Cross-Grade Comparisons among Statewide Assessments and NAEP

William D. Schafer

Min Liu

Hsiu-fei (Joyce) Wang

University of Maryland

Presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, April 10, 2007[1]

This project was funded by the Maryland State Department of Education (MSDE) under a contract to the Maryland Assessment Research Center for Education Success (MARCES). The opinions expressed do not necessarily reflect those of MSDE or of MARCES.

[1] This draft is a slightly corrected version from original one we presented in AERA. The main difference lies in the data used for grade 10 in North Carolina's state assessment: the old version used reading and math results for grade 10 in NC while the revised one uses Algebra I and English I instead because they are the indicators for AYP adopted by NC.

Cross-Grade Comparisons among Statewide Assessments and NAEP

Research Purpose

Several studies have addressed the (in)consistencies from state to state in the achievement levels established for the "No Child Left Behind (NCLB)"-mandated "proficiency" label. McLaughlin & Mello (2002) compared state results with those from National Assessment of Educational Progress (NAEP), Later, McLaughlin (2005) located states' primary-level "proficiency" standards for reading on the NAEP scale, Linn (2005a) presented graphical analyses showing state percents proficient for mathematics against NAEP results for 33 states, and Fuller, Gesicki, Kang, & Wright (2006) tracked state and NAEP percents proficient over time in 12 states. However, none of these has compared the overall state results in terms of percent proficient or better with the National Assessment of Educational Progress (NAEP) estimates for basic or better (or proficient or better) for a very broad sample of states.

Our initial goal was to collect for as many states as possible for the then-latest-available year (2005), percent proficient or better data in both reading (or English language arts) and mathematics and to present the results in graphic form. The presentation may be used by states to compare their results with those of NAEP or to compare their proficiency levels with those of other states with common references. They may also be used by researchers who are interested in comparing states with differing degrees of expectations for their students' achievement in order to satisfy NCLB's requirements on outcomes such as indicators of instructional change.

We also were interested in ways to index and compare features of state data across states, grades, and contents. These include such characteristics as degrees of moderation, degrees of idealism vs. realism of the state proficiency standards, and slopes of trend lines.

Background

The No Child Left behind (NCLB) Act requires that all states develop an annual educational assessment and accountability system in mathematics and reading/language arts in each of grades 3 through 8 and at least once in grade 10 through 12 by 2005-2006. NCLB also requires states to set challenging academic content standards and student academic achievement standards (i.e. performance standards) for their statewide assessments. To comply with NCLB, states must set at least three performance levels for their assessments: the first two, proficient and advanced are to correspond to high levels of achievement while the third level, basic, can provide a way to monitoring process toward the proficient level (Linn, 2005b).

Based on state-defined performance standards, all schools and districts are to be reviewed annually in terms of the percentages of students who achieved the proficient level or above and the results compared with the previous year. There are severe sanctions for schools that do not meet adequate yearly progress (AYP) targets for two or more years in a row. The AYP is set to assure that schools, districts and states will have 100% of the students at the proficient level or above by 2013-2014 as required by NCLB.

It is financially quite important for states to set appropriate performance standards and cut scores across grades for their statewide assessments, but accepted guidance on how to do that is lacking. It is not surprising to find that the state performance standards have a great state-to-state variability (Linn, 2005a). It is clearly not sensible to compare achievement results state by state, directly.

Thus, if we want to compare the performance standards of different statewide assessments, we need to find a common assessment related to them (McLaughlin & Mello, 2002). NAEP has been described as a "gold standard" for monitoring the educational progress of American students (Jones, Olkin, & American Educational Research Association., 2004) and it is "the only national assessment of achievement based on defensible samples" (Shepard, Glaser, Linn, & Bohrnstedt, 1993).

One of the fundamental goals for NAEP is that it could provide consistent information about student achievement. McLaughlin and Mello (2002) have compared math achievement results in different states by estimating the NAEP scale scores that correspond to state performance standards. But the state-level data in their analyses came from a sample of school-level state assessment scores. In our research, all of our state assessment data came directly from each state's department of education, which is more inclusive, and our analyses include reading as well as math.

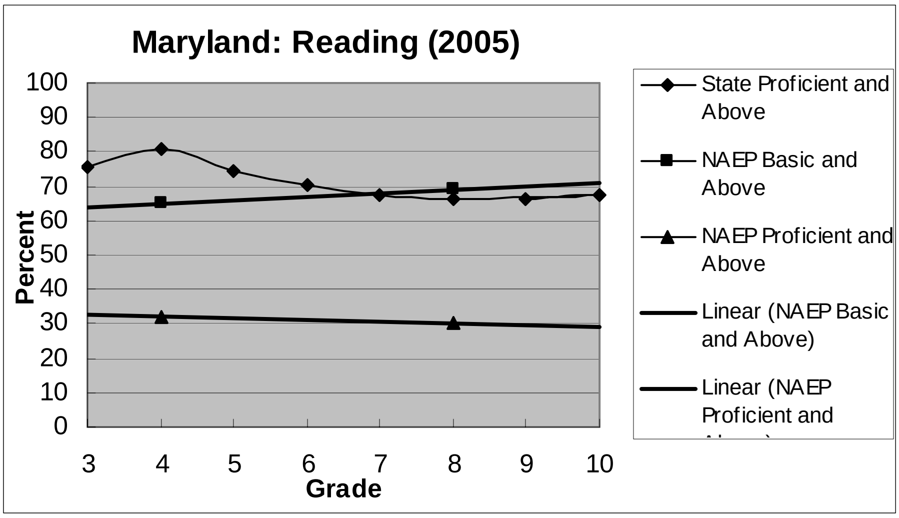
There are four achievement levels for three grades (4, 8 and 12) assessed by NAEP: Below Basic, Basic, Proficient, and Advanced. However, Pellegrino, Jones, and Mitchell (1999) have concluded that "collection of meaningful NAEP data in the twelfth grade is problematic given the insufficient motivation of high school seniors and their highly variable curricula and dropout rates," and the National Research Council (NRC) committee recommended that the NAEP should assess students in grades 10 or 11 instead of 12 (Pellegrino, Jones, & Mitchell, 1999). Therefore, we only used NAEP data for grade 4 and 8, ignoring grade 12.

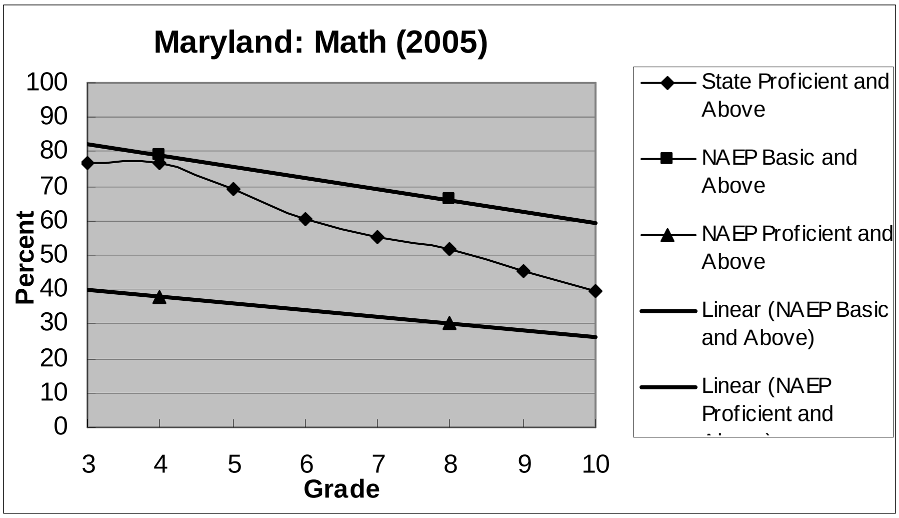
Of the four achievement levels reported by NAEP, any of them may be comparable to the term "proficient" as used in NCLB. Although the term "proficient" is used in NCLB and one might conclude that the intent of the Act was to convey a comparable degree of expectation for states, Mosquin and Chromy (2004) found state results most comparable with the NAEP Basic achievement level. In its report of the 2003 reading data, NCES (2003) focused attention on the NAEP Proficient achievement level. However, in a later, similar report, NCES (2005) focused attention on the NAEP Basic achievement level. We do not wish to enter this argument, so our comparisons used both NAEP Proficient and NAEP Basic as benchmarks for states.

Methodology

State assessment results for 2005 were compared graphically with NAEP (actual and extrapolated) across grades using a format adapted from Schafer (2005). All the NAEP data came from NAEP's "data explorer" webpage, http://nces.ed.gov/nationsreportcard/nde/criteria.asp, while state performance data were from the website of each state's department of education. The reader can check (and renew) the data by using corresponding links attached below each graph. All of these data are from 2005, the most recent year. We produced graphs for reading and math separately for each state.

We generated two graphs for each state. The graphs of Maryland are shown as a typical example.





Data source:

http://msp.msde.state.md.us/downloadindex.aspx?K=99AAAA

The data came from the Maryland School Assessment (MSA), which is a test of reading and math achievement that meets the requirements of the federal No Child Left Behind Act. The test is given each year in early March at grades 3 through 8. For MSA, there are three achievement levels: Advanced, Proficient, and Basic.

The first line labeled as "State Proficient and Above" was generated by a set of plots which represent Maryland's percentages of students who are proficient or above at each grade evaluated by the state assessments. During data collection, we found it common that we could not get state data for every grade, i.e., from grade 3 through grade 10, and therefore we interpolated or extrapolated plot points for those missing values using a linear relation between the students' performance percentage and grade level. Here, for the "State Proficient and Above" line of Maryland Reading, we interpolated one plot point, grade 9. For the "State Proficient and Above" line of Maryland Math, two plot points, grade 9 and 10, were found by extrapolation. The second line labeled as "NAEP Basic and Above" represents the percentage for students who are at the basic or above level in Maryland at each grade evaluated by NAEP. Similarly, the third line shows NAEP's percentage for Maryland's students who are proficient or above at each grade. Since we could use only NAEP data for grade 4 and grade 8 (the only tested grades in our range), we connected and extended these two plot points to get trend lines, which help to show the difference between state assessment and NAEP assessment. (The last two labels in the legend are for these NAEP trend lines). We can read all the state graphs in a similar way.

Results

I. 43 State Graphs

Data were available for 43 of the 50 states. Seven states' web sites did not provide sufficient information to include them in the study. Each Figure is numbered by state and further described using the state two-character acronym along with the content, reading (r) or math (m). The order is alphabetical by acronym.

Figure 1. AKr

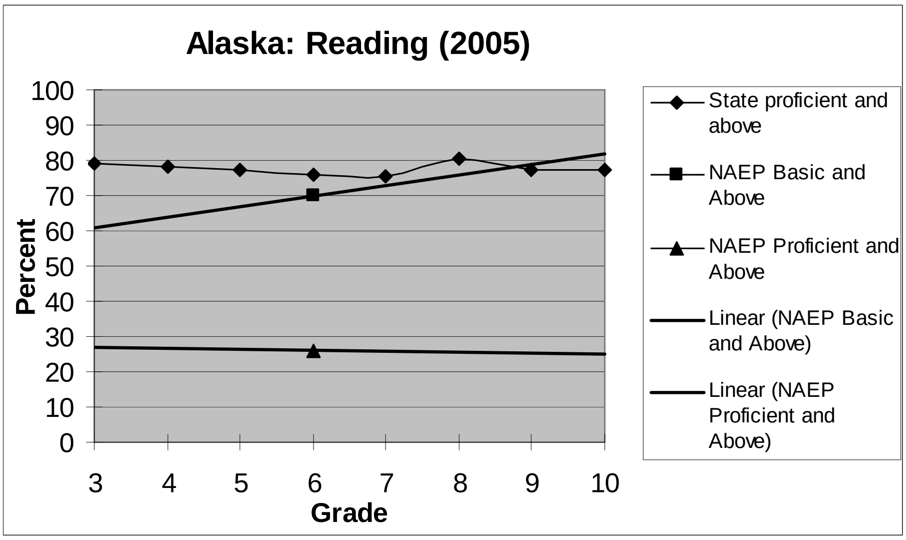
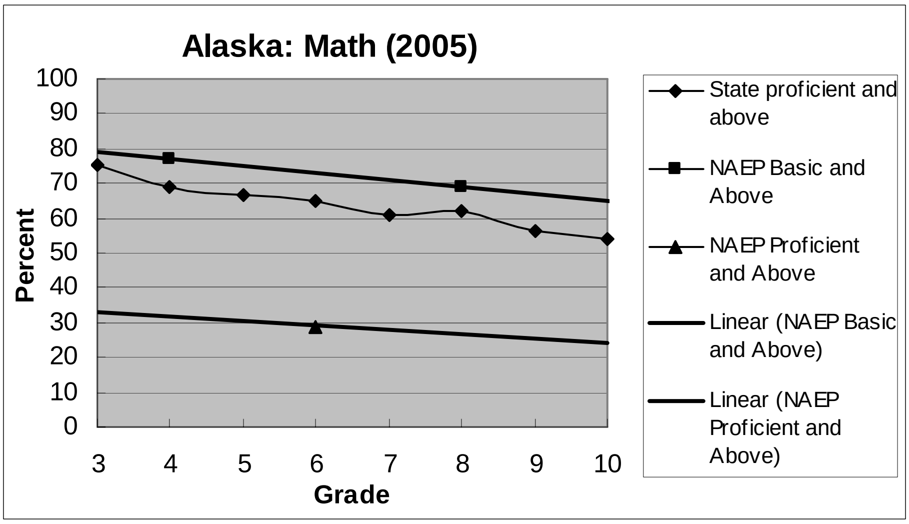


Figure 1.AKm



Data source:
http://www.educ.state.ak.us/tls/assessment/results/2005/2005StatewideSBA.pdf
The data come from Alaska Standard Based Assessments (SBAs) of the Spring 2005. There are four proficiency levels for SBAs: Advanced, Proficient, Below Proficient, and Far Below Proficient. We extrapolated Grade 10 in terms of Grade 3 to Grade 9 for both reading and math graphs.

