

REPORT OF THE ACHIEVING CLASSROOM EXCELLENCE II TASK FORCE

Adopted December 27, 2007

Dr. Jo Pettigrew and Dr. Janet Barresi, Chairs

Report of the Achieving Classroom Excellence II Task Force

April 15, 2008

Introduction and Purpose:

The Achieving Classroom Excellence (ACE) II Task Force was created pursuant to SB 921 of the 2007 Session, authored by Sen. Clark Jolley and Rep. Tad Jones. The study was authorized out of concern over differences in test scores by fourth- and eighth-grade students in reading and mathematics on the Oklahoma State Testing Program (OSTP) as compared with that seen on a representative sample of students on the reading and math assessments on the National Assessment of Educational Progress (NAEP). Students performed significantly lower on the NAEP than did the same age students on the OSTP.

2007 NAEP and OSTP Test Scores READING

	NAEP		OSTP
	Nation	Oklahoma	
4 th Grade			
Scale Score	220	217	
% at Proficient	24%	22%	86%
% at Advanced	7%	4%	4%
8 th Grade			
Scale Score	261	260	
% at Proficient	27%	25%	70%
% at Advanced	2%	1%	9%

2007 NAEP and OSTP Test Scores MATH

	NAEP		OSTP
	Nation	Oklahoma	
4 th Grade			
Scale Score	239	237	
% at Proficient	33%	30%	63%
% at Advanced	5%	3%	19%
8 th Grade			
Scale Score	280	275	
% at Proficient	24%	18%	54%
% at Advanced	7%	3%	23%

Report of the ACE II Task Force

Based on these differences in test scores several questions were raised concerning the rigor of state content standards as well as the rigor and structure of state assessments currently administered under the Oklahoma State Testing Program. The task force agrees that there are many factors that affect academic performance. Some of these are the curriculum that is utilized, effectiveness of classroom instruction, class size, school leadership, length and number of instructional days, as well as the level of school funding. While these and more may affect academic performance, it is the content and process standards that serve as the foundation upon which all teaching and learning are built. Oklahoma's content standards are definitive statements about what all children must know to be productive citizens and to compete in a national and global marketplace. If these state content and process standards are rigorous and if state guidelines for test construction result in assessments that achieve a high degree of alignment with those standards then it should follow that students who take both the state assessments and the NAEP should perform equally well on both examinations.

In addition, the reporting of test results should be of such a form so that they facilitate interpretation by all stakeholders. Classroom teachers and administrators should be able to use these test results as a tool to evaluate existing education programs and to modify others that will enhance learning. In like manner, government leaders and the public should be able to monitor school performance and outcomes. Parents should be able to easily monitor their own child's performance while comparing overall performance outcomes of their school with other schools.

With all of this in mind, the committee was charged with examining the apparent disparity in scores between Oklahoma's student performance on the National Assessment of Educational Progress with their performance on state mandated tests under the Oklahoma State Testing Program. Specifically, the task force was assigned to study the following issues and submit a report of findings and recommendations to the Governor and Legislature by December 31, 2007:

1. Comparison of the Priority Academic Student Skills with other states' curricular standards, primarily states that score highest on the National Assessment of Educational Progress (NAEP);
2. Alignment of the Priority Academic Student Skills with the National Assessment of Educational Progress (NAEP) standards;
3. Feasibility of realigning the state performance level standards to NAEP performance level standards;
4. Differences in achievement levels among states based on exclusion rates on the NAEP; and
5. Feasibility of aligning the cut scores on state-mandated tests to NAEP cut scores.

Membership:

The task force was comprised of seven members selected from among public and private school educators and members of the business community, excluding any elected officials. The members, their professional affiliation and appointing authorities were as follows:

- Keith Ballard, Ed.D., Executive Director – Oklahoma State School Boards Association, appointed by the Governor to replace Jo Pettigrew who resigned;
- Janet Barresi, DDS – Dentist, appointed by the Speaker of the House of Representatives;
- Susan Harris, Vice President – Tulsa Chamber of Commerce, appointed by the Chair of the House Education Committee;
- Phyllis Hudecki, Ed.D., – Executive Director - Oklahoma Business & Education Coalition, appointed by Co-President Pro Tempore of the Senate;
- Cleatta Johnson, M.Ed., – Licensed Professional Counselor and Retired Educator, appointed by the Co-chairs of the Senate Education Committee;
- Diana Leggett – Asst. Supt. of Curriculum, Instruction and Personnel, Stillwater Public Schools;
- Rick Martin, M.Ed. – Superintendent of Prague Public Schools, appointed by the Minority Leader of the House of Representatives; and
- Jo Pettigrew, Ed.D., – Education Consultant, appointed by the Governor, but resigned prior to the expiration of the task force.

Dr. Pettigrew was elected by the members to serve as chair of the task force; however, her acceptance of an appointment by Governor Henry to serve on another commission precluded her continued service on the ACE II Task Force, and she was unable to preside at the final two meetings of the task force when this report was adopted. The Governor then appointed Dr. Ballard to replace Dr. Pettigrew. Dr. Janet Barresi was elected by the members to serve as chair of the task force for the final meetings.

Meetings:

The task force held seven meetings from September 24, 2007, to December 27, 2007, and heard presentations and comments from the State Department of Education, school districts, testing vendors, the Regional Education Laboratory Southwest, the Southwest Educational Development Laboratory, and the National Assessment Governing Board. A list of all presenters follows:

Presenters:

- Dr. Mary Crovo – Deputy Executive Director, National Assessment Governing Board
- Dr. Vicki Dimock – Program Director, Southwest Educational Development Laboratory
- Debra Ensminger – Director of Student Assessment, Jenks Public Schools
- Shan Glandon – Director of Curriculum and Instruction, Jenks Public Schools

Report of the ACE II Task Force

- Dr. Cindy Koss – Asst. State Supt., Office of Standards and Curriculum, State Dept. of Education
- Diana Leggett – Asst. Supt. of Curriculum, Instruction and Personnel, Stillwater Public Schools
- Rick Martin – Superintendent, Prague Public Schools
- Dr. Maridyth McBee – Vice President, Assessment Services, Pearson Education
- Dr. Lisa McGlaughlin – Asst. Superintendent, Western Heights School District
- Dr. Dean Nafziger – Director, Regional Education Laboratory Southwest
- Todd Nelson – Director of Student Assessment, Union Public Schools
- Don Rader – Superintendent, Alva Public Schools
- Jennifer Stegman – Asst. Supt., Office of Accountability and Assessments, State Dept. of Education
- Becky Szlichta – Coordinator of Testing, Stillwater Public Schools
- Kerri White – Mathematics Curriculum Director, Office of Standards and Curriculum, State Dept. of Education

Findings:

The findings and recommendations of the task force are organized below under each of the items the task force was charged with studying.

ITEM 1: Comparison of the Priority Academic Student Skills with other states' curricular standards, primarily states that score highest on the National Assessment of Educational Progress (NAEP).

