<http://lesc.nmlegis.gov>

**SENATORS**

Cynthia Nava, Vice Chair  
Vernon D. Asbill  
Mary Jane M. Garcia  
Gay G. Kernan

**ADVISORY**

Andrew J. Barreras  
Ray Begaye  
Nathan P. Cote  
Nora Espinoza  
Mary Helen Garcia  
Thomas A. Garcia  
Dianne Miller Hamilton  
John A. Heaton  
Sheryl M. Williams Stapleton  
Jim R. Trujillo  
Teresa A. Zanetti



**ADVISORY**

Mark Boitano  
Carlos R. Cisneros  
Dianna J. Duran  
Lynda M. Lovejoy  
Howie C. Morales  
John Pinto  
William E. Sharer

Frances R. Maestas, Director

December 18, 2008

**MEMORANDUM**

**TO:** Legislative Education Study Committee

**FR:** Ms. Pamela Herman *PK*

**RE: WRITTEN REPORT: *STUDY EDUCATION ACCOUNTABILITY TARGETS, SJM 43***

---

Attached is a summary of the recommendations of the 2008 Assessment and Accountability Work Group, formed in response to SJM 43, *Study Education Accountability Targets*, which was passed by the Legislature in 2008. The memorial requests that the Legislative Education Study Committee (LESC), in collaboration with the Public Education Department (PED), the Legislative Council Service, public school superintendents, directors of special education, directors of bilingual education, and other appropriate educators, form a study group to meet during the 2008 interim to discuss the possible use of certain supplements or alternatives to New Mexico accountability targets and to report findings and recommendations to the LESL by December 2008.

A copy of the final report of the Work Group, including background information, a summary of the Work Group's activities, findings and recommendations, is also provided.

**2008 LESC ASSESSMENT AND ACCOUNTABILITY WORK GROUP, SJM 43**  
**SUMMARY OF WORK GROUP RECOMMENDATIONS**

1. The Public Education Department (PED), working with the department's Assessment and Accountability Task Force, should develop for the consideration of the Legislature, a state accountability system based to the extent possible on a growth model, separate from and complementary to the existing accountability system in statute and in the New Mexico Accountability Workbook approved by the US Department of Education (USDE) pursuant to the federal *No Child Left Behind Act of 2001* (NCLB).
2. To create an appropriate complementary high school accountability system, the New Mexico College and Career Ready Policy Institute (CCRPI) Working Team should be expanded to include some members of the 2008 LESC Assessment and Accountability Work Group. The goal of the CCRPI is to continue the secondary-postsecondary alignment work of the New Mexico American Diploma Project by developing an assessment and accountability system that measures how well high schools prepare students for college and careers.
3. To create a system to measure academic growth in grades 3 through 8, the Legislative Education Study Committee (LESC) should consider endorsing a memorial to request that PED convene a one-time work group of key stakeholders, including members of the PED Assessment and Accountability Task Force, to develop a New Mexico Value Table for the proposed complementary school accountability system that will measure student academic growth in grades 3 through 8 and differentially weight students' changes in performance from year to year on state standards-based assessments.
4. The new complementary accountability system envisioned by the Work Group should:
  - a. complement the existing accountability system by providing additional information about how schools are performing, without creating new ways for a school to fail under the existing system.
  - b. reward schools that achieve desired performance targets with public acknowledgement as well as site-based flexibility or autonomy (for curriculum and other decisions) if financial rewards are not possible.
  - c. produce reports that are easily understood by school staff, parents, students, and community members, and that do not duplicate existing report cards. Further, there should be clear, effective communication with the public and schools about the purpose of the new system and how it works.
  - d. be timed and designed to leverage existing personnel and data reporting resources, so as not to overburden already strained district and state accountability resources.
  - e. continue to disaggregate reported data by key demographic groups in the school. Policymakers should remain alert to viable suggestions for ways to avoid duplicate counting of students.

- f. track and report achievement of English language proficiency within a five-year period by schools' English language learners (ELL) populations, as measured by the New Mexico English Language Proficiency Assessment (NMELPA).
  - g. include the results for all students who take alternative assessments for students with disabilities, both the assessment currently being developed by PED to be based on grade level content standards but modified performance standards, as well the New Mexico Alternate Performance Assessment (NMAPA). It should give schools credit for the actual proficiency those students demonstrate without the arbitrary caps provided in federal rule. PED should evaluate the impact of the new assessment, once implemented, to determine if other metrics for the achievement of students with disabilities should be considered in the complementary accountability system.
  - h. report results that recognize the differing needs and opportunities of students by school by creating "peer groups" of schools with similar school size and percentages of low income students, ELLs, and student mobility. A report card format should be developed that shows how each school ranks in comparison with its peer schools in Value Table calculations, including achievement of English language proficiency by ELLs and performance of students with disabilities based on alternative assessments.
  - i. recognize that special schools, such as the New Mexico School for the Deaf, should be judged by achievement of their special missions.
  - j. provide for a limited, voluntary pilot of the complementary assessment and accountability system for grades 3 through 8 to determine if it is workable and if it provides useful information to evaluate and improve the performance of public schools.
5. Regarding the existing accountability system under NCLB, the Work Group recommends the following:
- a. PED should continue to pursue approval of a growth model by USDE once NCLB has been reauthorized;
  - b. PED should consider applying to USDE for use of a differentiated accountability system, if permitted under a reauthorized NCLB; and
  - c. The LESC and the executive branch should consider continuing to participate, in coordination with organizations such as the National Conference of State Legislatures, the New Mexico Coalition of School Administrators, the New Mexico School Boards Association, the National Governors Association, and other groups, in the redesign and reauthorization of NCLB in order to eliminate the unintended negative consequences of the act as it currently stands.