Figure 2. ALr

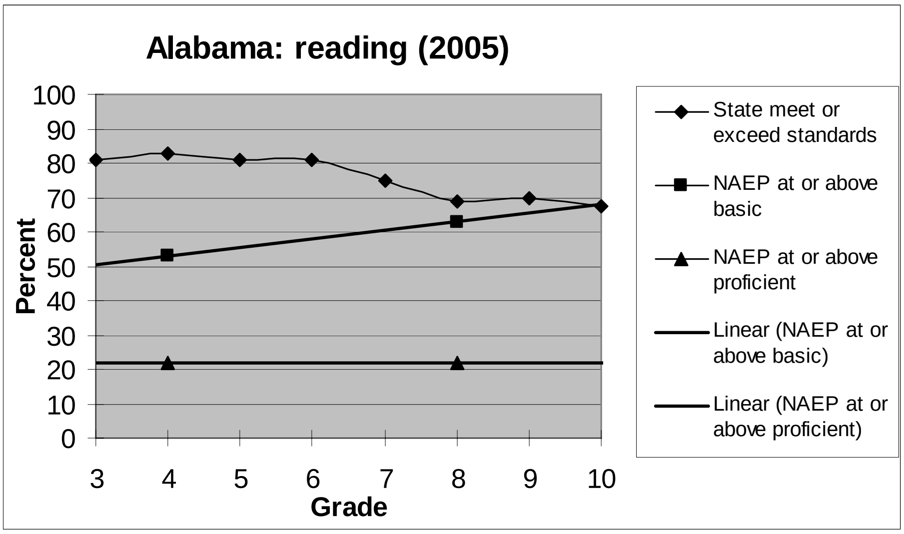
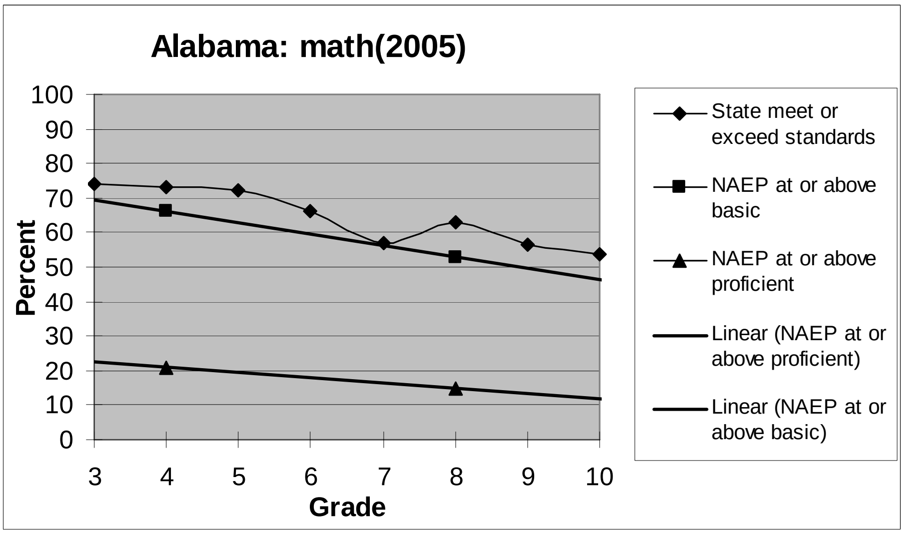


Figure 2. ALm



Data source:
http://www.alsde.edu/Accountability/2005Reports/AL2005ARMT_0624051.pdf?lstSchoolYear=3&lstReport=2005Reports%2FAL2005ARMT_0624051.pdf
Data come from the Alabama Reading and Mathematics Test (ARMT). Academic achievement levels define how well students master the state's academic content standards. The State Board of Education adopted four achievement levels for ARMT (http://www.alsde.edu/Accountability/2005Reports/AccountabilityFAQs2005.pdf?lstSchoolYear=3&lstReport=+2005Reports%2FAccountabilityFAQs2005.pdf): Level IV – Exceeds Academic Content Standards; Level III – Meets Academic Content Standards; Level II – Partially Meets Academic Content Standards; and Level I – Does Not Meet Academic Content Standards. For purposes of NCLB, Level III is considered proficient, or at grade level. Therefore, a student scoring at Level III or Level IV meets the proficiency standard required of NCLB. A student scoring at

Level I or Level II does not meet the proficiency standard required of NCLB. Here, for both reading and math graphs, we extrapolated 2 plots in the first lines, grade 9 and grade 10, from grade 3 to grade 8.

Figure 3. ARr

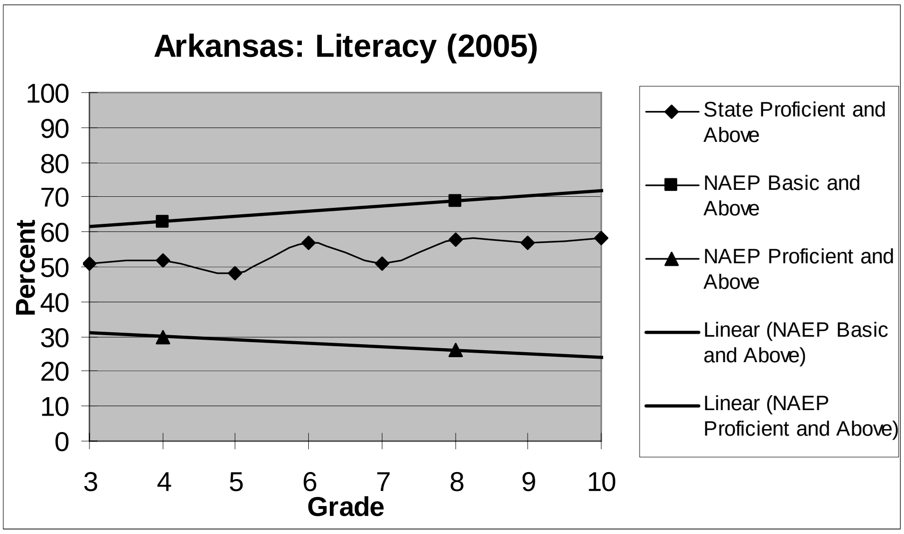
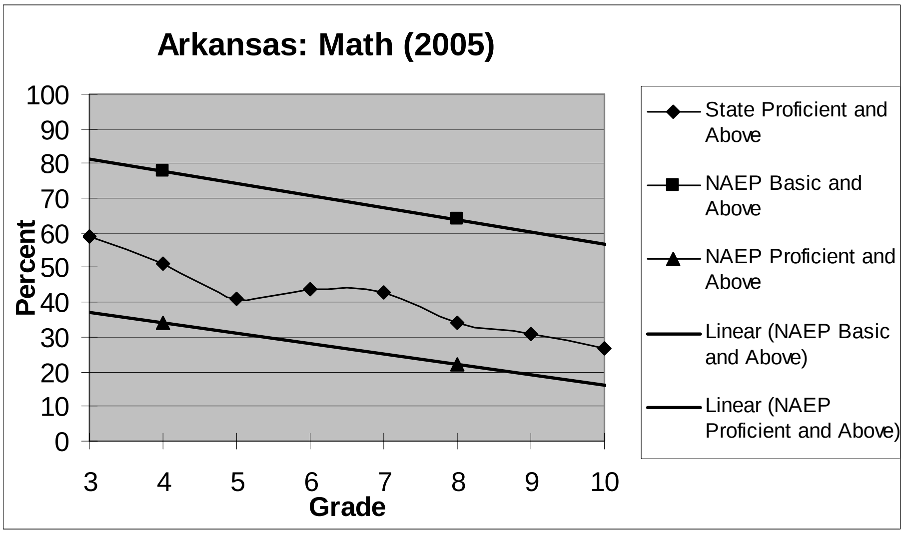


Figure 3. ARm



Data source:
http://130.184.43.9/reportcards/state05.php
In Arkansas, criterion-referenced tests (CRT) on Literacy (Comparable for Reading in other states) and Math are administered at Grades 3- 8 (Benchmark Exams). Students' achievement levels are: Advanced, Proficient, Basic and Below Basic. Data from Grades 9 and 10 were obtained by extrapolation based on data of grades 3 to 8.

Figure 4. AZr

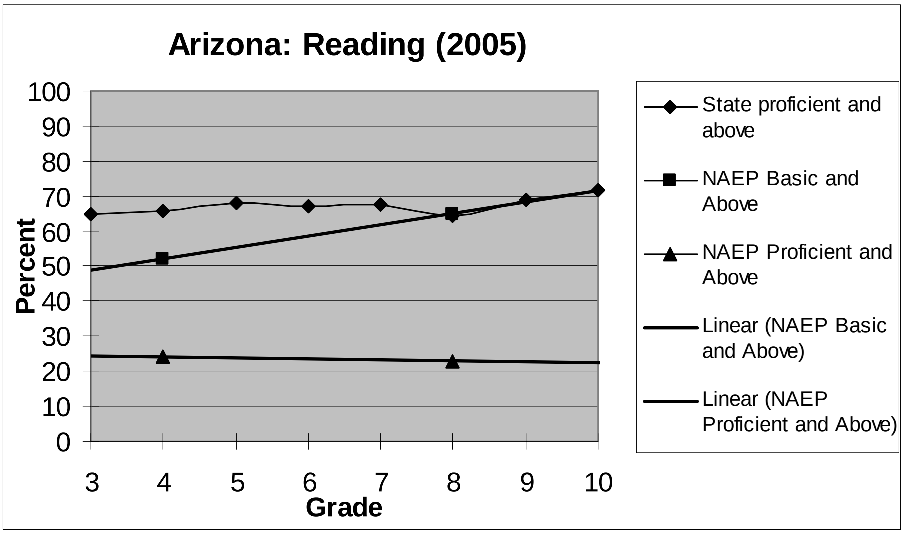
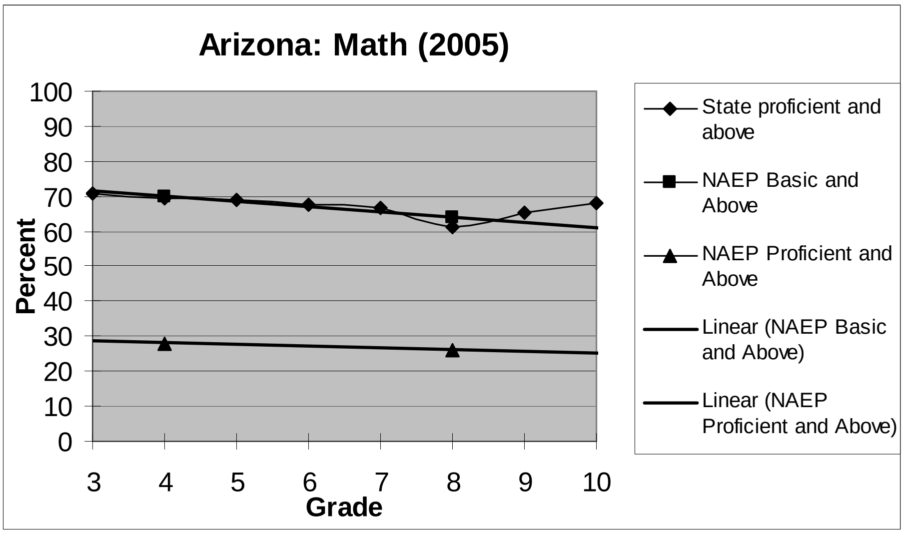


Figure 4. AZm



Data source:
http://www.ade.state.az.us/srcs/statereportcards/statereportcard04-05.pdf
Data come from the 2004-2005 State Report Card, which was provided by Arizona Department of Education as part of compliance with the federal No Child Left Behind Act. Arizona's Instrument to Measure Standards (AIMS) provides educators and the public with information regarding the progress of Arizona's students toward mastering Arizona's reading, writing and mathematics standards. There are four performance levels for AIMS: Falls Far Below Standards, Approaches Standards, Meets Standards, and Exceeds Standards. Students falling in the Meet Standards level or the Exceeds Standards level are considered as “have passed” AIMS and “reached” proficient level. Here, we interpolated Grade 4, 6, 7 and 9 from Grade 3, 5, 8 and 10 for both reading and math.