We compared Oklahoma's Priority Academic Student Skills (PASS) with other states' curricular standards indirectly by way of review of national studies conducted by the U.S. Chamber of Commerce, Achieve, the American Diploma Project, Thomas B. Fordham Foundation, Regional Education Laboratory Southwest, The Education Sector, and various colleges and universities. None of these organizations have completed a comprehensive comparative review of K-12 curricular standards in each state. Most of these studies involved an indirect comparison by comparing Oklahoma's standards with its student's performance on the NAEP.

In its report, *Leaders and Laggards: A State-by-State Report Card on Educational Effectiveness*, Oklahoma was given an overall grade of "C" for the rigor of our standards. Oklahoma's English and Math standards were given a grade of "C" while our science standards were given a grade of "F". However, Oklahoma was only one of eight states that has aligned high school graduation requirements with college and workplace expectations. While no specific information was given concerning the methodology used to derive these grades for each state, the information in this report shows that 11 states scored higher than Oklahoma and 19 states had the same grade of "C" as Oklahoma. The states of New York, California, Indiana and Massachusetts received an overall grade of "A".

Oklahoma's curricular standards are developed by committee with support from the State Department of Education. The committee is comprised of classroom teachers, curriculum

Report of the ACE II Task Force

directors, higher education professionals and when possible, individuals from Oklahoma's business community whose area of expertise is in the subject area being developed. The State Department of Education is also responsible for setting test construction guidelines, developing the tests through an independent contractor, defining performance level descriptors, setting test cut scores and administering the test.

The NAEP framework is similar but not identical in form to state content standards. The framework was developed with the help of educators, curriculum directors, higher education professionals as well as members of the business community from across the country. It has gone through exhaustive independent review. The NAEP framework, much like state content and process standards is meant to be a statement about what all children should know and be able to do in each subject and at each grade level assessed.

The State Department of Education presented a comparison of state content and process standards to the NAEP framework using the Surveys of Enacted Curriculum (SEC). The comparison covered math and reading in grades 4 and 8. The SEC is a data tool which was principally designed to provide educators a subjective self analysis of the alignment of their own teaching practices with state content standards. The tool was developed by the Wisconsin Center for Ed Research under the sponsorship of the Council of Chief State School Officers (CCSSO). According to the CCSSO web site, CCSSO and project partners in 26 states are assisting educators to implement applications of the Surveys. While teacher practices and instructional content are self-reported, standards and assessments are analyzed by independent content experts, such as state level curriculum specialists from other states, university faculty, and business professionals. Therefore, cohort partners from other states performed the coding for this analysis.

The analysis shows that Oklahoma's mathematics standards are more closely aligned to the NAEP framework for Grade 4 than are those of Massachusetts which scored higher on the NAEP assessment (OK - 234, MA - 247). According to the SEC comparison, Oklahoma's mathematics standards for Grade 4 have a high degree of alignment with NAEP with a score of 0.329 as opposed to that of Massachusetts with a score of 0.279. A comparison of alignment scores for other states that outperformed Oklahoma on the NAEP is included in the table below.

Grade 4 Mathematics

	2005 NAEP score	SEC alignment score
Oklahoma	234	0.329
Massachusetts	247	0.279
Idaho	242	0.294
Montana	241	0.206
Iowa	240	0.257

A similar comparison for Grade 8 Mathematics does not reveal accurate information due to the fact that Oklahoma's Grade 8 Mathematics Process Standards were not included in the Surveys of Enacted Curriculum study when it was originally conducted. This oversight by the Wisconsin Center for Education Research resulted in a low alignment score for the Grade 8 mathematics standards as compared to the NAEP framework.

Report of the ACE II Task Force

The SEC shows that Oklahoma's standards, along with the standards of most other states, include a wider variety of content in Language Arts than is included in the NAEP Reading framework. Both the Grade 4 and Grade 8 Language Arts PASS standards include content not covered by the NAEP Reading framework such as Vocabulary, Writing Process, Writing Components, Writing Applications, Listening and Viewing, as well as topics with minimal coverage in the NAEP Reading framework such as Phoneme Awareness, Language Study, Speaking and Presenting. It should be noted however that NAEP has a separate Writing framework and assessment. The broader body of content used by most states results in lower alignment scores, but comparisons between states shows Oklahoma with a higher alignment score than many states that outperform Oklahoma on the NAEP.

Grade 4 Reading

	2005 NAEP score	SEC alignment score
Oklahoma	214	0.231
Vermont	227	0.223
Maine	225	0.208
Ohio	223	0.212
Idaho	222	0.215
Wisconsin	221	0.185

Grade 8 Reading

	2005 NAEP score	SEC alignment score
Oklahoma	260	0.217
Vermont	269	0.192
Ohio	267	0.173
Idaho	264	0.145
Indiana	261	0.205

The SEC cannot be used to compare states on science at this time since the new NAEP Science frameworks have not yet been included in the Surveys of Enacted Curriculum study.

While the SEC appears to be a useful tool for classroom teachers and administrators to use to assess the effectiveness of instruction and alignment to state standards, some task force members expressed concern about the lack of widespread use of this tool nationally. There was also concern over the subjective nature of the analysis for purposes of comparing state curricular standards to the NAEP framework.

The Regional Educational Laboratory – Southwest (REL-SW) at Edvance Research conducted a study of each of the five states in its region. The purpose of the study was to do a prospective comparison of each state's assessment standards in science with the 2009 NAEP science examination. It was designed in an effort to alert state education officials so they could determine whether or not they wanted to make changes to their own state assessment standards and specifications in order for these assessments to be more closely aligned to the NAEP.

Report of the ACE II Task Force

Values applied range from 0 or no alignment to 3 or perfect alignment. They are as follows:

Oklahoma

- Grade 4 – 1.24
- Grade 8 – 1.53
- Grade 12 – 1.24 (all content); 1.92 (life science)

Arkansas

- Grade 4 – 2.0
- Grade 8 – 2.1
- Grade 12 – 1.3 (all content); 2.1 (life science)

New Mexico

- Grade 4 – 2.2
- Grade 8 – 2.1
- Grade 12 – 2.3

Texas

- Grade 4 – 2.0
- Grade 8 – 1.6
- Grade 12 – 1.6 (all content); 1.8 (life science)

Louisiana

- Grade 4 – 2.6
- Grade 8 – 2.1
- Grade 12 – 2.5

REL-SW notes: “In comparing Louisiana benchmarks and grade level expectations with the NAEP, the overall alignment ratings for elementary, middle and high school are generally very high. The combination of Louisiana’s benchmarks and grade level expectations at all grade levels aligns very well with the NAEP content statements, because the grade level expectations often parallel NAEP statements in their level of detail.”

RECOMMENDATIONS:

The State Department of Education should work in partnership with an independent, third party contractor such as Achieve, Inc. to perform a comprehensive crosswalk of Oklahoma’s PASS as compared to other state’s standards. This independent and comprehensive study would allow not only a comparative analysis to other state’s standards but would also be anchored by an analysis of state content and process standards to other national standards such as those seen in NAEP and the American Diploma Project.