**State of New Mexico**  
**LEGISLATIVE EDUCATION STUDY COMMITTEE**

**REPRESENTATIVES**

Rick Miera, Chair  
Roberto "Bobby" J. Gonzales  
Jimmie C. Hall  
Mimi Stewart  
Thomas E. Swisstack  
W. C. "Dub" Williams

State Capitol North, 325 Don Gaspar, Suite 200  
Santa Fe, New Mexico 87501  
PH: (505) 986-4591 FAX: (505) 986-4338  
<http://lesc.nmlegis.gov>

**SENATORS**

Cynthia Nava, Vice Chair  
Vernon D. Asbill  
Mary Jane M. Garcia  
Gay G. Kernan

**ADVISORY**

Andrew J. Barreras  
Ray Begaye  
Nathan P. Cote  
Nora Espinoza  
Mary Helen Garcia  
Thomas A. Garcia  
Dianne Miller Hamilton  
John A. Heaton  
Sheryl M. Williams Stapleton  
Jim R. Trujillo  
Teresa A. Zanetti



**ADVISORY**

Mark Boitano  
Carlos R. Cisneros  
Dianna J. Duran  
Lynda M. Lovejoy  
Howie C. Morales  
John Pinto  
William E. Sharer

Frances R. Maestas, Director

**FINAL REPORT OF THE**  
**2008 LESC ASSESSMENT AND ACCOUNTABILITY**  
**WORK GROUP**

IN RESPONSE TO  
***SJM 43, STUDY SCHOOL ACCOUNTABILITY TARGETS***

**LEGISLATIVE EDUCATION STUDY COMMITTEE**  
**DECEMBER 2008**

**SJM 43, STUDY SCHOOL ACCOUNTABILITY TARGETS**  
**FINAL REPORT OF THE 2008 LESC ASSESSMENT AND ACCOUNTABILITY**  
**WORK GROUP**

*The extremity of the unrealistic expectations in [the NCLB] approach becomes clearer if one imagines trying something similar in another area of public concern, say, hospital quality. We would first set standards for “sufficiently healthy outcomes,” using arbitrary and different methods in different states that yielded different answers and that were not based on any evidence about what current medical technology could produce. We would then tell all hospitals, regardless of their circumstances—for example, the age or the health status of the patients they take in, the pool of available specialists in their geographic area, the resources available to them, and so on—that they had a set time, say a dozen years, to reach the point at which all patients would be discharged “sufficiently healthy.” They would be rewarded or punished along the way on the basis of whether they were making linear progress toward this goal. It’s hard to imagine such a proposal even getting serious consideration.*

-- Daniel Koretz, Ph.D., *Measuring Up* (2008)

## **Introduction**

In 2008, the Legislature passed SJM 43, *Study School Accountability Targets* (see Attachment 1). The memorial requested that the Legislative Education Study Committee (LESC), in collaboration with the Public Education Department (PED), the Legislative Council Service, public school superintendents, directors of special education, directors of bilingual education, and other appropriate educators, form a study group to meet during the 2008 interim:

- to discuss the possible use of short-cycle assessments as an indicator of student progress;
- to recommend an accountability mechanism for specific subpopulations based on a growth model;
- to study the impact of using the federal Office of Special Education targets in lieu of New Mexico accountability targets;
- to examine opportunity-to-learn factors as a companion to New Mexico accountability targets; and
- to report its findings to the appropriate interim committee of the Legislature by December 2008.

During the 2008 interim, therefore, the LESC convened the 2008 LESC Assessment and Accountability Work Group to address the requests in the memorial (see Attachment 2). This report contains background information, a summary of the Work Group’s activities, and its findings and recommendations.

## **Background**

According to the Education Commission for the States (ECS), accountability means “holding key individuals and groups responsible for student achievement through systematic collection, analysis, use and reporting of valid and reliable information.” States started putting accountability systems in place in the 1990s, in response to the standards movement, an initiative in many states to define clearly what students should know and be able to do, to report how well

students in each school did in meeting those standards, and to establish consequences for schools based on those results. Since the start of the new millennium, the federal *No Child Left Behind Act of 2001* (NCLB) built upon those individual state efforts to require all states to implement a uniform, inclusive standards-based assessment and accountability system that meets strict federal guidelines in order to continue receiving federal Title I funds.

Because for the first time NCLB established a national system of increasingly serious sanctions for schools and school districts that fail to reach their accountability targets, over the seven years since its passage, NCLB has been credited with transforming the landscape of public education. According to ECS, the law set deadlines for states to expand the scope and frequency of student testing, revamp their accountability systems, and guarantee that every teacher in core academic classrooms meets the statutory definition of “highly qualified” in their subject areas. NCLB requires states to make and report demonstrable annual progress in increasing the percentage of students in each of several demographic categories who test proficient in reading and math, toward a goal of universal student proficiency in grades 3-8 and one high school grade by 2014.

To address the requirements of NCLB, in 2003, the Legislature passed a comprehensive LESC-endorsed school reform bill that included the *Assessment and Accountability Act*, which established a system of assessment and accountability that complies with NCLB by focusing on adequate yearly progress (AYP) as an accountability measure, using standards-based (also known as “criterion-referenced”) rather than norm-referenced assessments; and establishing a rating system for schools and school districts that fail to make AYP that tracked in most particulars the mandates of the federal law. As required by the federal law, New Mexico has an approved Accountability Workbook that sets forth in detail how its assessment and accountability system works, and as of June 2008, the state’s assessment system was fully approved by the US Department of Education (USDE).