Figure 5. CAr

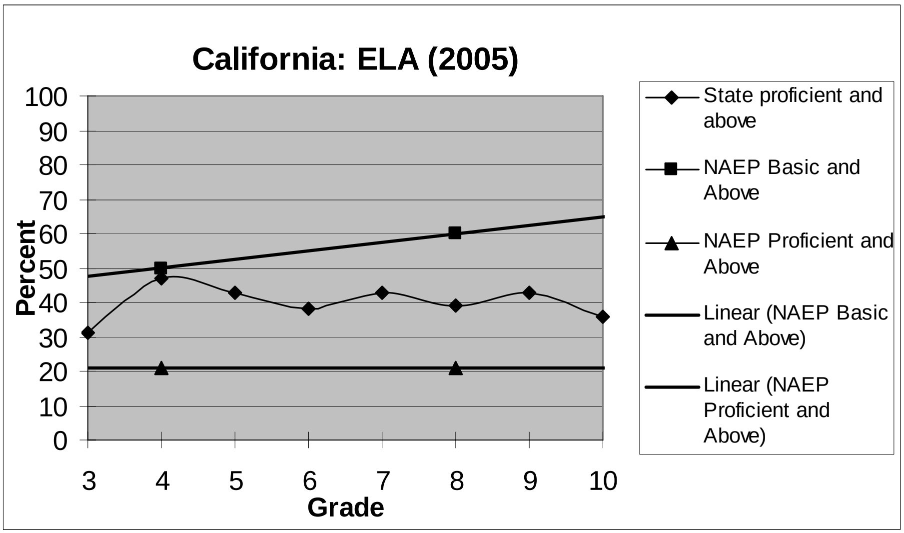
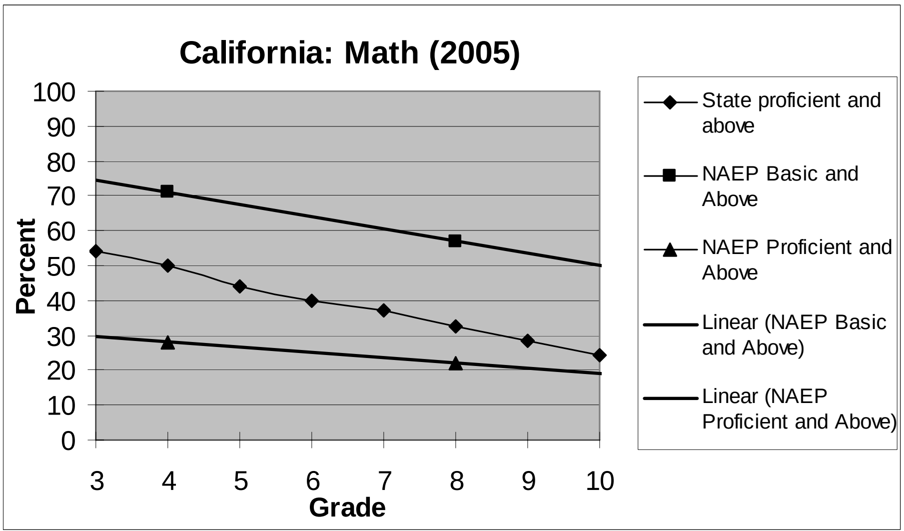


Figure 5.CAm



Data source:
http://star.cde.ca.gov/star2005/viewreport.asp?ps=true&lstTestYear=2004&lstTestType=C&lstCounty=&lstDistrict=&lstSchool=&lstGroup=1&lstSubGroup=1
The California Standards Tests (CST) in English-language arts, mathematics, science, and history-social science are administered to students in California public schools. Except for a writing component that is administered as part of the grade four and seven English-language arts tests, all questions are multiple-choice. Students' performances are categorized into five levels: Advanced, Proficient, Basic, Below Basic, and Far Below Basic. The 2005 CSTs in English-Language Arts were required for students who were enrolled in Grade 2 to 11; CSTs in Math were required for students who were enrolled in Grade 2~9. Data for English-language arts are complete, while for math, data on grades 8 and 10 are interpolated (extrapolated) based on data from grades 3-7.

Figure 6. COr

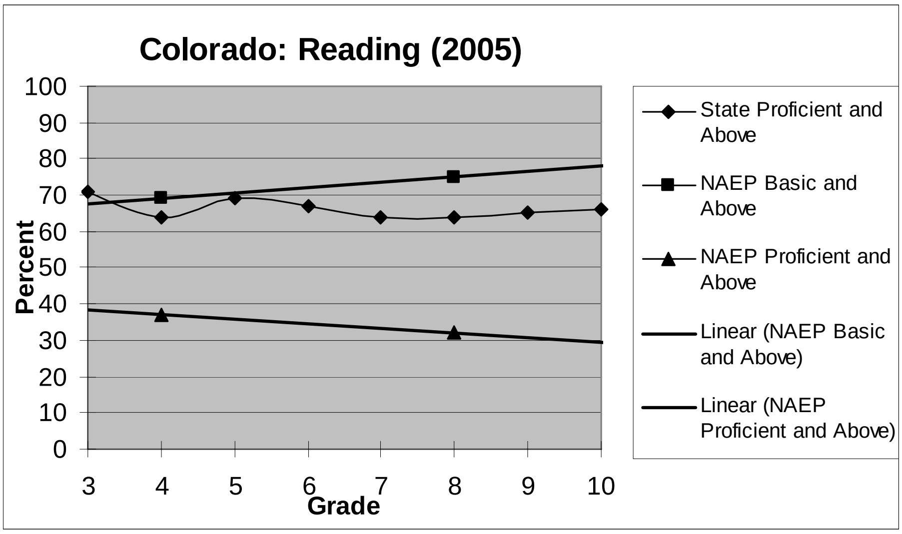
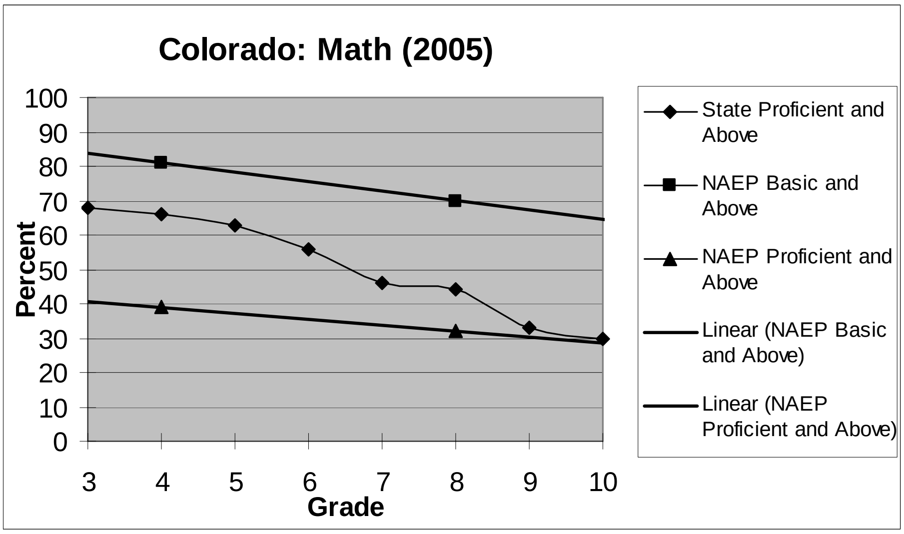


Figure 6. COm



Data source:
http://www.cde.state.co.us/cdeassess/documents/csap/csap_summary.html
The Colorado Student Assessment Program (CSAP) is administered in grades 3 through 10. Students' achievement levels consist of "Advanced" (Performance Level 4), "Proficient" (Performance Level 3), "Partially Proficient" (Performance Level 2) and "Unsatisfactory" (Performance Level 1). All data of 2005 from Grade 3 to 10 were available and no interpolation/extrapolation was needed.

Figure 7.DEr

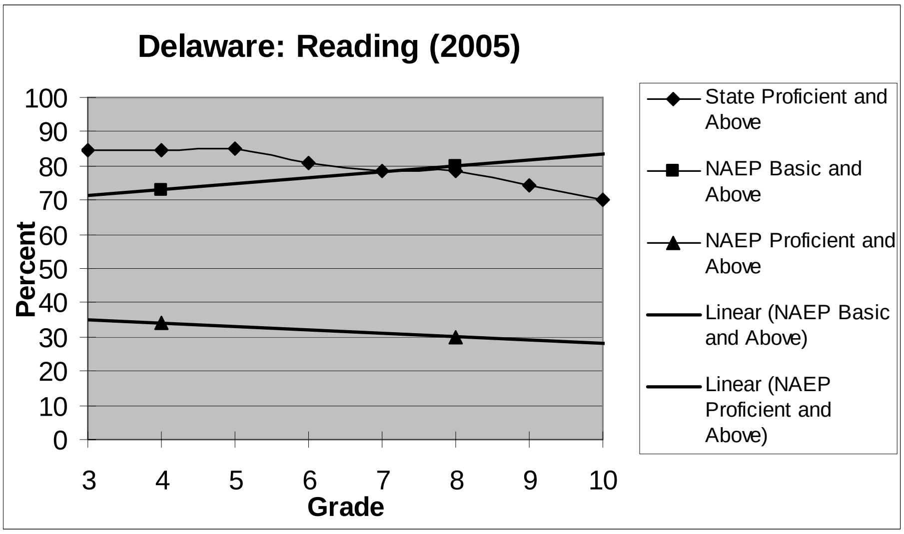
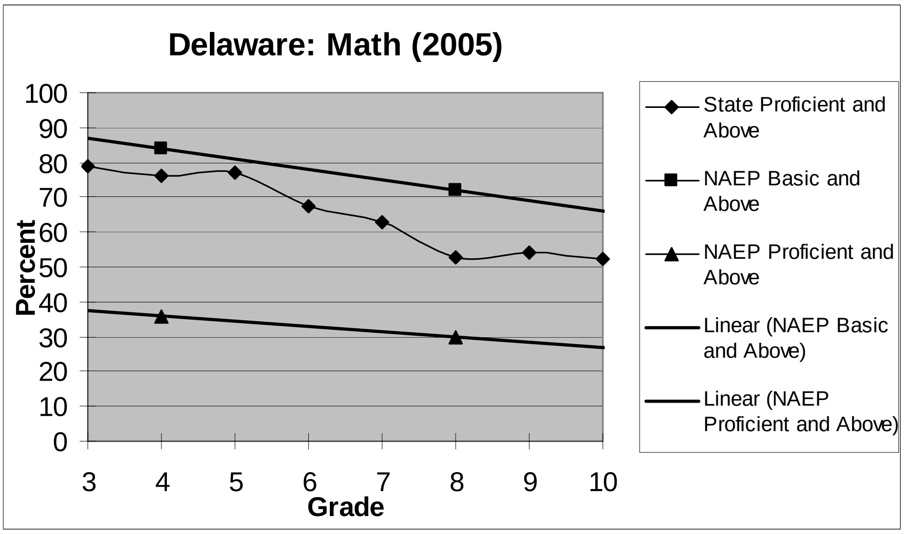


Figure 7. DEm



Data source:
http://www.doe.state.de.us/files/pdf/de_edreportcard200405.pdf
Data comes from Delaware Education State Report Card in 2004-2005. The Spring Delaware Student Testing Program (DSTP) was conducted for reading, mathematics and writing in grades 3, 5, 8 and 10; as well as the Fall and Spring 2004 science and social studies DSTP for grades 4, 6, 8 and 11. The DSTP Student Performance Levels include five categories: level 1-Well Below the Standard (Needs Significant Improvement); level 2- Below the Standard (Needs Improvement); level 3- Meets the Standard (Good Performance); level 4-Exceeds the Standard (Very Good Performance); level5- Distinguished (Excellent Performance). For NCLB purpose, level 3 and level 4 are considered as “proficient” level. Here, we interpolated grades 4, 6, 7 and 9 from grades 3, 5, 8 and 10 for both reading and math.