ITEM 2: Alignment of the Priority Academic Student Skills with the National Assessment of Educational Progress (NAEP) standards.

Report of the ACE II Task Force

As stated above, state assessments should exhibit a high level of alignment with rigorous state content and process standards. A complete and independently performed comparison of Oklahoma's state standards (PASS) to the NAEP framework and other national standards provides the anchor necessary to make definitive statements about the quality of state standards. However, the task force is concerned that focusing only on the degree of alignment of state standards to the NAEP framework, while useful, would not fully address the issue of achievement gaps. The committee heard evidence that there are other factors that impact student achievement such as teacher quality, the level of school funding, whether or not the standards are being taught in the classroom, the particular curriculum being utilized, vertical alignment of instruction, time on task as well as the number of instructional days, professional communication and teaming within each school, the importance placed on NAEP assessment at school sites, logistics of testing large amounts of students on computers in a specific window of time and the actual format of the tests themselves.

The NAEP is developed and implemented as a result of the work of two separate organizations that are working under federal congressional authority. The National Assessment Governing Board (NAGB) serves as a policy arm and through an external contractor it develops the framework, sets achievement levels, directs communications and disseminates information through various avenues including its web site.

The National Center for Education Statistics (NCES) serves as the operations arm of the NAEP. Its duties include item development, sampling and data collection, design analysis and reporting, materials distribution and scoring, state service center and web technology.

It was noted by some task force members that this separation of powers between NAGB and its role as a policy oriented agency and NCES as the operations division, provides an important component of reliability and validity to not only the development of the NAEP itself but also to the administration of the test and reporting of its results. In addition this separation of duties would place the administration of the testing program in a neutral entity, thus separating the functions so the entity responsible for making progress in student achievement is not managing the accountability function. The policy agency could then concentrate on improving teaching and learning by providing technical assistance and managing curriculum content standards, while an operations entity handles the evaluation of students and reporting of assessment results.

The task force heard testimony that the NAEP assessment is constructed in such a way as to include a large percentage of items that have a greater depth of knowledge and many constructed response questions. In April of 2007, the ACE I Steering Committee recommended new guidelines for test construction that would eventually result in a test that is aligned more closely to the NAEP both in format and content. Among the steering committee's recommendations is the use of constructed-response test items. Constructed response questions assess several aspects of a student's knowledge of the subject matter including extension of knowledge as well as abstract reasoning, synthesis and analysis. These recommendations on test construction guidelines wait consideration by the State Board of Education.

As stated in section 1, The Regional Educational Laboratory – Southwest presented their research report, *Aligning Science Assessment Standards: Oklahoma and the 2009 National Assessment of Educational Progress (NAEP)*. Test specifications for both Oklahoma and the NAEP were used by REL-SW as a means of evaluating alignment. According to REL-SW tests should reflect a high degree of alignment with state standards, therefore comparing test specifications could be used as a tool to evaluate alignment. The alignment study “was designed to give policymakers and educators a head start if they choose to make changes in state assessment standards and specifications to develop an assessment system more closely aligned to that used for NAEP”. The summary of the report stated, “Reviewers found Oklahoma to be generally unaligned with the NAEP. Oklahoma’s standards, on the whole, are less detailed and contain less content than the NAEP. The majority of the NAEP content statements are unaddressed by the content standards and objectives in Oklahoma’s test specification documents”.

In the comparison of state standards to the NAEP framework, alignment was rated on a scale from one to three, with three indicating the state standards fully address or exceed NAEP content. According to the study, the majority (82%) of Oklahoma grade 5 science standards do not address NAEP grade 4 content. For grade 8, 53% of all NAEP content is not addressed by the Oklahoma objectives in the grade 8 test specifications document. The overall alignment rating for Oklahoma science content at grade 8 and the NAEP grade 8 is 1.53. At grade 12, 80% of all NAEP content is not addressed, with an overall alignment rating for Oklahoma science content in biology and NAEP grade 12 of 1.24. However, at grade 12 the NAEP tests cumulative knowledge in all of the sciences while the Oklahoma End of Instruction Examinations are limited for the time being to Biology. It would follow then that as a result of the study it was recommended that “If state policymakers wish to increase the alignment between the state assessments and the NAEP, areas to consider are adding physical science and Earth and space science to the high school examination and including a wider variety of test item types.”

The REL-SW is currently engaged in an alignment study for mathematics.

RECOMMENDATIONS:

In order to increase the level of alignment between assessments in the Oklahoma State Testing Program and NAEP, the committee recommends that at a minimum, the State Board of Education adopt all of the recommendations of the Workgroup on Curriculum Alignment, Assessment and Cut Scores of the ACE I Steering Committee approved in April of 2007. We also agree with this same workgroup that the addition of constructed response questions would be a significant improvement to the quality of Oklahoma state tests and to alignment with the NAEP as well other national examinations.

We request that guidelines be adopted that require the employment of out-of-state educators to independently grade state constructed response questions. This would be a process that mirrors that used in grading written responses on Advanced Placement Examinations and SAT examinations. The use of independent and impartial out-of-state educators to grade the examination would yield a higher level of reliability and validity to score results.

Report of the ACE II Task Force

In addition to the above recommendations from the ACE I Steering Committee, this committee also recommends that the state legislature consider the Oklahoma State Department of Education's funding request with regard to test item development. In its FY 2009 and FY 2010 budget requests, the Oklahoma State Department of Education has stated, "The first step in strengthening the state assessments, in order to emphasize critical thinking and reasoning skills and align state tests with the NAEP, requires additional funding relative to the number of open-ended and short constructed-response items to be included on the OCCT. To include two short constructed-response items on reading and mathematics assessments in Grades 3-8 and on EOI exams, testing vendors project an amount of \$3.1 million in Fiscal Year 2009 for test-item development and field testing, and \$5.2 million in Fiscal Year 2010, when the items would become fully operational. The additional cost is primarily because of human scoring needed for this type of test question."

FY 2009 Requested Funding	\$3,100,000
FY 2010 Requested Funding	\$5,200,000

The current budget for testing for the OSTP is approximately \$15 million per year with \$10.85 million of those dollars coming from state appropriations. This additional funding would be an important first step in improving state assessments.

The task force recommends that all Oklahoma content and process standards at all grades be revised to achieve a high degree of alignment with national standards including the NAEP. The task force recognizes that there is currently a schedule for regular review of content and process standards for each grade and subject. It is during these regularly scheduled events that those state standards found to be deficient by independent review, be revised to an appropriate level of alignment with the NAEP and other national assessments.

The task force also recommends that membership of the revision committee for state content and process standards include at least 15% membership from the business community with occupations that align with the subject area being reviewed. It is also recommended that teachers from a higher grade than the grade being considered as well as representatives from higher education serve on this revision committee. The presence of these representatives will help maintain focus on high expectations for skill development and academic achievement at each grade so students will be well prepared for the next grade.