Researchers and commentators have reported both positive and negative changes in public education since NCLB was enacted. The Center on Education Policy (CEP) has issued several reports annually regarding NCLB, its implementation and its effects, and reports the following changes in public education since passage of the law<sup>1</sup>:

- Since 2002, reading and math achievement on state tests has gone up in most states based on the percentage of students scoring at the proficient level, with gains tending to be larger in the elementary and middle school grades than in high school.
  - Gains on the National Assessment of Education Progress (NAEP), the “nation’s report card,” have shown similar trends as state tests, although the gains have tended to be smaller.
  - In New Mexico between 2005 and 2007, on state reading tests, elementary students made proficiency gains that may not have been statistically significant; middle school students made moderate-to-large proficiency gains; and high school students posted moderate-to-large losses in proficiency. In math, students made moderate-to-large proficiency gains in elementary and middle school and slight gains in high school.

---

<sup>1</sup> CEP does not, however, attribute these changes entirely to the passage of the NCLB, since at least some represent extensions of trends that pre-date the act.

- Nationally, in states with sufficient state test data to make a judgment, achievement gaps between white and minority students have narrowed more often than they have widened. NAEP results showed the same trends except in eighth grade math, where gaps have tended to widen instead of narrow.
  - In New Mexico between 2005 and 2007, gaps on state tests between Hispanic and white students narrowed in elementary and middle school reading and high school math; and widened in middle school math and high school reading. Gaps between Native American and white students narrowed in elementary reading, and widened in high school reading and math. Gaps between low-income and other students narrowed in elementary reading and math, stayed the same or widened in middle school, and narrowed in both subjects in high school.
- Schools and districts are paying much more attention to achievement gaps and the learning needs of traditionally underperforming groups of students.
- Students are spending more time taking more tests.
- A substantial majority of school districts appear to be spending more time on the subjects tested, reading and math, at the expense of other subjects, particularly social studies.
- Schools are paying more attention to aligning curriculum and instruction with standards and assessments, using test data more widely to modify instruction, using educational research more often to inform decisions about instructional strategies, increasing the quality and quantity of teacher professional development, and providing more intensive instruction and intervention for low-achieving students.
- The percentage of schools nationwide that are missing their annual accountability targets is increasing steadily. In New Mexico, according to PED, the percentage of all public schools that did not make AYP in spring 2008 was 68.2 percent, up from 54.5 percent in spring 2007, 54.1 in 2006, and 52.8 percent in 2005.

NCLB was slated for reauthorization by Congress in 2007; however, while over 100 measures have been introduced to reauthorize the act, none has been enacted. According to a presentation by a representative of the National Conference of State Legislatures to the LESC in April 2008, the act will remain in full force so long as Congress continues to appropriate funds. Changes to NCLB that have been proposed by education interest groups and members of Congress include the following:

- Modify AYP by using a “growth model” that acknowledge improvements in individual student achievement toward the goal of universal proficiency by a target deadline.
- Offer greater flexibility and allowing expanded use of alternative assessments to measure the academic achievement of students with disability and English language learners (ELLs).
- Employ “multiple factors” rather than one annual test to measure achievement.
- Develop more easily understood reporting systems for school accountability.

## Work Group Activities

The LESC Assessment and Accountability Work Group met four times in half-day or full-day sessions in Santa Fe and Albuquerque during the 2008 interim to consider the requests in SJM 43. Among the key topics that the group addressed were:

- Use of short-cycle assessments for school accountability: The Work Group agreed that while short-cycle assessments have proven valuable as tools to diagnose student needs and address the differentiated instructional needs of a diverse student population, assessment experts caution against using such assessments for purposes they were not designed to serve, such as school accountability.
- Use of the targets established for state plan accountability to the US Office of Special Education for academic achievement of students with disabilities: PED representatives explained that targets developed for statewide program accountability would be difficult or impossible to apply to individual schools for reasons of group size, among others. Further, some members of the Work Group felt that establishing different accountability targets for students with disabilities would do those students a disservice. Instead, the Work Group focused its deliberations on the question of how the achievement of those students should most appropriately be assessed.
- Need for technical assistance: At its first meeting, the Work Group agreed that given the technical nature of the topic, its discussions could be furthered with help from a nationally recognized expert in the development of state assessment and accountability systems. LESC staff approached the National Center for the Improvement of Educational Assessment (NCIEA), a nonprofit organization that works with states and other educational agencies to design and implement effective assessment and accountability policies and programs. The NCIEA has provided its services to the states of Louisiana, Vermont, Oregon, Pennsylvania, California, Nevada, Wyoming, Alaska, Massachusetts, Maryland, and Ohio. Dr. Richard Hill, the founding director and chairman of the board of NCIEA, traveled to New Mexico for a one-day presentation and discussion with the Work Group that included a framework for approaching decisions about accountability systems, and a description of the Value Table model that the Work Group recommends New Mexico adopt (see Attachment 3 and Work Group Recommendation 3).
- Consensus: The Work Group attempted as much as possible to reach its decisions by consensus of all the members present, and the recommendations below represent that consensus.