Figure 8. FLr

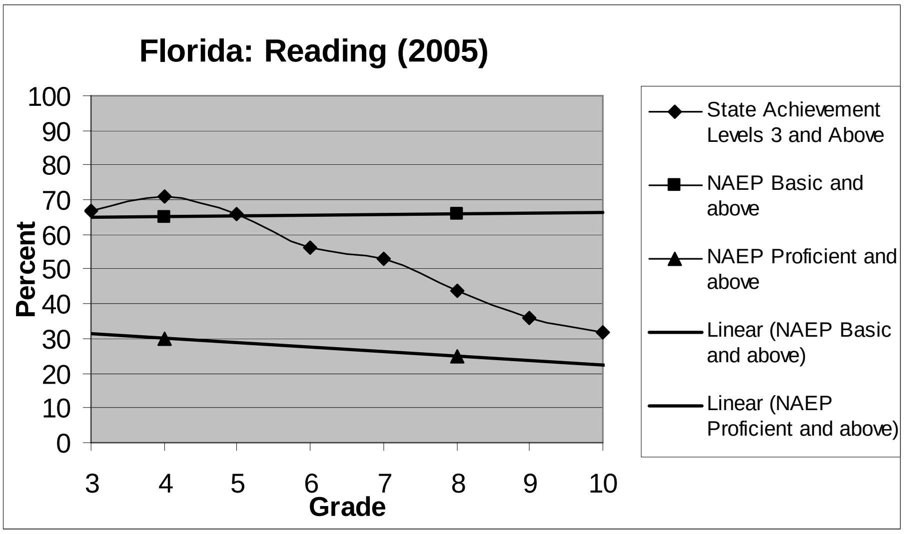
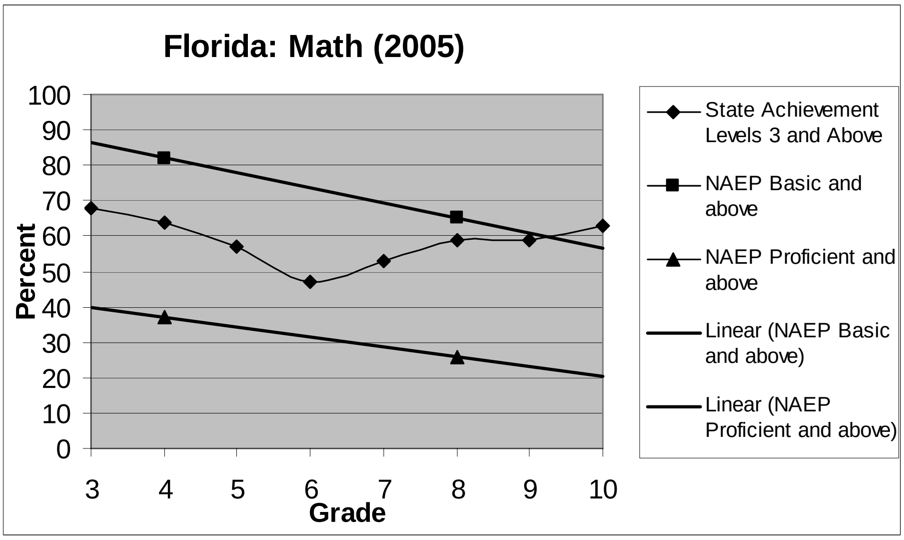


Figure 8. FLm



Data source:
http://fcat.fldoe.org/index.asp#reports
The FCAT (Florida Comprehensive Assessment Test), administered to students in grades 3-11, contains two basic components: criterion-referenced tests (CRT), measuring selected benchmarks in Mathematics, Reading, Science, and Writing from the Sunshine State Standards (SSS); and norm-referenced tests (NRT) in Reading and Mathematics, measuring individual student performance against national norms. Student achievement levels are numbered from Level 1 to Level 5 (the larger the number, the higher the achievement level). For NCLB purpose, level 3 and above are considered as "proficient" level. Since all data for 2005 are complete, no interpolation/extrapolation was needed.

Figure 9. GAr

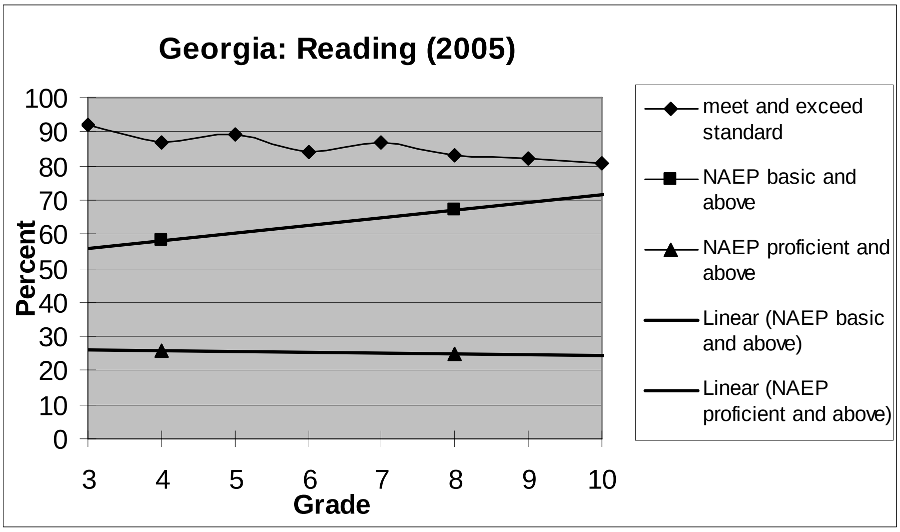
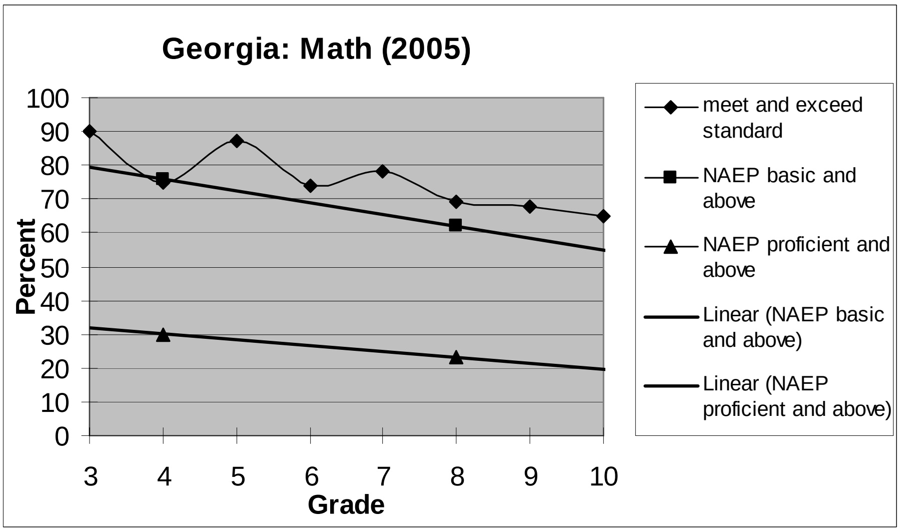


Figure 9. GAm



Data source:
http://reportcard2005.gaosa.org/k12/reports.aspX?ID=ALL:ALL&TestKey=C*4&TestType=qcc
The Georgia Criterion-Referenced Competency Tests (CRCT) are state-mandated end-of-year assessments. These tests measure the content and skills required by Georgia's Quality Core Curriculum (QCC) in the areas of reading, English/language arts, and mathematics in grades 1 through 8 and in science and social studies in grades 3 through 8. The achievement level can be classified as: Level 1,"Does Not Meet Standard"; Level 2 , "Meets Standard"; Level 3, "Exceeds Standard" and a Level 4. For NCLB purpose, level 3 and level 4 are considered as "proficient" level. For both reading and math graphs, we extrapolated 2 plots in the first lines, grade 9 and grade 10 from grades 3 to Grade 8.

Figure 10. HIr

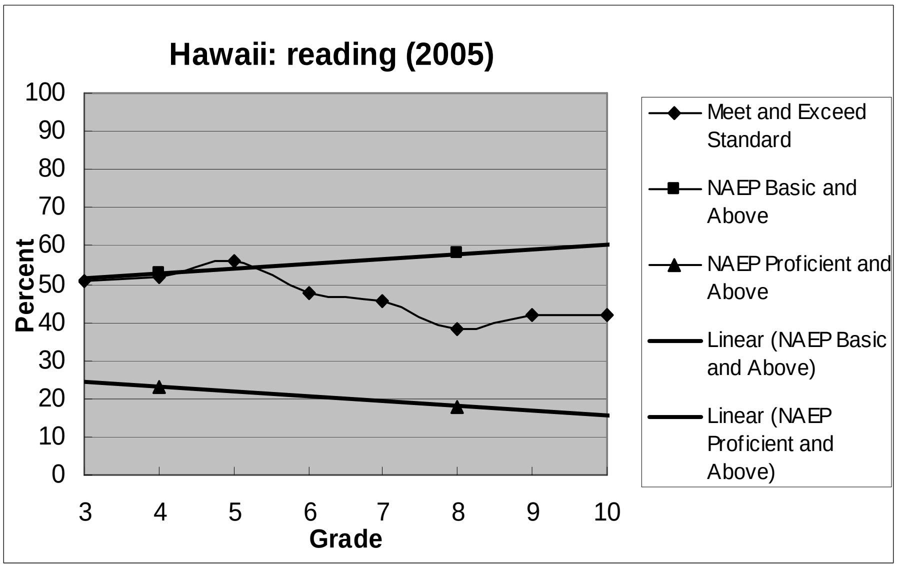
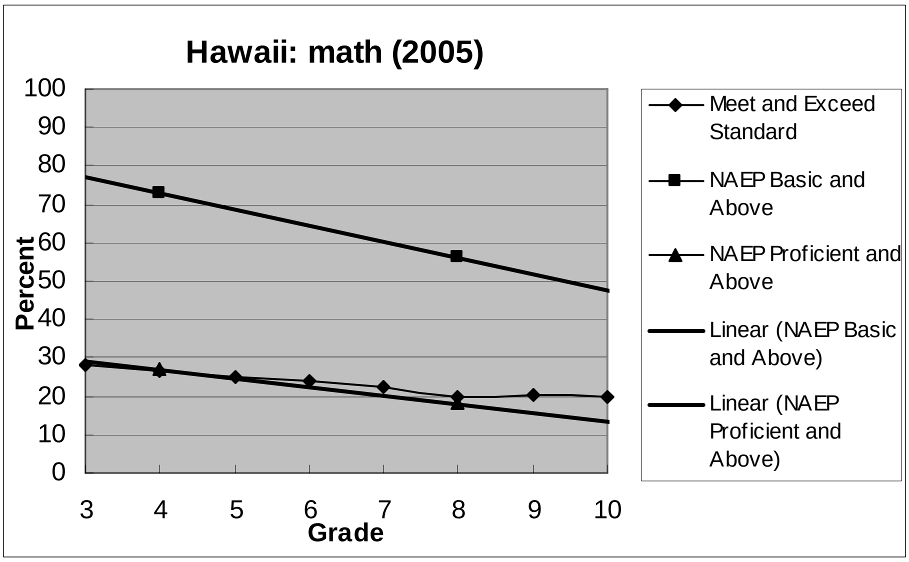


Figure 10.HIm



Data source:
http://arch.k12.hi.us/, then choose "Hawaii State Assessment: 2005 State Results"
The Hawaii Statewide Assessment (HSA) includes customized standards-based items aligned to Hawaii Content and Performance Standards II as well as components of the Stanford Achievement Test, 9th Edition. These results are compiled from the Spring, 2005 administration of the HSA. Students' performances are categorized into four levels: Exceed, Meet, Approaches, and Well Below. For NCLB purpose, meet and exceed standard are considered as "proficient." Data are available for Grade 3, 5, 8 and 10, while the data for Grade 4, 6, 7 were obtained by interpolation based on data from the other grades.