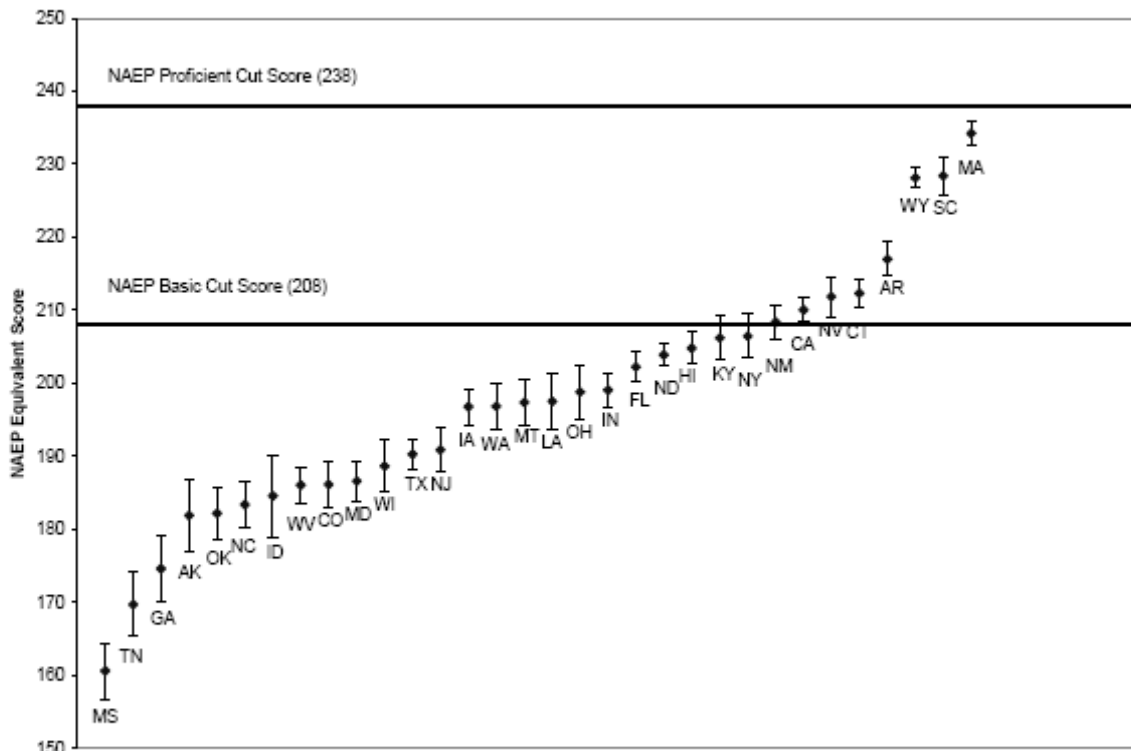
The structure of the NAEP on the federal level is a model to consider in Oklahoma. The structure provides separation of duties between NAGB as the policy arm and NCES as the operations arm. Given the importance of insuring the state has a valid and reliable accountability system to accurately track the progress of our students, policymakers are urged to give consideration to separating the duties and functions of the state agency to provide autonomy and instill confidence in the integrity of our education data and accountability systems.

ITEM 3: Feasibility of realigning the state performance level standards to NAEP performance level standards.

The National Center for Educational Statistics issued a report titled *Mapping 2005 State Proficiency Standards onto the NAEP Scales*. The study mapped state proficiency standards in reading and mathematics in grades 4 and 8 on to the appropriate NAEP scale. Data from the 2004-2005 academic year was used for the study. It was noted that across all states evaluated there was wide variation when comparing the NAEP score equivalents to the states' proficiency standards. With this in mind the study concluded: "There is a strong negative correlation between the proportions of students meeting the states' proficiency standards and the NAEP score equivalents to those standards, suggesting that the observed heterogeneity in states' reported percents proficient can be largely attributed to differences in the stringency of their standards. **There is, at best, a weak relationship between the NAEP score equivalents for the state proficiency standard and the states' average scores on NAEP. Finally, most of the NAEP score equivalents fall below the cut-point corresponding to the NAEP Proficient standard, and many fall below the cut-point corresponding to the NAEP Basic standard.**"

Figures 2-5 illustrate these results.

Figure 2. NAEP score equivalents of states' proficiency standards for reading, grade 4: 2005



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment, and National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

Report of the ACE II Task Force

Figure 3. NAEP score equivalents of states' proficiency standards for reading, grade 8: 2005

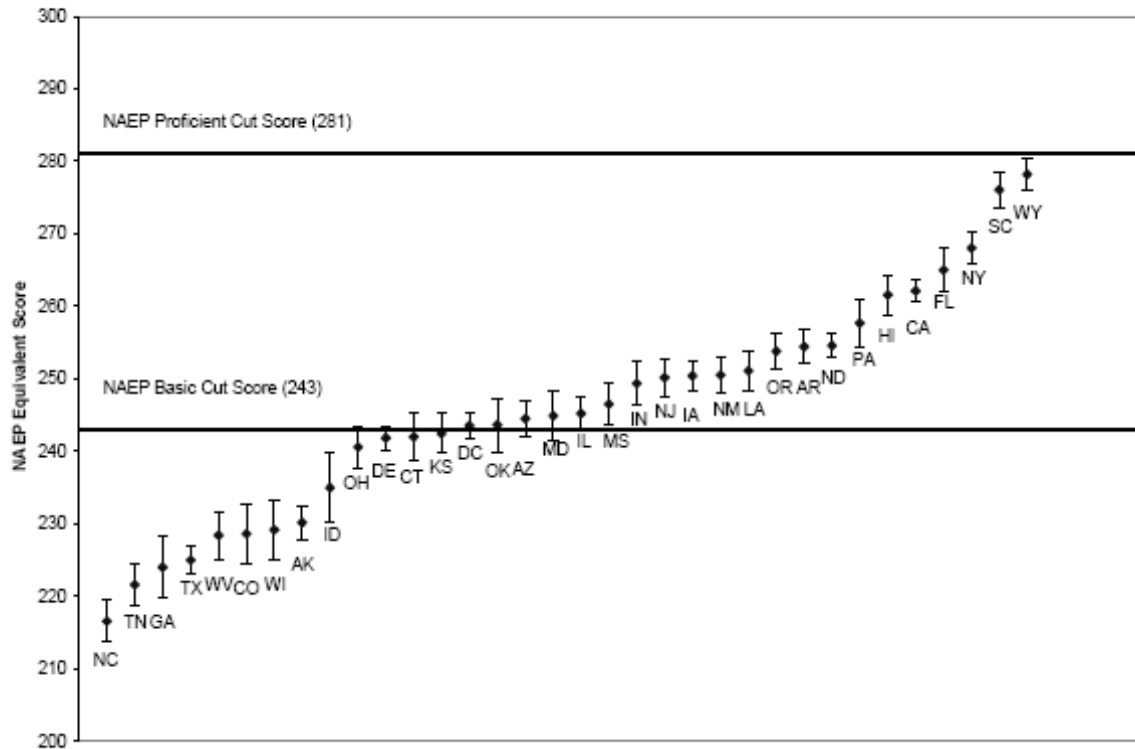


Figure 4. NAEP score equivalents of states' proficiency standards for mathematics, grade 4: 2005