## Work Group Recommendations

Based on discussions and input at the four Assessment and Accountability Work Group meetings, the Work Group has agreed on the following recommendations to the LESC:

1. *PED, working with the department's Assessment and Accountability Task Force, should develop for the consideration of the Legislature a state accountability system based to the extent possible on a growth model, separate from and complementary to the existing accountability system in statute and in the New Mexico Accountability Workbook approved by the USDE pursuant to NCLB.*

- a. According to the Center for Assessment and the Council of Chief State School Officers (CCSSO), approximately one-third of US states, as well as some large urban school districts, have state accountability systems separate from the systems required by NCLB. These systems are designed to provide information that is not reflected in AYP reports but that are important to the state for accountability purposes, and to trigger different consequences than those provided for in federal law.
  - b. The Work Group concurred that New Mexico needs an assessment and accountability system that produces information useful to highlight schools' strengths and weaknesses and helpful in improving instruction. The Work Group agreed that the existing accountability system does not serve those purposes because it uses a flawed metric—AYP—to over-identify schools in need of assistance.
2. *To create an appropriate complementary high school accountability system, the New Mexico College and Career Ready Policy Institute<sup>2</sup> (CCRPI) Working Team should be expanded to include some members of the 2008 LESC Assessment and Accountability Work Group. The goal of the CCRPI is to continue the secondary-postsecondary alignment work of the New Mexico American Diploma Project by developing an assessment and accountability system that measures how well high schools prepare students for college and careers.*
- a. The Work Group agreed that (1) a system based on growth measures is not practicable in high school where state standards-based assessments are administered only in eleventh grade, and (2) the many facets of college and career readiness, rather than performance on one standardized test, are more appropriate accountability measures for high schools, whose mission is to prepare all students, despite their widely diverse goals, interests and future plans, for successful lives after graduation. The CCRPI Working Team will spend the next year in an effort to establish metrics and systems for high school accountability based on college and career readiness.
3. *To create a system to measure academic growth in grades 3 through 8, the LESC should consider endorsing a memorial to request that PED convene a one-time work group of key stakeholders, including members of the PED Assessment and Accountability Task Force, to develop a New Mexico Value Table for the proposed complementary school accountability system that will measure student academic growth in grades 3 through 8 and differentially weight students' changes in performance from year to year on state standards-based assessments (see Attachment 3).*
- a. The Work Group heard a presentation from a representative of the National Center for the Improvement of Educational Assessment describing a Value Table accountability model that measures schools based on changes in student assessment scores across years. According to the Center, such a system is relatively easy to understand, computationally simple, and able to incorporate all the performance levels (Basic, Nearing Proficiency, Proficient, and Advanced) in the state's existing system. The Value Table is a grid that assigns points for each change from one level to another from year to year. A Value

---

<sup>2</sup> The CCRPI is an 18-month joint initiative of Achieve, Inc., the National Governors Association Center for Best Practices, Education Counsel, Jobs for Progress, and the Data Quality Campaign to assist the eight states that were accepted to participate through a competitive process, including New Mexico, "to put an assessment and accountability system in place that fully reflects and supports the goal of all students graduating college and career-ready."

Table can be designed to recognize changes in performance not currently rewarded in the existing system. For example, it can award more points for a student who moves two levels than only one, and can also recognize students who move from Basic to Nearing Proficiency, or from Proficient to Advanced. A school's score would be based on the average number of points its students earn.

4. *The new complementary accountability system envisioned by the Work Group should:*

a. *complement the existing accountability system by providing additional information about how schools are performing, without creating new ways for a school to fail under the existing system.*

b. *reward schools that achieve desired performance targets with public acknowledgement as well as site-based flexibility or autonomy (for curriculum and other decisions) if financial rewards are not possible.*

- The Work Group recommends implementing an effective form of public recognition that extends beyond simply publishing report cards in local newspapers, to include publicity that actively involves and engages students and teachers, and that encourages like-situated schools to emulate successful approaches.

c. *produce reports that are easily understood by school staff, parents, students, and community members, and that do not duplicate existing report cards. Further, there should be clear, effective communication with the public and schools about the purpose of the new system and how it works.*

- If this recommendation is implemented, redundant reports cards, such as those currently required in statute, should be eliminated. The Work Group suggested that the data produced by the new accountability system might prove valuable in the Educational Plan for Student Success (EPSS) process by providing more accurate and nuanced information about a school's strengths and weaknesses than the current AYP system.

d. *be timed and designed to leverage existing personnel and data reporting resources, so as not to overburden already strained district and state accountability resources.*

- For example, the reports could be issued in winter or spring, off-cycle from the NCLB required AYP reports.

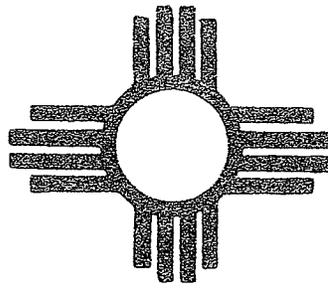
e. *continue to disaggregate reported data by key demographic groups in the school. Policymakers should remain alert to viable suggestions for ways to avoid duplicate counting of students.*

- The Work Group agreed that assessments provide their greatest value when they supply teachers and administrators with useful diagnostic information for individualized instruction, but still believed that disaggregation of achievement data by ethnic, racial and low income groups can be useful. Disaggregated data identifies achievement gaps and changes in achievement gaps, and can help focus attention on

what works or doesn't work in schools that are outliers in the performance of their traditionally underachieving populations.