Figure 11. IDr

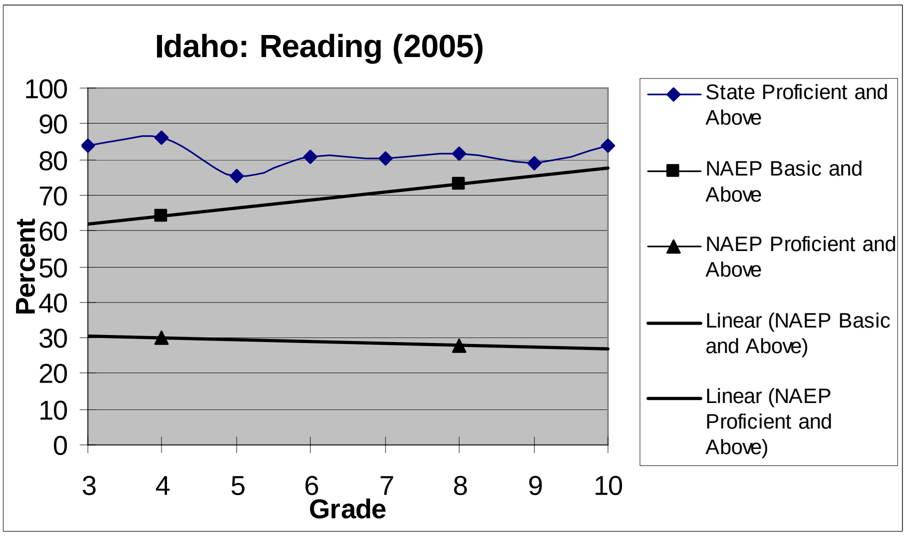
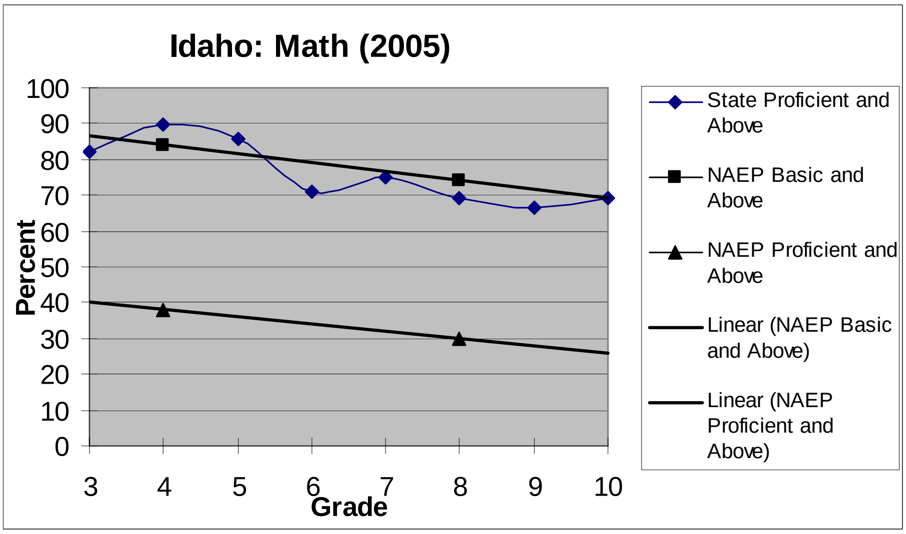


Figure 11. IDm



Data source:
http://www.sde.state.id.us/ipd/aypassessment05/default.asp
Achievement standards have been created in six subject areas for years K-12: math, science, social studies, language arts/communications, health and humanities. Districts reported student performance in four levels of performance: below basic, basic, proficient, and advanced. Students performing at the advanced and proficient levels met or exceeded the standards. Students performing at the basic and below basic levels did not meet the standards. Data for grade levels 3 to 8 and 10, were available for 2005. Data for grade 9 were unavailable, so linear interpolation was used to generate the results.

Figure 12. ILr

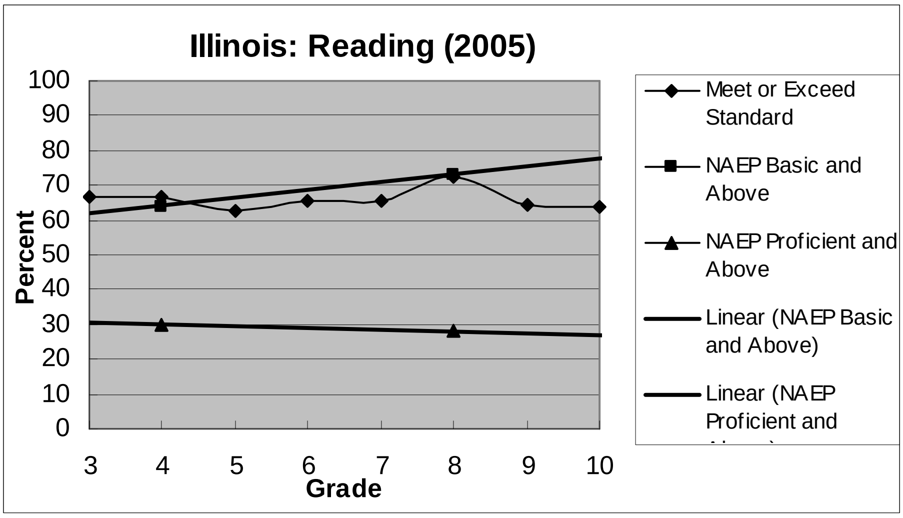
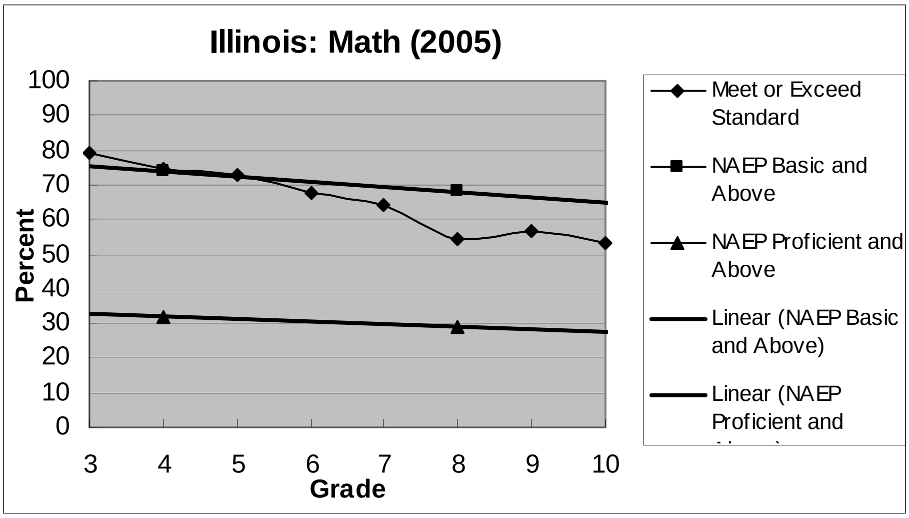


Figure 12. ILm



Data source:
http://webprod1.isbe.net/ereportcard/publicsite/getSearchCriteria.aspx
Data come from 2005 Illinois State Report Card which is required by state and federal laws. The Spring 2005 Illinois Standards Achievement Test (ISAT) was administered to students at grades 3, 5, and 8 in reading and mathematics. It can give us information about students' attainment of the Illinois Learning Standards. Students are categorized into four achievement levels: Exceeds Standards (Advanced knowledge and skills), Meets Standards (Proficient knowledge and skills), Below Standards (basic knowledge and skills), and Academic Warning (limited knowledge and skills). We interpolated grades 4, 6, 7, 9 and 10 from grades 3, 5, 8, and 11 for both reading and math.

Figure 13. INr

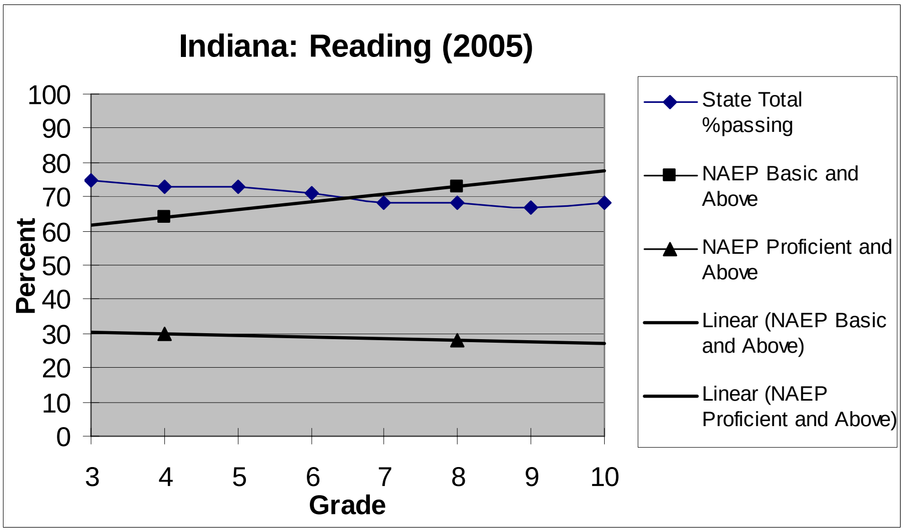
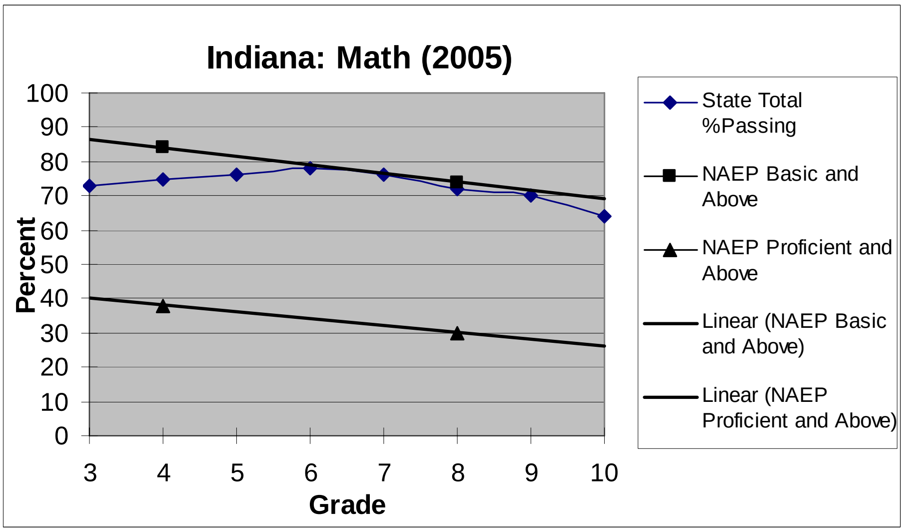


Figure 13. INm



Data source:
http://www.doe.state.in.us/reed/newsr/2005/12-December/051214prOverview.pdf
Indiana Statewide Testing for Educational Progress-Plus (ISTEP+) results provide the number of state total percent passing, as indicated the percentage of students who met the state's academic standards and "proficient" level required by NCLB in grades 3-10 English/language arts and mathematics. Since data for all grade levels, 3 through 10, were available, linear interpolation and extrapolation were not used to generate results.