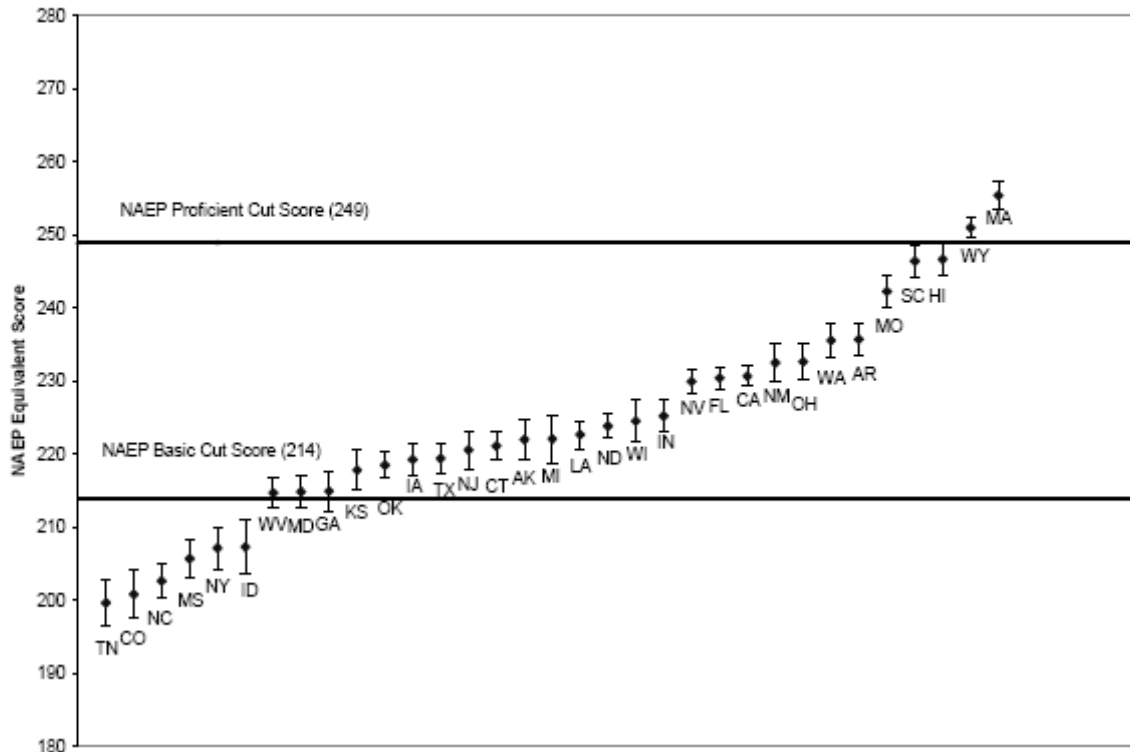
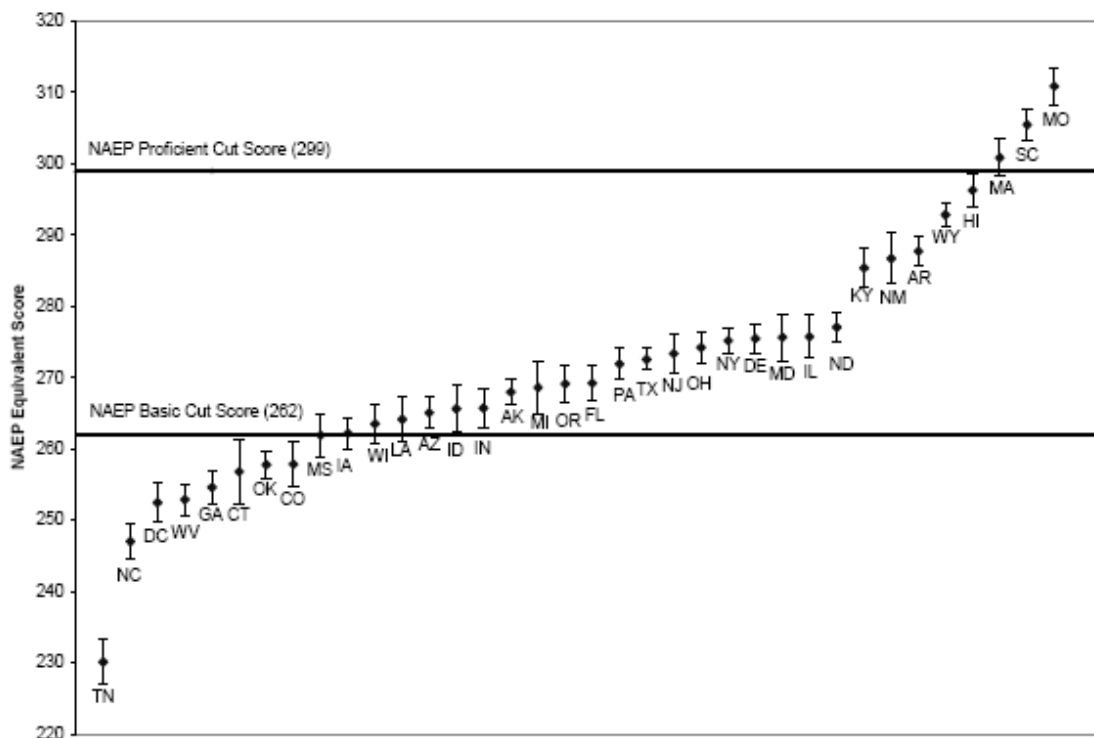


Figure 5. NAEP score equivalents of states' proficiency standards for mathematics, grade 8: 2005



This report found a strong negative correlation between the proportions of students meeting the state proficiency standards and the NAEP score equivalents. They concluded that this was due largely to the differences in the stringency of the Oklahoma state standards as compared to the NAEP.

NAEP performance descriptors recognize levels of achievement as Basic, Proficient and Advanced. They are defined as follows:

- **BASIC** – denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- **PROFICIENT** – represents solid academic performance. Students reaching this level have demonstrated competency over challenging subject matter.
- **ADVANCED** – represents superior performance.

As illustrated, most of the NAEP score equivalents for Oklahoma students tested fell below the cut point corresponding to the NAEP performance standard designated as basic. The task force noted the low number of Oklahoma students scoring at or above the level of proficient on the NAEP. Indeed, the NAEP score equivalent for Oklahoma was significantly below the NAEP basic score for 4th grade reading and was just at the basic level for 8th grade

Report of the ACE II Task Force

reading. Oklahoma was just above basic for 4th grade math and just below basic for 8th grade mathematics.