- The group discussed its concerns about multiple counting of students under the current AYP system, particularly those most likely to be struggling academically (e.g., low-income minority ELL students, but was unable to develop a statistically workable recommendation to address that problem. Regarding concerns about the reported stigmatization of two groups that may cause a school to fail to make AYP – students with disabilities and ELLs – the Work Group has recommendations below to make it possible to more accurately and usefully measure the performance of these groups for state accountability purposes.
- f. *track and report achievement of English language proficiency within a five-year period by schools' ELL populations, as measured by the New Mexico English Language Proficiency Assessment (NMELPA).*
- The consensus of the Work Group is that a well-documented correlation exists in New Mexico school districts between ELL students' achieving English language proficiency and their ability to demonstrate proficiency on content area assessments. The Work Group felt that schools should be judged in part on their ability to bring ELL students to English language proficiency within a time period whose length is supported by educational research, not the three-year period currently required in federal law.
- g. *include the results for all students who take alternative assessments for students with disabilities, both the assessment currently being developed by PED to be based on grade level content standards but modified performance standards, as well the New Mexico Alternate Performance Assessment (NMAPA). It should give schools credit for the actual proficiency those students demonstrate without the arbitrary caps provided in federal rule. PED should evaluate the impact of the new assessment, once implemented, to determine if other metrics for the achievement of students with disabilities should be considered in the complementary accountability system.*
- It was the consensus of the Work Group that the most appropriate way to evaluate the achievement of students with disabilities is by measuring achievement of the goals in their individualized education plans (IEP). In many cases finding an objective way to measure such achievement can be difficult and costly. However, where the IEP specifies that a student take the NMAPA or the modified performance assessment, the complementary accountability system should report the actual results of those assessments. The results should not be limited by the 1.0 percent proficiency cap on the NMAPA and the 2.0 percent proficiency cap on the modified performance assessment currently imposed by the federal AYP rules.
- h. *report results that recognize the differing needs and opportunities of students by school by creating "peer groups" of schools with similar school size and percentages of low income students, ELLs, and student mobility. A report card format should be developed that shows how each school ranks in comparison with its peer schools in Value Table calculations, including achievement of English language proficiency by ELLs and performance of students with disabilities based on alternative assessments.*

- i. recognize that special schools, such as the New Mexico School for the Deaf, should be judged by achievement of their special missions.*
  - j. Provide for a limited, voluntary pilot of the complementary assessment and accountability system for grades 3 through 8 to determine if it is workable and if it provides useful information to evaluate and improve the performance of public schools.*
5. *Regarding the existing accountability system under NCLB, the Work Group recommends the following:*
- a. PED should continue to pursue approval of a growth model by USDE once NCLB has been reauthorized;*
  - b. PED should consider applying to USDE for use of a differentiated accountability system, if permitted under a reauthorized NCLB; and*
  - c. The LESC and the executive branch should consider continuing to participate, in coordination with organizations such as the National Conference of State Legislatures, the New Mexico Coalition of School Administrators, the New Mexico School Boards Association, the National Governors Association, and other groups, in the redesign and reauthorization of NCLB in order to eliminate the unintended negative consequences of the act as it currently stands.*



The Legislature  
of the  
State of New Mexico

48th Legislature, Second Session

LAWS 2008

CHAPTER \_\_\_\_\_

SENATE JOINT MEMORIAL 43

Introduced by

SENATOR CYNTHIA NAVA  
SENATOR GAY G. KERNAN



1 A JOINT MEMORIAL

2 REQUESTING THAT THE LEGISLATIVE EDUCATION STUDY COMMITTEE,  
3 THE PUBLIC EDUCATION DEPARTMENT AND OTHERS MEET IN THE  
4 INTERIM TO STUDY ACCOUNTABILITY TARGETS.

5  
6 WHEREAS, the people of New Mexico share a vision of a  
7 world-class education system that prepares their children for  
8 success in the twenty-first century; and

9 WHEREAS, the New Mexico legislature has laid the  
10 essential groundwork needed to support an educational system  
11 with enhanced educational opportunities for all New Mexico  
12 students; and

13 WHEREAS, public schools in New Mexico have made progress  
14 in reducing the achievement gap; and

15 WHEREAS, New Mexico ranks second in the nation for  
16 education reform, school choice and children's access to free  
17 and reduced-cost breakfast and ranks fourth in the nation for  
18 a strong nutrition policy; and

19 WHEREAS, New Mexico is one of four states leading the  
20 nation in innovative school improvement strategies; and

21 WHEREAS, New Mexico ranks in the top ten states in grade  
22 three reading comprehension and in oral reading fluency for  
23 students in grades one, two and three; and

24 WHEREAS, New Mexico is one of only nine states with an  
25 approved highly qualified teacher plan and is one of eleven

1 states to have standards-based assessments aligned to strong  
2 content standards; and

3 WHEREAS, New Mexico is recognized for standards and  
4 benchmarks that are well aligned to advanced placement and  
5 college entrance examinations; and

6 WHEREAS, the secretary of public education and the  
7 educational community assert that it is not just assessments  
8 that will reform education in New Mexico, but that schools  
9 must be accountable for performance, instruction and  
10 improvement; and

11 WHEREAS, scientific research supports the use of  
12 formative "short-cycle" assessments to inform instruction for  
13 students; and

14 WHEREAS, the public education department supports the  
15 use of short-cycle assessments to address the differentiated  
16 learning needs of a diverse student population; and

17 WHEREAS, there were four hundred forty schools that did  
18 not make adequate yearly progress pursuant to New Mexico and  
19 federal accountability standards, largely because of the  
20 performance of various subpopulations;

21 NOW, THEREFORE, BE IT RESOLVED BY THE LEGISLATURE OF THE  
22 STATE OF NEW MEXICO that the legislative education study  
23 committee, in collaboration with the public education  
24 department, the legislative council service, public school  
25 superintendents, directors of special education, directors of

1 bilingual education and other appropriate educators, form a  
2 study group to meet during the interim to discuss the  
3 possible use of short-cycle assessments as an indicator of  
4 student progress and to recommend an accountability mechanism  
5 for the specific subpopulations based on a growth model; and

6 BE IT FURTHER RESOLVED that the study group study the  
7 impact of using the office of special education targets in  
8 lieu of New Mexico accountability targets; and

9 BE IT FURTHER RESOLVED that the study group examine  
10 opportunity-to-learn factors as a companion to New Mexico  
11 accountability targets; and

12 BE IT FURTHER RESOLVED that the study group report its  
13 findings to the appropriate interim committee of the  
14 legislature by December 1, 2008; and