Figure 14. KSr

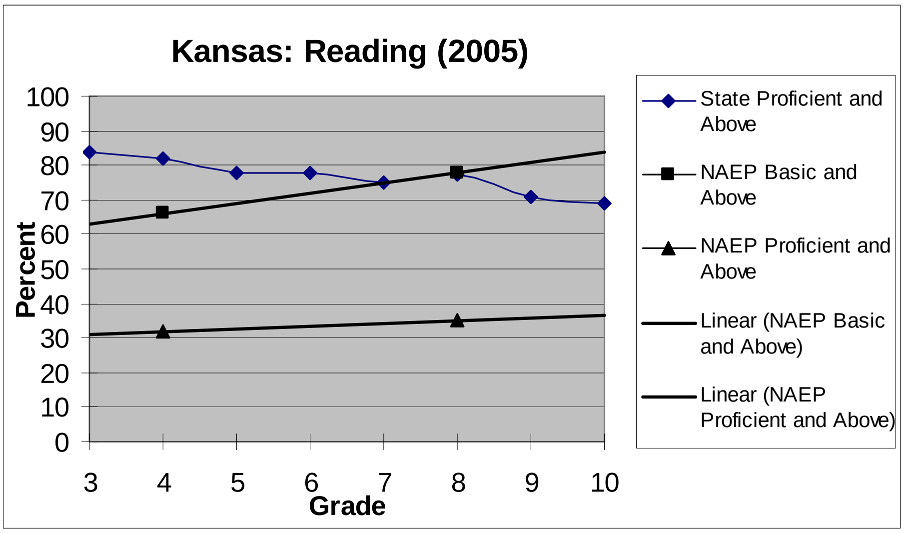
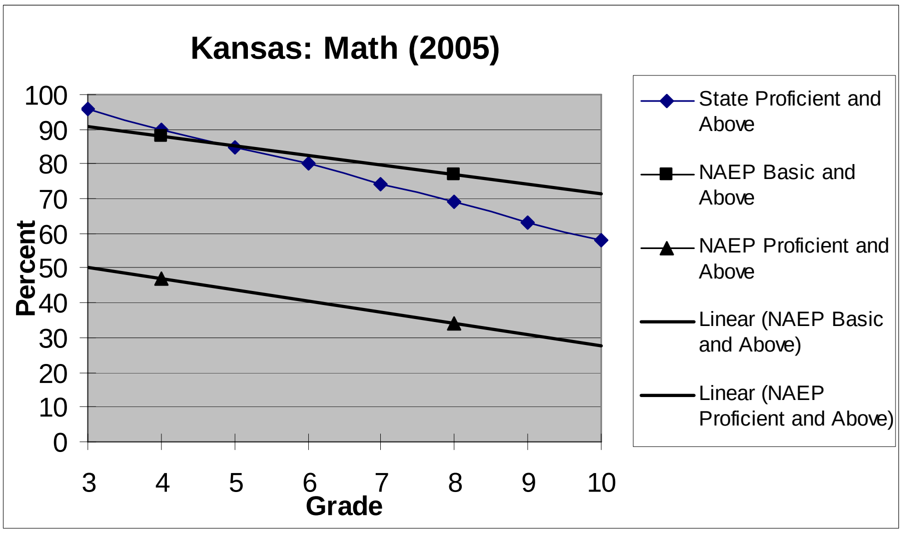


Figure 14. KSm



Data source:
http://www.ksde.org/assessment/state_tables_2005.pdf
The Kansas assessments—regular, alternate scored against grade level standards, and alternate scored against alternate standards—report student progress in five performance levels including: exemplary, advanced, proficient, basic and unsatisfactory. Students performing at exemplary, advanced and proficient levels met or exceeded the standards. Students performing at the basic and unsatisfactory levels did not meet the standards. In 2005, all Kansas public school students in grades 5, 8 and 11 were assessed on state-approved mathematics and reading standards. However, data for grades 3,4,6,7,9,10 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results.

Figure 15. KYr

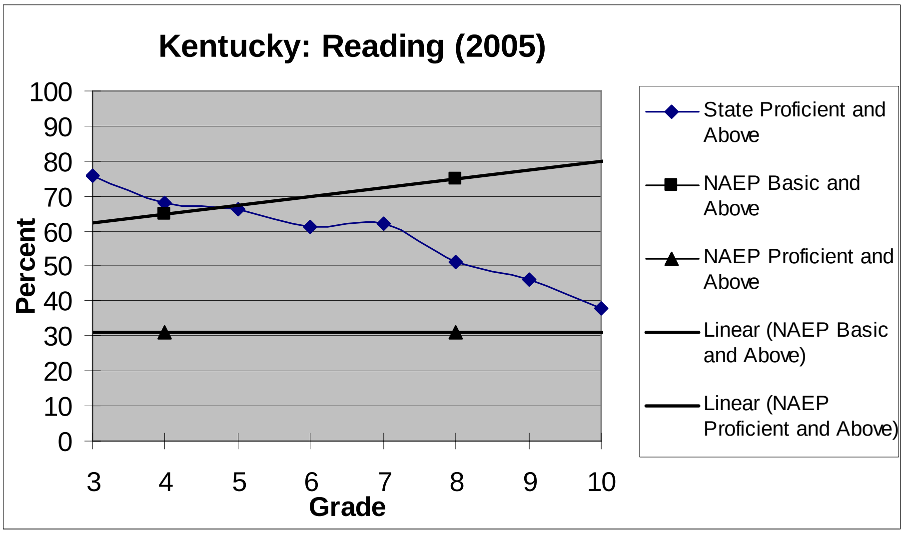
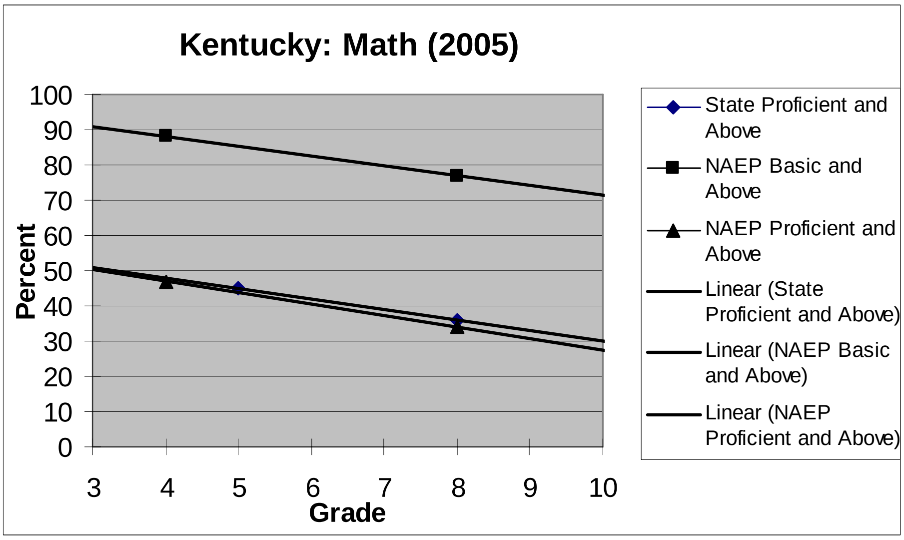


Figure 15. KYm



Data source:
http://apps.kde.state.ky.us/secure_cats_reports_05/index.cfm?fuseaction=main.display_regionstate
In Spring 2005, Kentucky Core Content Test (KCCT) were administered in selected grades. For Reading these are Grade 4, 7, and 10; while for Math they are Grade 5 and 8. Students' achievements were categorized into four levels: Distinguished, Apprentice, Proficient, and Novice. Data for Reading for Grade 3, 5, 6, 8 and 9 were interpolated based on data of Grade 4, 7 and 10. Data for Math or Grade 3, 4, 6, 7, 9, and 10 are interpolated (extrapolated) based on data from grades 5 and 8, which therefore constitutes the three straight lines on Math graph.

Figure 16. LAr

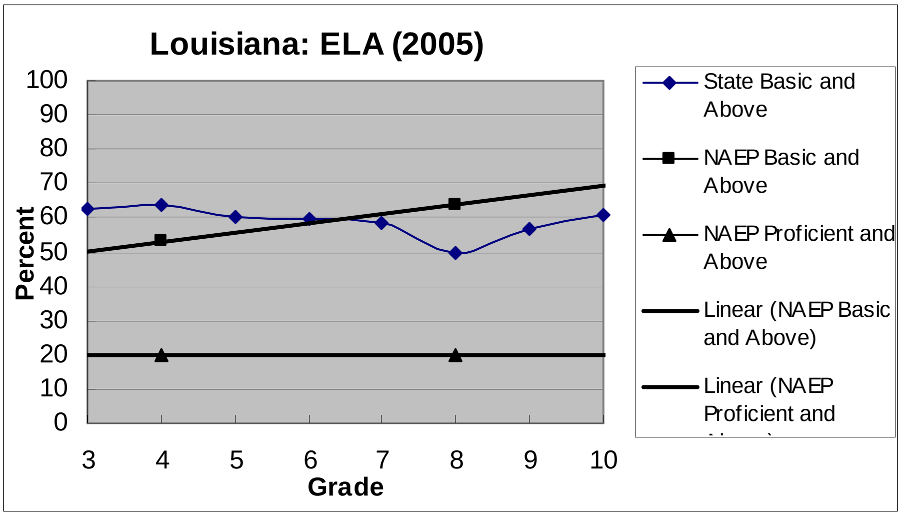
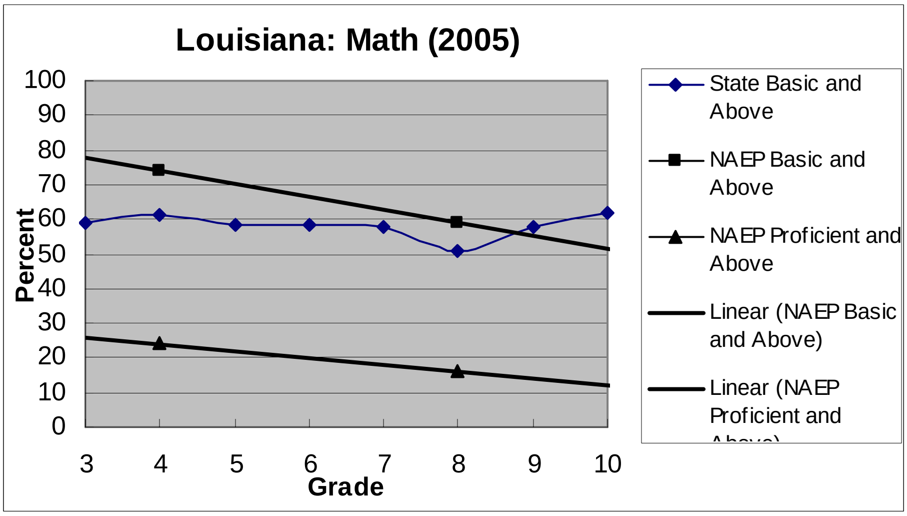


Figure 16. LAm



Data source:
http://www.doe.state.la.us/lde/uploads/7714.pdf
Our data come from the Spring 2005 Louisiana's Criterion-referenced testing (CRT), which is constituted by the Louisiana Educational Assessment Program for the 21st Century (LEAP 21) and the Graduation Exit Examination for the 21st Century (GEE 21). LEAP is administered in the 4th and 8th grades while GEE in the 10th and 11th grades. The GEE determines whether students are eligible to graduate from high school while LEAP is used to determine whether students advance to the 5th and 9th grades. There are five achievement ratings students can attain: Unsatisfactory (don't have fundamental knowledge and skills), Approaching Basic (partially fundamental knowledge and skills), Basic (fundamental knowledge and skills), Mastery (competency over challenging subject matter and well prepared for the next level of schooling), and Advanced (superior performance beyond the proficient level of mastery). The state's goal is for all students to achieve at the Basic level. For NCLB purpose, Basic level and above are considered "proficient". State data were available only for Grade 4, 8 and 10, and therefore we interpolated other grades in each graph.

Figure 17. MAr

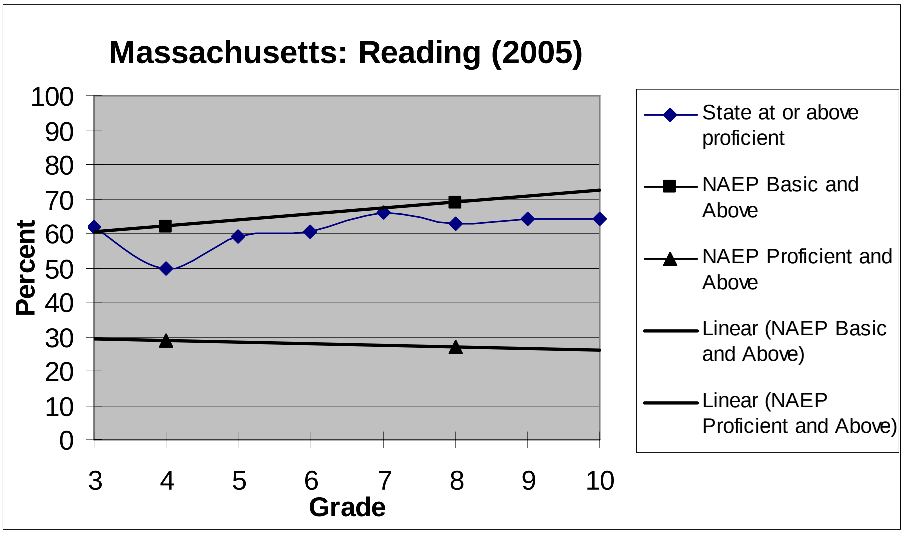
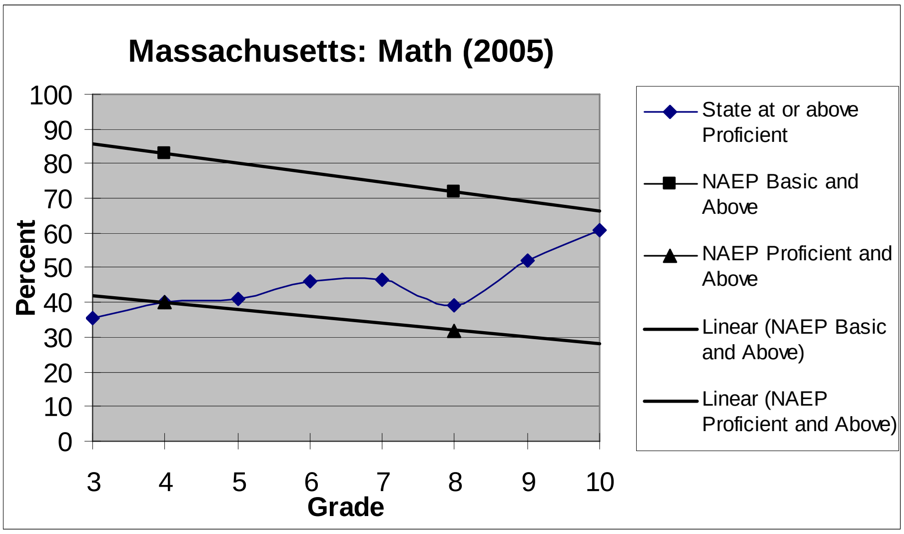