In April of 2007, the ACE I Steering Committee unanimously passed the following recommendation by the Workgroup on Curriculum Alignment, Assessment and Cut Scores that pertains to performance level descriptors. This recommendation awaits consideration by the State Board of Education. The recommendation reads as follows:

Performance Level Descriptors

Performance Level Descriptors shall specify the amount of knowledge and/or skills required to achieve an outcome or classification. These descriptors shall be utilized as part of the process of setting cut scores during the review and revision of existing assessments as well as the development of all new assessments. They shall also be used where appropriate in reporting assessment scores.

1. **Advanced** – The student demonstrates superior performance on challenging subject matter.
2. **Proficient** – The student demonstrates mastery of appropriate grade-level subject matter and that students are ready for the next grade, course, or level of education, as applicable.
3. **Limited Knowledge** – The student demonstrates partial mastery of the essential knowledge and skills appropriate to their grade level, course, or level of education as applicable.
4. **Unsatisfactory** – The student does not perform at least at the limited knowledge level.

The ACE II Task Force notes the similarity in performance level descriptors between that designated as Limited Knowledge by the ACE I Steering Committee and that designated as Basic by the National Assessment Governing Board. The same observation is true for the term Proficient on the ACE I recommendations and that defined as Proficient by the National Assessment Governing Board.

RECOMMENDATIONS:

The ACE II Task Force urges that the Oklahoma State Board of Education adopt the above definitions of performance level descriptors as recommended by the ACE I Steering Committee as well as the use of these descriptors when setting OSTP test cut scores and in reporting test results. When focusing on the descriptors for Proficiency, some ACE II Task Force members expressed the belief that the term “mastery of appropriate grade-level subject matter” should address the question, “what ought to be known”. All shared the belief that this descriptor should demand “that students be ready for the next grade, course, or level of education as applicable”.

State panels should be established to review and revise the performance level descriptors for each subject and grade level. They shall assure that these descriptors align with those

Report of the ACE II Task Force

established by NAEP performance setting processes. These panels should have the same composition as that seen in the NAEP review panel.

ITEM 4: Differences in achievement levels among states based on exclusion rates on the NAEP.

Student performance for the 2007 NAEP show Oklahoma student performance overall below the national average at the level, except in 4th grade mathematics where Oklahoma students scored one percent above in the basic category, but six percent below for proficient. When the percentage of Oklahoma students identified as students with disabilities or English language learners are compared to the national average as well as high performing states, very similar exclusion rates are revealed as is demonstrated in the table below. Data for the lowest performing states are also included and their exclusion rates range from very high in the District of Columbia to very low in Mississippi. Therefore the task force did not find a difference in achievement on the NAEP based on exclusion rates alone.

2007 NAEP Performance of public school students compared to public school students with disabilities and English language learners identified, excluded and accommodated as a percentage of all students by state									
4 th Grade Reading									
State/ Juris.	NAEP Score		% Overall Excluded	Students with Disabilities			English Language Learners		
	% at or above Basic	% at or above Proficient		% Ident.	% Excl.	% Accom.	% Ident.	% Excl.	% Accom.
DC	39	14	14	15	11	3	9	4	4
Miss.	51	19	2	11	2	4	1	#	#
Okla.	65	27	7	15	7	5	5	1	1
Nation	66	32	6	14	5	6	11	2	2
NJ	77	43	7	14	5	7	4	2	1
Mass.	81	49	6	18	5	10	6	2	1
8 th Grade Reading									
DC	48	12	13	18	12	4	4	2	1
Miss.	60	17	3	9	6	4	#	#	#
Okla.	72	26	7	16	6	5	3	1	#
Nation	73	29	5	13	5	6	7	2	1
NJ	81	39	7	15	5	8	4	2	1
Mass.	84	43	7	18	6	10	4	2	#

Report of the ACE II Task Force

4 th Grade Mathematics									
DC	49	14	6	14	5	8	8	2	5
Miss.	70	21	1	10	1	6	1	#	#
Nation	81	39	3	14	3	8	11	1	3
Okla.	82	33	5	14	5	6	5	#	1
NJ	90	52	2	14	2	11	4	#	3
Mass.	93	58	5	18	5	11	6	1	2
8 th Grade Mathematics									
DC	34	8	10	17	9	6	4	1	2
Miss.	54	14	2	11	2	6	#	#	#
Okla.	66	21	8	14	8	4	4	1	1
Nation	70	31	4	13	4	6	7	1	2
NJ	77	40	3	14	3	11	4	1	2
Mass.	85	51	9	17	9	6	3	1	1

Indicates the value rounds to zero.

The term “exclusion rates” usually applies to student groups that are excluded from reporting traditional testing data. Many of these students, those with various disabilities and English Language Learners, take the test but do so with various forms of accommodations provided to them during the assessment.

The term “exclusion rates” has been applied to another statistical group which is more accurately known as “n” size. The term “n” size applies to individual groups of students whose test results are disaggregated for ethnicity, race, sex and income levels. The statistical information affords educators the opportunity to assess the impact of instructional techniques for these groups. Results for these groups enter into accountability requirements for NCLB. The task force reviewed the November 2006 report by the Education Commission of the States, “Minimum Subgroup Size for Adequate Yearly Progress (AYP), State Trends and Highlights”. This report compared exclusion rates for subgroups based on ethnicity, income, learning challenges, and English Language Learners. The report states,

“The “n” size must be large enough to ensure statistically reliable information and prevent personal information from being revealed. Schools and districts are held accountable only for the student groups that met the minimum subgroup number. If a state chooses an “n” of 35, for example, a school with only 20 English language learners (ELL) in the tested grades would not be held accountable for this group of students. The test results for these students, however, would factor into the district’s (or possibly the state’s) AYP calculation and results.”

The task force noted that the report stated that “In 2006, states with the largest subgroup numbers included Oklahoma (52); California, Texas, Virginia and West Virginia (50); Illinois,

Rhode Island and Tennessee (45).” Oklahoma’s use of an “n” size of 52 has resulted in an exemption of nearly 62,000 children from accountability from school sites around Oklahoma. While these students do factor into district and state AYP calculations, it gives an altered impression of individual school’s actual performance and could lead districts to a false conclusion regarding the effectiveness of educational programs offered to these students.

RECOMMENDATIONS:

The task force recommends that Oklahoma’s minimum subgroup number, or “n” number, be reduced to no more than 30 students. This number is used by the majority of states. This will more accurately report performance at the school level for accountability purposes as well as provide school level faculty and administrators the information they need to properly focus curriculum and instruction to assist all students in each subgroup.

ITEM 5: Feasibility of aligning the cut scores on state-mandated tests to NAEP cut scores.

At present, Oklahoma uses the bookmark methodology to set cut scores. The technique is also used for setting cut scores on the NAEP. The method requires several rounds of reviews of test questions that are aligned from least difficult to most difficult by the vendor who produced the test. A subsequent round allows participants to move their bookmark after reviewing impact data from students who took the test. During each round participants set bookmarks at each level that best equates to the performance level descriptor for that particular subject. Therefore, there is a bookmark set that delineates performance levels between unsatisfactory and limited knowledge (known as “Basic” on the NAEP), between limited knowledge and satisfactory (“Proficient” on the NAEP) as well as a mark delineating the boundaries between satisfactory and advanced performance. These bookmarks are specific to the test and test items and can not be set at the same linear point as on the NAEP simply because the test questions are different on the NAEP than that on the OSTP.