15 BE IT FURTHER RESOLVED that copies of this memorial be  
16 transmitted to the secretary of public education, the chair  
17 of the legislative education study committee and the director  
18 of the legislative council service. \_\_\_\_\_

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

s/Diane D. Denish

Diane D. Denish, President  
Senate

s/Lenore M. Naranjo

Lenore M. Naranjo, Chief Clerk  
Senate

s/Ben Lujan

Ben Lujan, Speaker  
House of Representatives

s/Stephen R. Arias

Stephen R. Arias, Chief Clerk  
House of Representatives

**MEMBERSHIP OF THE ASSESSMENT AND  
ACCOUNTABILITY WORK GROUP, SJM 43**

PUBLIC EDUCATION DEPARTMENT

Dr. Catherine Cross Maple  
Deputy Secretary, Learning and Accountability

Mr. Carlos Martinez  
Assistant Secretary  
Assessment and Accountability Division

Ms. Denise Koscielniak  
Program Director, Special Education Bureau

Ms. R. Sue Gronewold  
Deputy Director, Special Education Bureau

PUBLIC SCHOOL SUPERINTENDENTS AND  
DELEGATES

Mr. Bill Green  
Superintendent, Quemado Independent Schools

Dr. Rhonda Seidenwurm  
Superintendent, Clovis Municipal Schools

Ms. Carrie Nigreville  
IRC/EPSS Coordinator, Clovis Municipal Schools

Ms. Bobbie J. Gutierrez  
Superintendent, Santa Fe Public Schools

Ms. Lynn VanderLinden  
Director, Assessment and Accountability  
Santa Fe Public Schools

DIRECTORS OF SPECIAL EDUCATION  
AND BILINGUAL EDUCATION

Dr. Susan Wilkinson Davies  
Director of Student Services, Jemz Valley Public  
Schools

Ms. Yvonne Lozano  
Associate Superintendent of Educational Services  
Gadsden Independent Schools

Ms. Carolyn R. Lindau  
Compliance Officer, Gadsden Independent Schools

Dr. Ed Monaghan  
Assistant Superintendent of Learning Services  
Gallup-McKinley County Public Schools

Ms. Glenda Rodriguez  
Special Education Director  
Las Cruces Public Schools

OTHER APPROPRIATE EDUCATORS

Ms. Gloria Hale  
Indian Education Director, Grants-Cibola County  
Schools

Ms. Lora Harlan  
Clovis Board of Education Member  
Clovis Municipal Schools

Mr. Andy Lotrich  
Early Childhood and K-12 Specialist  
AFT-New Mexico

Ms. Rose-Ann McKernan  
Executive Director, Research, Development and  
Accountability  
Albuquerque Public Schools

Ms. Sharon Morgan  
President, National Education Association-NM

Ms. Arlene Trujillo  
Principal, Deming Cesar Chavez Charter School

LEGISLATIVE COUNCIL SERVICE

Ms. Jonelle Maison  
Senior Bill Drafter, Legislative Council Service

LESC STAFF

Ms. Frances Maestas  
Director, Legislative Education Study Committee

Ms. Pamela Herman  
Senior Research Analyst, Legislative Education  
Study Committee

**Using Value Tables for a School-Level Accountability System**

Paper presented at the NCME Annual Conference  
April, 2006

Richard Hill  
The National Center for the Improvement of Educational Assessment

RECEIVED  
VIA E-MAIL

SEP 15 2008

**Background**

The goal of this activity was to create a statewide school-level accountability system that judged schools on the basis of changes in student scores across years. In order to make the system effective (i.e., to create a system that changed school-level behavior in desired ways), it was felt that the system should have the following characteristics:

1. It would be simple to understand. Schools cannot fairly be held accountability to a system that they do not understand.
2. It would be computationally simple. State policy-makers felt that schools would make positive changes only if they were able to compute the results to which they were being held accountable. Under the desired system, a school would be able calculate in advance what its results would be, in contrast to other systems that might require weeks or months before the school found out whether its students had grown sufficiently by the end of the school year.
3. It would use the performance levels currently used in the state's assessment system rather than scaled scores. The state had chosen performance levels as its primary reporting statistic because they wanted those scores to develop meaning to school-level personnel over time. They wanted the accountability system to reinforce the use of these statistics, and therefore wanted them to be at the center of it.

Systems such as hierarchical linear models were rejected for this system because they fail all three criteria. While such systems are straightforward enough on their surface, it is hard for most people other than the most devoted data analysts to understand precisely how the models work. The computations involved are complex, requiring sophisticated software and computer far beyond the range of most schools. Finally, they employ scaled scores. We decided to create a different method of measuring school growth.

**Developing a Value Table**

As an initial step was to consider an accountability system that compared the achievement level a student earns one year to that student's achievement level the previous year, and then assign a numerical value to that change. Higher values would be assigned to results that are more highly valued. We decided to called this matrix of assigned values a Value Table.

Table 1 shows the Value Table we first considered for the state's accountability system. Note that the state uses five performance levels with which to report its results. The levels are ordered from lowest to highest, with the third level being considered "proficient" for purposes of No Child Left Behind (NCLB) accountability. This table seemed fairly straightforward to develop. When students maintained their performance from one year to the next, the school

earned 100 points; the school earned 50 additional points each time a student went up one performance level, and lost 50 points for each performance level that the student went down.