Figure 17. MAm



Data source:
http://www.doe.mass.edu/mcas/2005/results/summary.pdf
The Massachusetts Comprehensive Assessment System (MCAS), the Commonwealth's statewide assessment program for students educated with Massachusetts public funds, was developed to fulfill the requirements of the Education Reform Law of 1993. Performance levels include advanced, proficient, needs improvement, and warning. Students performing at proficient and advanced levels met or exceeded the standards. In the spring of 2005, 521,635 Massachusetts students in grades 3, 4, 5, 6, 7, 8, and 10 participated in the administration of the MCAS tests. Reading was administered at grade 3 only; English language arts were at 4, 7 and 10 grades; therefore, data for grades 5, 6, 8, and 9 in reading were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades. On the other hand, in 2005 mathematics was

administered at 4, 6, 8 and 10 grades. Thus, linear interpolation and extrapolation were used to generate the results for 3, 5, 7, and 9 grades.

Figure 18. MDr

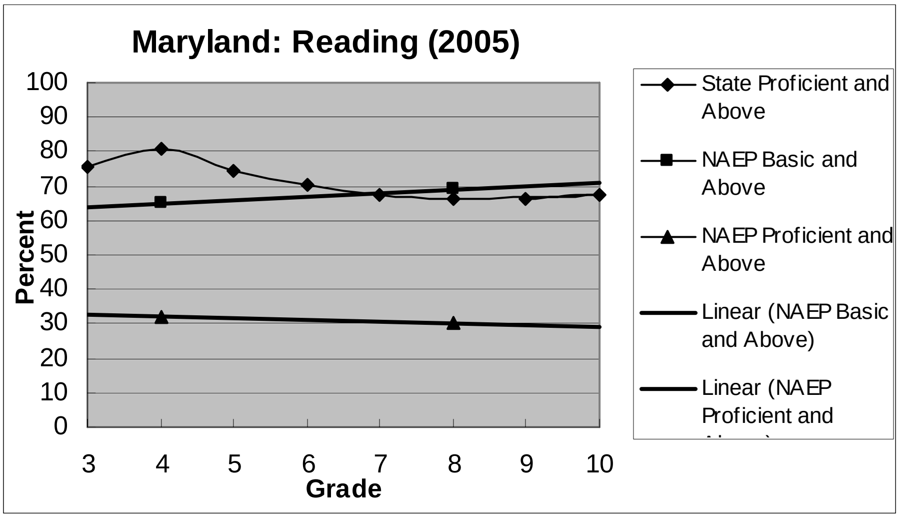
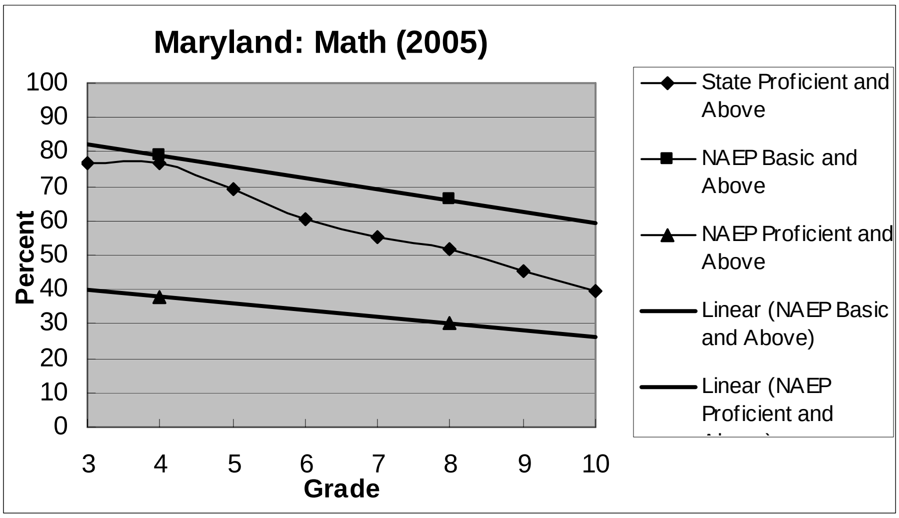


Figure 18. MDm



Data source:
http://msp.msde.state.md.us/downloadindex.aspx?K=99AAAA

The data came from the Maryland School Assessment (MSA), which is a test of reading and math achievement that meets the requirements of the federal No Child Left Behind Act. The test is given each year in early March at grades 3 through 8. For MSA, there are three achievement levels: Advanced, Proficient, and Basic. See the introduction to the methodology for more details.

Figure 19. MEr

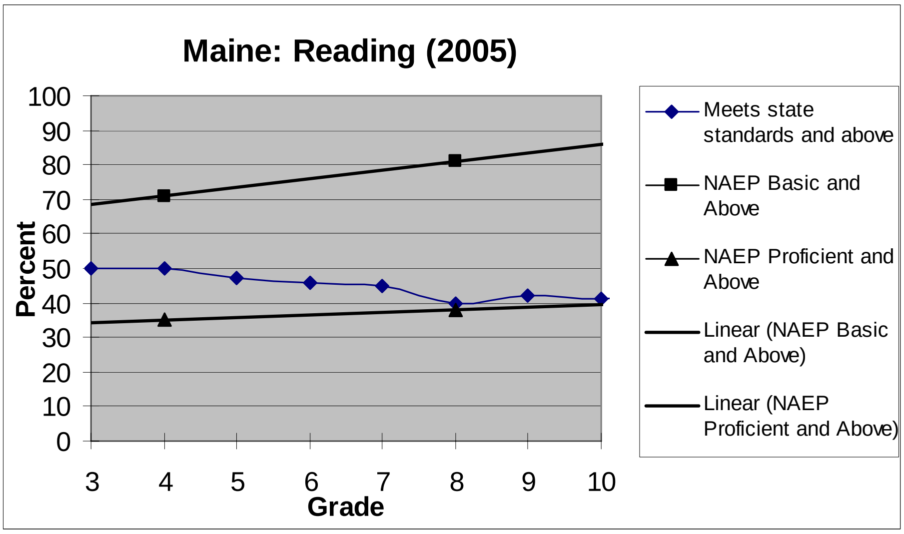
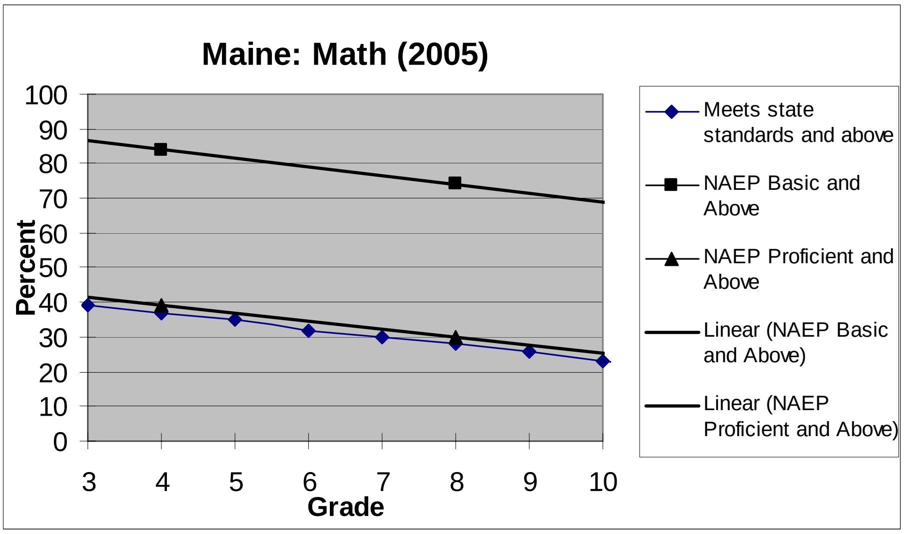


Figure 19. MEm



Data source:
http://www.maine.gov/education/mea/edmea.htm
The Maine Educational Assessment (MEA) tests in reading, writing, mathematics and science/technology are developed to measure the performance of students in achieving the expectations of Maine's Learning Results. The MEA assessment results report student progress in four performance levels: exceeds standards, meets standards, partially meets, and do not meet standards. In the 2005 school year, the Maine Educational Assessment (MEA) was administered to 14,328 grade 4 students, 16,551 grade 8 students, and 15,703 grade 11 students. Additionally, data for grades 3, 5, 6, 7, 9 and 10 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades.

Figure20. MIr

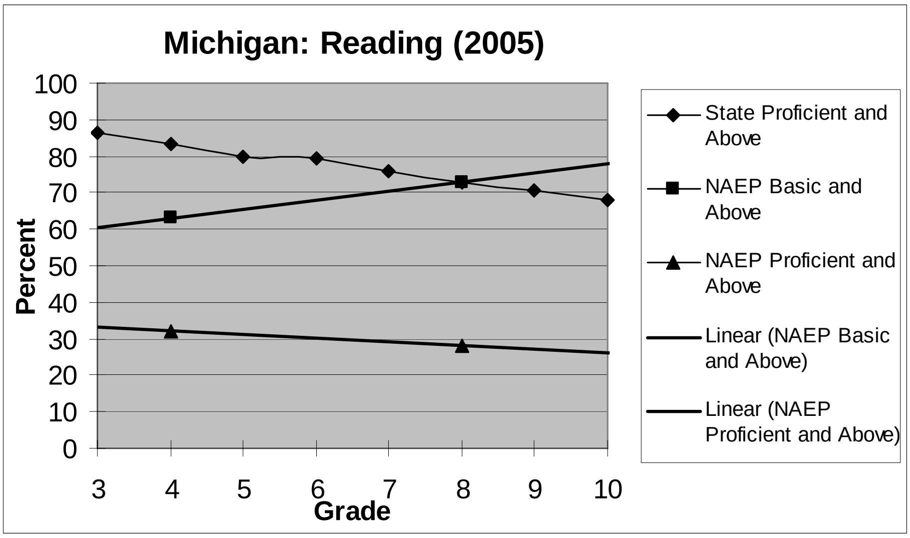
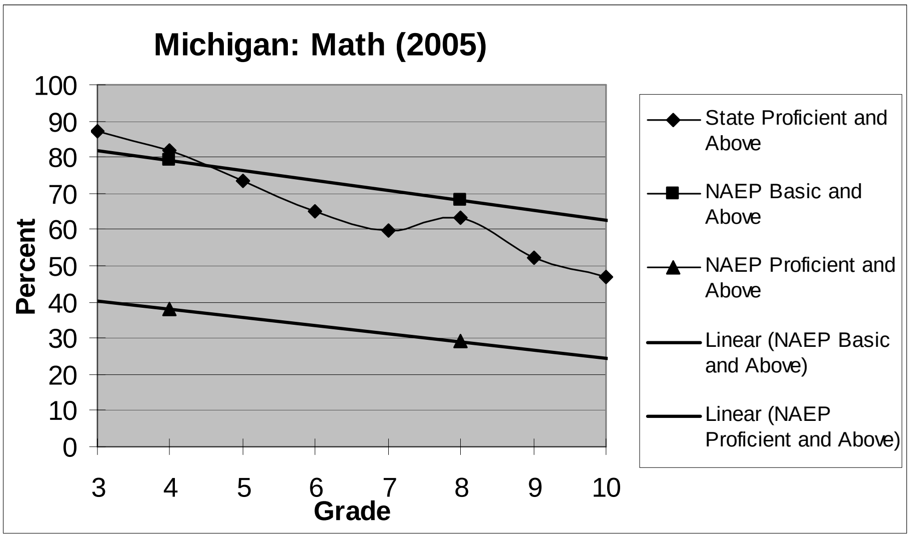


Figure20. MIm



Data source:
http://www.michigan.gov/documents/FALL_2005_STATEWIDE_MEAP_RESULTS_151913_7.pdf
Our data come from Statewide Michigan Educational Assessment Program (MEAP) of the Fall 2005. There are four achievement levels: level 1-Exceeded, level 2- Met, level 3- Basic, and level 4- Apprentice. Students in the level 1 & level 2 are taken as "have passed" the MEAP test. We extrapolated grade 9 and 10 from grade 3 to grade 8 for both reading and math.