Relative to this point, the committee appreciated the presentation by Dr. Mary Crovo, Deputy Executive Director of the National Assessment Governing Board. As part of her presentation, Dr. Crovo discussed the process of setting achievement levels (cut scores) on the NAEP. Under the guidance of NAGB, the process includes first the formation of a panel which is comprised of 70% educators and 30% non-educators. Of the educators, 55% includes classroom teachers for the grade and subject being considered. Approximately 15% of the educator group are “other educators with knowledge of the subject matter and students at the grade level of the assessment”. Non-educators make up the remaining 30% of the group and may be representatives of the business community or are professionals working in the particular subject area. The process of setting achievement levels also utilizes the “bookmark” method. Training of panelists is extensive and all panelists must take the test.

The reporting of test results is a related matter that has been the source of confusion and debate. In a report by the Education Sector, Oklahoma is one of twenty states judged “not transparent” in making information on cut scores available and understandable to educators and the public. In their July 2006 report “Making the Cut: How States Set Passing Scores on Standardized Tests”, they accessed the Oklahoma State Department of Education

web site. A similar technique was used on each state. If they could find the state's cut scores then the state was deemed to be transparent. If the state failed to display cut scores or only reported scaled scores then that state was deemed to be not transparent. The task force noted the recommendations in the report:

1. Make the score setting process and the results more transparent and accessible.

Among other things they recommend that states describe how cut scores are set, the range of scores the judges considered as well as information about what kinds of people participated in the process.

2. Include outside representatives on score setting panels to improve alignment and help ensure rigor.

In addition to grade level panelists, the state should include educators from subsequent grades (as well as higher education) so the panel can stay focused on setting achievement levels that will yield success at the next grade level.

3. Validate tests in an ongoing manner.

A regular and ongoing review of state standards and tests should be made to ensure that they are all aligned to state policy goals. They should also validate test results by comparing student performance to other national tests such as the NAEP as well as student performance to real-world competencies such as reading an editorial in the newspaper, writing an essay, or making sense of a graph.

The report goes on to state that print and broadcast media should report more than just test scores and should also report on how cut scores are determined and what a score of proficient means.

Data-driven decision-making has been proven to be an effective tool as educators focus on student-specific needs. Clear, concise and useable information is invaluable for all educators as they strive to provide academic excellence to all students. To this end, district-level student assessment and curriculum directors as well as superintendents provided information to the task force from the district perspective. Their concerns were similar to those criticisms offered by national interests. The presenters expressed concern over lack of communication between the State Department of Education and parents as well as educators on definitions of performance level descriptors and student scores. In other words, parents and the public are confused when a student is described as having achieved "satisfactory" performance when that student has answered only 45 to 64 percent of the test items correctly. District representatives also expressed frustration over the difficulty in interpreting testing outcomes for their district and subgroups of students. A theme was noted in that most of the speakers requested that data be presented in a more "user friendly" and "transparent" format for district officials, principals, classroom teachers and parents as well as other interested stakeholders in the community to use as they monitor student progress.

Report of the ACE II Task Force

A presentation by Dr. Vicki Dimock, Program Director of SEDL, addressed this issue. She cited an explanation of the use of scale scores used in interpretation of test results using the Test Interpretation Manual for the OSTP, grades 3-8. It states:

“Oklahoma Performance Index (OPI) scores are reported on a scale from 400 to 990. OPI scores, also called scale scores, are more accurate than “percent correct” scores because they factor in the difficulty level of the test and correct for possible guessing. OPI scores are based on percent correct scores but are reported on a scale of 400 to 990 so that they mean the same thing from one year to the next. Because tests have different questions on them in different years, a test one year could be slightly more or less difficult than the next year. OPI scores take into account this difference in difficulty and report scores on a common scale so that OPI scores mean the same thing from one year to the next. For example, students one year may need to answer 37 questions correctly to obtain an OPI score of 750. If the test the next year is a little more difficult, students may only need to answer 35 questions correctly to obtain the same 750 OPI score. This way, scores for groups of students can be accurately compared from one year to the next using OPI scores.”

While this explanation is reminiscent of the interpretation of scaled scores on the ACT or SAT, educators who testified and task force members appeared frustrated over the lack of communication between their districts and the State Department of Education as to this process. One presenter explained that it is frustrating to have to go to the Test Interpretation Manual to try to understand how scale scores are computed and then converted to cut scores.

On a related matter of interpretation of test results is the concern that when it is revealed that a student answers less than half of the test items correctly on an OSTP assessment, yet receives a rating of “Satisfactory”, that it appears that students do not have to answer a majority of the test items correctly in order to pass the test. Dr. Dimock addressed this issue when she cited a study by Rotherman in 2006, “Making the Cut: How States Set Passing Scores on Standardized Tests”. He points out that “On a difficult test, a cut score that represents answering correctly 65 percent of the test items may in fact be much more challenging than “D” work. Conversely, on an easy test a score of 80 percent may not reflect a high level of learning” Yeager, in an October 2007 report for the Education Sector titled “Understanding the NAEP: Inside the Nation’s Report Card” states: NAEP scores are not as simple to interpret as pure percentage scores or letter grades, e.g. 95 percent is an “A,” 85 percent is a “B.” A NAEP score of 220 is not 10 percent better than a score of 200, because there is no single formula to convert raw scores on test sections to scale scores for the test as a whole. Instead the weight of each individual question in contributing to the scale score is determined by that year’s student data. Additionally, changes in NAEP scores from one testing to the next may be only 1-2 points, but can be statistically significant due to the large sample size.

Understanding this fact could prove confusing for classroom teachers as they interpret student performance on formative tests during the school term as a way of preparing

Report of the ACE II Task Force

students for the OSTP. Formative tests, especially those constructed by the teacher could be graded in the same way they grade for their classroom instructional tests. In other words, that teacher could apply a grade based on a percentage correct rather than apply a scale score to the test result. Therefore, a teacher may get a false or misleading interpretation of how well their students are or are not prepared for the OSTP assessment.

RECOMMENDATIONS:

Oklahoma panels that set cut scores should have the same composition in membership as that seen on the NAEP. This membership will reflect a broader scope of participants including more members of the business community and higher education.

The process of training these panels should be reviewed and revised to emphasize the importance of applying performance level descriptors when making decisions about setting bookmarks. The panel should address the question, “what ought to be known” at that particular level and “will students performing at this level be prepared for the next grade level or content course”.

The “consequential data” or the questions of “consequential validity” introduced during the process of setting cut scores should be weighted carefully against the questions of student preparedness for the next grade level or content course. It is therefore recommended that student preparedness must be the key consideration in establishing cut scores for achievement levels.