Table 1

A Value Table Initially Considered  
(Value Table A)

Year 1 Performance Level	Year 2 Performance Level				
	I	II	III	IV	V
I	100	150	200	250	300
II	50	100	150	200	250
III	0	50	100	150	200
IV	-50	0	50	100	150
V	-100	-50	0	50	100

However, when we began to apply data to this Value Table, it was clear that it had some significant deficiencies. It was most obvious when we looked at the current changes in student performance across years. Table 2 provides the percentages of students who performed at each performance level in 2003, given the students' result in 2002. The results in the last column, "Average Growth Score," provide the results for all students statewide for each performance level. Thus, for example, the average growth score earned in 2003 by all students performing at Level I in 2002 was 120.5, while the average growth score earned by all students performing at Level V was 63.5. The lower the performance level in 2002, the higher the number of growth points the student was likely to earn. For the two extreme levels, the result is rather obvious; Level I students cannot earn fewer than 100 points, while Level V students cannot earn more than 100 points. As a result, the average for Level I students has to be higher than that for Level V students. But the trend holds up even for the middle levels; the average growth score for students at any level is lower than the comparable score for students at a lower level.

Table 2

Percentage of Students at Each Performance Level in Year 2,  
Given Performance Level in Year 1

Year 1 Performance Level	Year 2 Performance Level					Average Growth Score
	I	II	III	IV	V	
I	64	27	8	0	0	120.5
II	24	43	32	1	0	105.0
III	4	18	64	13	1	94.5
IV	0	2	39	51	8	82.5
V	0	0	10	53	37	63.5

It was obvious that regression was playing a role here that needed to be taken into account. However, once we started looking at the potential of using Value Tables, it became clear that an

additional possibility could—and should—be taken into account. Rather than valuing all gains and losses equally, it would be possible to value some outcomes more highly than others. For example, under NCLB, all gains below Level III are considered inconsequential, gains from below Level III to above Level III are highly valued, and gains above Level III are once again considered inconsequential (Note that some can—and do—have values that differ greatly from those of NCLB, but at least NCLB has clearly stated what its goals are). In contrast, the Alaska state legislature is considering awarding cash rewards to schools based on the changes in student performance from year to year, and the proposed legislation specifically states that schools should be rewarded, in part, for their success in moving Proficient students to Advanced, and maintaining students who already are Advanced at that level. Regardless of how one might feel about the contrasting priorities of NCLB and Alaska, it is clear that the value systems behind each piece of legislation are considerably different, and therefore the Value Tables that reflected each set of values would be just as different. Thus, the use of Value Tables not only permitted, but indeed required, that policy-makers explicitly state what educational outcomes they valued most highly, and that the Value Table that should be used in each state needed to be tailored to the values of the policy makers.

As a result, we returned to the state and reviewed our first proposed Value Table with state policy makers. During that meeting, they created two additional rules for us to follow: (1) No value should be less than zero, and (2) the value for any student that was at Level I in the second year should be zero. In addition, they told us that students that maintained performance at higher levels should be assigned more points than students that hold their own at lower levels. With that direction, we created the Value Table presented in Table 3 (the Average Growth Score is appended to each column).

Table 3

An Alternative Value Table, with Observed Results for Each Year 1 Performance Level when Applied to Actual Statewide Data across Years (Value Table B)

Year 1 Performance Level	Year 2 Performance Level					Average Growth Score
	I	II	III	IV	V	
I	0	200	250	300	230	74.0
II	0	100	130	180	230	86.4
III	0	50	100	150	200	94.5
IV	0	20	70	120	180	103.3
V	0	0	40	100	160	116.2

Value Table B overshoots the mark when regression is taken into account. Students who start at lower performance levels tend to earn the fewest points, while students at the higher levels earn more. Whereas the correlation between growth score and starting status scores was  $-.23$  for Value Table A, it was  $.61$  for this Value Table.

Thus, our next thought was to create a Value Table that we could call “neutral;” that is, one that took regression into account and produced average scores for each Year 1 performance level that were roughly equal. Then, we would ask policy makers to tweak that neutral table,

rather than creating one from scratch that we knew would not take regression into account. There are two sources of regression that one might consider. The first is due to measurement error within year, and the second is the regression of students across years. The first source certainly should be taken into account, and can be computed readily if one knows the reliability of the tests being used. The second source is questionable. We know that, even if we had true scores, some students would move from, say, Level II to Level III from one year to the next even if they had average instruction—it just would have been their year to grow. Students don't grow evenly every year, even if the instruction they have is constant from year to year. Some will have a banner year in, say, grade 3, and then perhaps show little growth from grade 3 to grade 4, while other students will show the opposite pattern. Thus, there will be some "churn" even if instructional effectiveness is constant for all students, and it likely will occur to a greater extent than one might think, since an underestimate of a student's performance level one year automatically introduces error into the measurement of gain. Doran and Cohen (2005) showed that linking error could lead to significantly greater uncertainty in the measurement of growth than researchers were used to seeing. On the other hand, looking at the actual results of how students change levels from one year to the next almost certainly overstates the amount of true regression going on. There certainly are some students who grow from Level II to Level III because they had truly superior instruction during that year, not because of regression effects. So if one applied statewide observed results to a truly neutral Value Table, one should expect somewhat higher scores for students at the lowest levels and lower scores for students at the highest levels. We had not considered that at the time we started this work, so we did not take that into account when we proposed a third Value Table for the state. At that time, we thought it was neutral; now, we would question whether it over-corrects for regression.

However, given our understanding of the regression issue at that time, we proposed the Value Table presented in Table 4.

Table 4

Another Alternative Value Table, with Observed Results for Each Year 1 Performance Level when Applied to Actual Statewide Data across Years (Value Table C)

Year 1 Performance Level	Year 2 Performance Level					Average Score
	I	II	III	IV	V	
I	0	200	400	400	400	86.0
II	0	100	150	200	250	93.0
III	0	50	100	150	200	94.5
IV	0	10	60	110	160	92.5
V	0	0	20	90	120	94.1

This table produced results with which the policy makers in the state felt comfortable. Although we almost certainly have overstated the amount of regression in our calculations of the "average score" for each Year 1 Performance Level, and thereby understated the difficulty

teachers with lower performing students will have achieving the same scores from this Value Table as teachers with higher performing students, policy makers were satisfied with that. One of the considerations that led to their acceptance of this Value Table is that there is a perception within the state that lower performing students probably already are receiving less effective instruction, and the fact that their scores tended to be lower was probably an appropriate reflection of that fact.

### **Correlations with Status Scores and Reliability**

If growth scores are supposed to be a measure of the effectiveness of schools, and it is presumed that schools in higher socio-economic areas (which have higher-achieving students) provide, on average, more effective schooling than schools in lower socio-economic areas, then there should be some moderate positive correlation between the status scores of schools and their growth scores. Researchers conducting HLM studies have reported negative correlations between status and the measures they are computing, which raise questions about the validity of those results for measuring school effectiveness. The correlations between status and growth scores were -0.23 when Table A was applied, +0.61 for Table B, and +0.44 for Table C. That last result seems to be the most reasonable, lending further support to the validity of the Table C results.

We also looked at the reliability of the growth scores using Table C. We computing multiple scores for each school by drawing samples with replacement and compared the values we got for schools under pairs of draws for all schools with 20 students or more. The reliability of status scores was 0.99. We had previously looked at the stability of improvement scores (comparing the status scores attained by one cohort with the status scores obtained by the next year's cohort) under similar conditions and found the reliability of the improvement scores to be 0.87. The reliability of the growth scores was 0.94—a value considerably lower than the status scores, but also considerably higher than the improvement scores. This is an additional indication that measures of growth may be more appropriate for school-level accountability than improvement scores.

### **Comparisons of Different Value Tables and of Value Tables to Other Measures of Growth**

It is of interest to know what the relationships are between this method of measuring growth and others that are being proposed. For this reason, staff at the Center for Assessment took data from one state and calculated schools' growth scores using each of three Value Tables, and compared those results to two other more traditional ways of computing student growth. The first alternative was a two-level analysis of covariance, using students' first year test scores as the covariate ("ANCOVA"), in which each school's score was expressed as the deviation from the overall predicted status. The second was a two-level hierarchical linear model that estimated slope parameters for schools, expressed as a deviation from the average slope for students statewide ("HLM slope"). Further details on these analyses are provided in Appendix A. The correlations of school scores among several statistics are reported in Table 5.

Table 5

## Correlations among Several Measures of School Growth

	ANCOVA	HLM Slope	Value Table A	Value Table B	Value Table C
Year 1 Status	.70	-.19	-.20	.65	.44
Year 2 Status	.88	.12	.08	.82	.64
ANCOVA		.57	.56	.93	.85
HLM Slope			.98	.53	.67
Value Table A				.54	.69
Value Table B					.95

First of all, the correlations show that it matters which Value Table you choose, so decisions about what values to insert into the table should not be taken lightly. Further, the correlations show that Value Table B provides essentially the same information as HLM slope, while Value Table C is a fairly close match to ANCOVA. Interestingly, when policy makers had an opportunity to define what they truly wanted in an accountability system, they favored Value Table C over the other two. That result, in turn, would imply that if policy makers truly understood what results statisticians were providing in their “growth” analyses, they might not find them as acceptable as they think they do. Note that ANCOVA results were well correlated with the Year 1 status scores, meaning that the schools that have high-achieving students are likely to continue to be judged successful if this method of assessing student growth is chosen. HLM slope results, on the other hand, are negatively correlated with schools’ starting positions. Again, if these statistics are supposed to measure teacher effectiveness, and policy makers believe there is a moderate tendency for better teachers to be located in higher scoring schools, then these results suggest that the statistics being calculated by ANCOVA and HLM Slope are not the ones policy makers would choose if they truly understood the procedures.

### Next Steps

Our initial analyses show relatively low correlations among school growth scores depending on the Value Table chosen. We also know that some Value Tables that appear on their surface to be appropriate turn out to be poorly correlated with other school-level statistics that should be indicating school effectiveness. This suggests that the process for establishing the Value Table to be used in a state school-level accountability system needs to be better understood than it does now. For example, we have developed a procedure similar to standard setting that allows policy-makers to articulate the values they wish to see reflected in their accountability system. We do not yet know how to create a Value Table that accurately reflects those values.

While our approach to school-level accountability is much simpler than HLM models, it uses just one year’s worth of prior test scores (and just one content area from that one year). We know that researchers have shown that student-level scores are much more accurately predicted when one uses multiple years of data from several content areas. Thus, it almost certainly is true that Value Tables would not be an adequate substitute for HLM procedures if the goal were

student-level predictions. However, we also know that the prediction errors tend to average out when results are reported at the teacher and school levels, so it is unclear whether the added complexity of HLM models provides significantly better predictability at those levels. Therefore, our next step will be to collect student-level data from another state for which more sophisticated HLM predictions have been made. We will apply a series of Value Tables to those same data and compare the correlations, at the student, teacher and school level, much as we have done in Table 6 in this paper. Only when we have completed such a study will we know to what extent the added complexity of HLM models changes the measures of growth from this simple model, and of these simple models, which ones best parallel the HLM models. Such information should suggest an appropriate course of action for policy makers.

Finally, to the extent that measures of school growth generated from some Value Tables correlate highly with the results of HLM models, a careful look should be taken at the outliers. We would need to question why some schools have a high relative rank under one model and a low rank under another. The answers to those questions should shed light on which of the statistical procedures is better at ranking out schools according to the values established by the state policy-makers.