Transparency in reporting cut scores, and test results is paramount so that all stakeholders including educators, parents and those in the community at large can easily understand and interpret student performance. Transparency will provide educators with an understanding and awareness critical to targeting performance deficits which will in turn increase student performance on both state assessments and the NAEP. Cut scores and the process used to determine scale scores should be released to educators as part of the assessment reports. District Directors of Testing and Curriculum should receive ongoing training regarding this matter. They will then be able to assist all stakeholders in their districts to interpret the impact of test outcomes and plan appropriately to improve programs or adjust instructional efforts to best benefit all students. Educators should not have to refer to the state technical manual to determine this important information.

The State Department of Education, together with the Office of Accountability should team together to assure that the public are aware of the cut score setting process and what cut scores mean in terms of determining student achievement levels. The media should be made aware of this process as well. In addition, the media should be educated so they may report to the public in a meaningful way, state test results and compare those results to student performance on the NAEP. This would include a comprehensive and in-depth explanation of what the performance level of “Proficient” means. It should be noted that a recommendation similar to this was approved by the ACE I Steering Committee and a \$1 million budget item is requested by the State Department of Education for the purpose of dissemination of information to the public and media regarding the OSTP.

GENERAL RECOMMENDATIONS TO IMPROVE STUDENT ACHIEVEMENT AND STUDENT PREPAREDNESS:

The important work of aligning both our state standards and testing to NAEP and other national organizations is necessary to assure that Oklahoma's students are well prepared to participate in a global economy. In view of this, Oklahoma should continue to be an active participant in the American Diploma Project network and participate in the next generation of state accountability and assessment systems as developed under the ADP.

The SDE, OCTP, OTAC, K20 Center, Center for Effective Schools, Regents for Higher Education, and MC3 as well as all teacher professional organizations should work together as partners to develop a professional development plan for state educators that will:

- Systematically build capacity for content area specialists, particularly math and reading specialists, and provide a more focused sustainable system that will assure that all students reach optimum performance for each grade or subject. As a result of the ACE I recommendations the State Department of Education has made a preliminary budget request to establish a program that would provide training for reading and math coaches that would serve statewide. It is not clear as of the date of this report who would administer the program;
- Reallocate existing resources and allocate new funds to provide sustained, job-embedded professional development that is ongoing throughout the year;
- Focus on the "Professional Teaching and Learning Cycle" as defined by SDE; and
- Provide professional development, time and support for peer coaching (instructional coaches in content areas).

The Legislature should allocate additional funds and additional time in the extended year school calendar for Professional Development for teachers and administrators.

Administrator preparation programs and all SDE or professional association sponsored professional development programs should target the development of leadership skills that specifically address the utilization of data for school improvement, cultural change, establishing high expectations for student outcomes, and monitoring classroom instruction to address expectations and achievement.

College or university teacher preparation programs should include in their course curricula instruction on the interpretation of statistical data regarding student performance on state and national assessments and the effective application of that data in a meaningful way such that it will positively impact instruction for each student.

The State Department of Education should take the necessary steps to qualify for eligibility to use the Growth Model for Accountability which is now open to all states. This entails having a student identification number for tracking each student across the state and having consistent cut scores from grade to grade and from year to year as well as reduction

Report of the ACE II Task Force

of our “n” size from its present level of 52. The Growth Model will enable districts and schools to be held accountable for student growth toward set standards which will increase the opportunity to focus on change in individual student achievement.

Parents, community members and those in the business community should have access to transparent and clearly stated information regarding student testing outcomes. This information should be easy to interpret. This will allow all stakeholders to become a stronger and more knowledgeable partner in assisting Oklahoma’s youth to achieve their potential and to enrich our state.

Report of the ACE II Task Force

APPENDIX

Legislation enacted creating the ACE II Task Force:

ENROLLED SENATE

BILL NO. 921

By: Jolley of the Senate

and

Jones of the House

An Act relating to schools; creating the Achieving Classroom Excellence II Task Force; stating issues that task force shall study; providing for membership, appointment, election of chair, quorum, staff support, and travel reimbursement; requiring compliance with Oklahoma Open Meeting Act and Oklahoma Open Records Act; directing task force to submit report of findings and recommendations by certain deadline; providing for noncodification; providing an effective date; and declaring an emergency.

BE IT ENACTED BY THE PEOPLE OF THE STATE OF OKLAHOMA:

SECTION 1. NEW LAW A new section of law not to be codified in the Oklahoma Statutes reads as follows:

A. There is hereby created to continue until December 31, 2007, the Achieving Classroom Excellence II Task Force. The task force shall study the following issues:

1. Comparison of the Priority Academic Student Skills with other states' curricular standards, primarily states that score highest on the National Assessment of Educational Progress (NAEP);

2. Alignment of the Priority Academic Student Skills with the National Assessment of Educational Progress (NAEP) standards;

Report of the ACE II Task Force

3. Feasibility of realigning the state performance level standards to NAEP performance level standards;

4. Differences in achievement levels among states based on exclusion rates on the NAEP; and

5. Feasibility of aligning the cut scores on state-mandated tests to NAEP cut scores.

B. The Achieving Classroom Excellence II Task Force shall consist of seven (7) members who shall be selected from among public and private school educators and members of the business community, but shall not include any elected officials, appointed by:

1. The Governor;
2. The President Pro Tempore of the Senate;
3. The Co-President Pro Tempore of the Senate;
4. The Speaker of the House of Representatives;
5. The Minority Leader of the House of Representatives;
6. Agreement of the Co-Chairs of the Senate Education Committee; and
7. The Chair of the House Education Committee.

C. Appointments to the task force shall be made by August 31, 2007. The member appointed by the Governor shall convene the first meeting of the task force by September 30, 2007. Members of the task force shall elect a chair from among the membership. A majority of the members of the task force shall constitute a quorum to transact business, but no vacancy shall impair the right of the remaining members to exercise all of the powers of the task force.

D. Staff support for the task force shall be provided by the State Senate and the Oklahoma House of Representatives.

E. Members of the Achieving Classroom Excellence II Task Force shall receive no compensation for serving on the task force, but shall be reimbursed by their respective appointing authorities for their necessary travel expenses incurred in the

Report of the ACE II Task Force

performance of their duties in accordance with the State Travel Reimbursement Act.

F. The proceedings of all meetings of the task force shall comply with the provisions of the Oklahoma Open Meeting Act and the Oklahoma Open Records Act.

G. The task force shall study the subject matter specified in subsection A of this section and submit a report of findings and recommendations to the Governor and Legislature by December 31, 2007.

SECTION 2. This act shall become effective July 1, 2007.

SECTION 3. It being immediately necessary for the preservation of the public peace, health and safety, an emergency is hereby declared to exist, by reason whereof this act shall take effect and be in full force from and after its passage and approval.

Passed the Senate the 8th day of March, 2007.

Passed the House of Representatives the 20th day of April, 2007.

Approved by the Governor the 30th day of April, 2007.