Figure 21. MNr

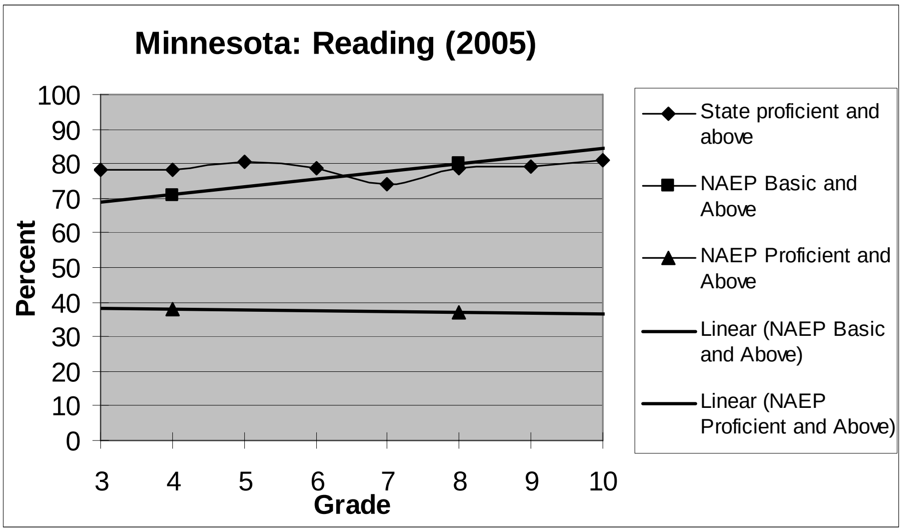
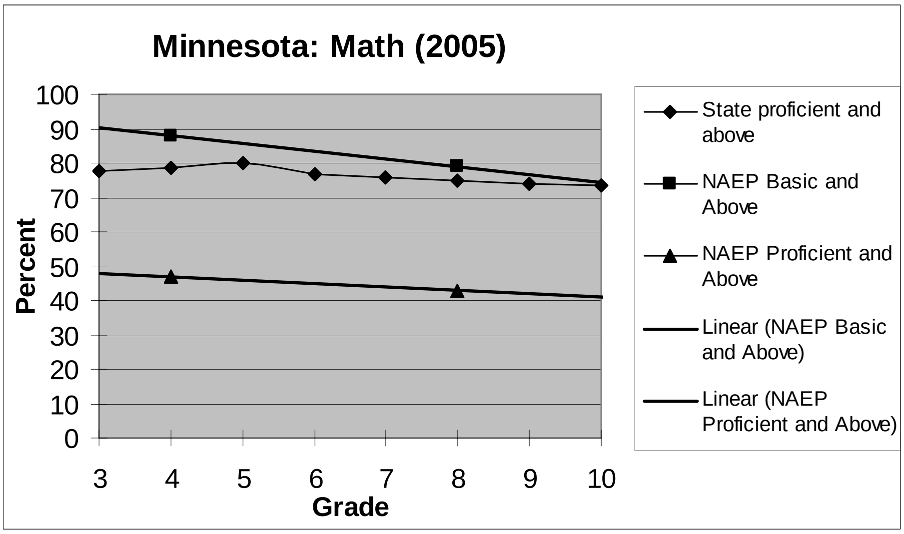


Figure 21. MNm



Data source:
http://education.state.mn.us/mde/static/MCA2005Public_Filter9.TAB
The MCA-IIs (Minnesota Comprehensive Assessments—Series II) are the state reading and mathematics tests that meet the requirements of No Child Left Behind. These assessments are specific to Minnesota. There are 5 levels of achievement in Minnesota, labeled 1 to 5, with 5 being the highest level. For AYP purposes, Level 3 currently corresponds to the Federal benchmark of proficient. Here, we interpolated grades 4, 6, 8 and 9 from grades 3, 5, 7 and 10 for Reading and interpolated grades 4, 6, 8, 9 and 10 from grades 3, 5, 7 and 11 for Math.

Figure 22. MOr

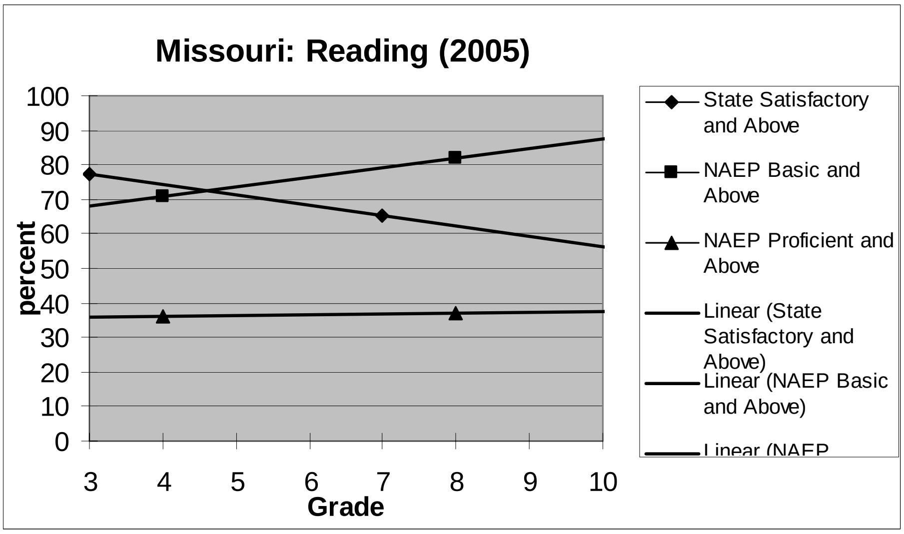
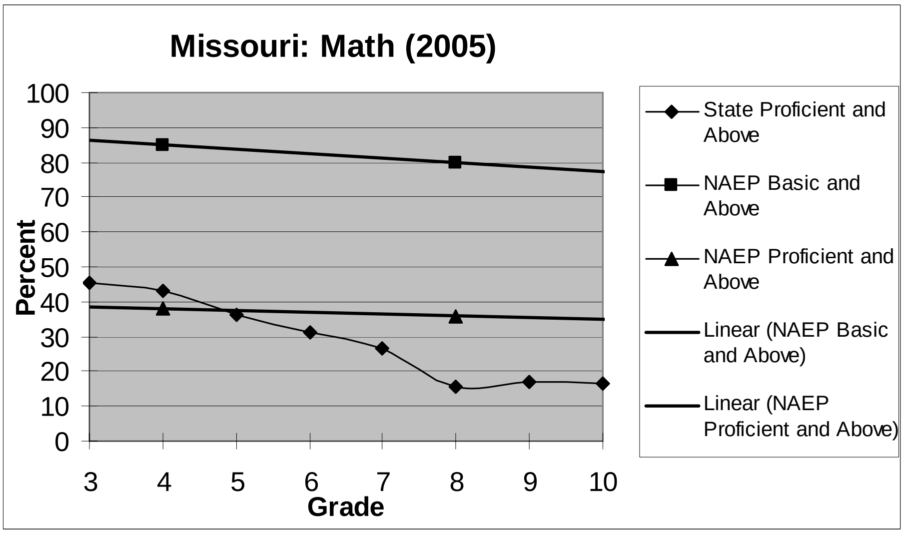


Figure 22. MOm



Data source:
http://dese.mo.gov/divimprove/assess/stateresults.html.pdf;
http://dese.mo.gov/divimprove/assess/State_MAP2005_Reading.pdf
The Missouri Assessment Program (MAP), in Spring 2005, included required assessments in Math for grades 4, 8, 10. In this assessment, students' achievement levels are categorized into "Step 1", "Progressing", "Near Proficiency", "Proficiency" and "advanced". Data for grades 3, 5-7, and 9 were interpolated based on data already known. For Reading, the achievement levels are "proficiency", "satisfaction", and "un-satisfaction." MAP reading 2005 only has data available for Grade 3 and 7 and so the "State Satisfactory and Above" line is a straight trend line based on only two plots.

Figure 23. MSr

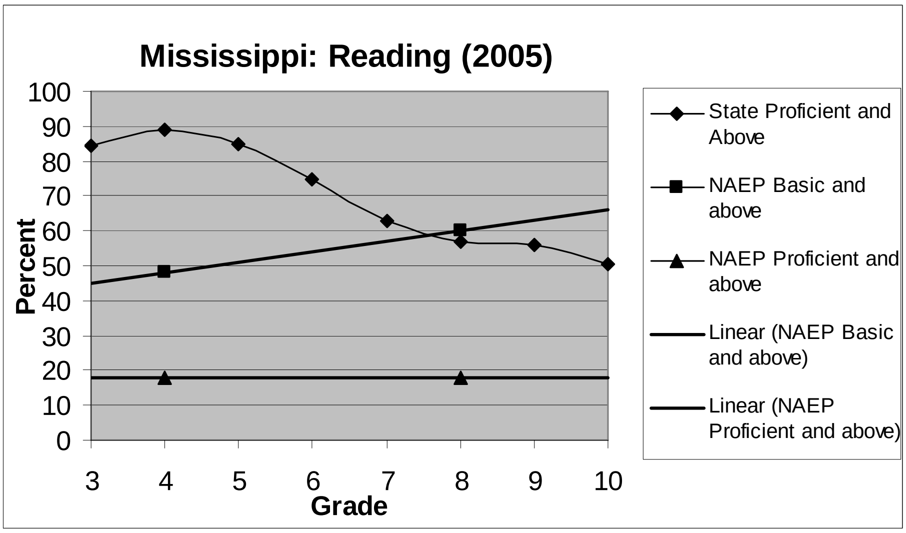
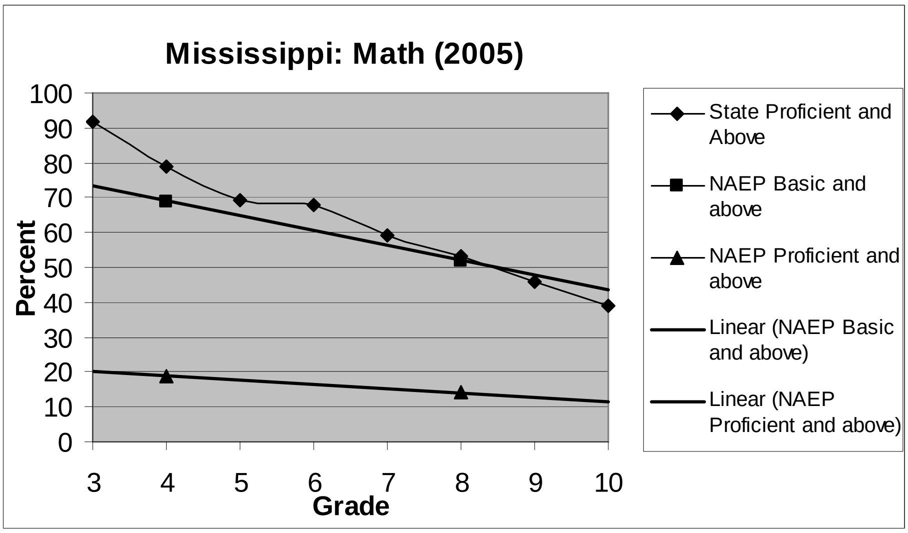


Figure 23. MSm



Data source:
http://orsap.mde.k12.ms.us:8080/MAARS/indexProcessor.jsp
Our data comes from Mississippi Curriculum test (MCT), which includes three subjects (reading, language, and mathematics) across grades 2 through 8. Student performance is classified into four different levels: Minimal, Basic, Proficient and Advanced. Here, for both reading and math graphs, we extrapolate 2 plots in the first lines, grade 9 and grade 10, in terms of grade 3 to 8.

Figure 24. MTr

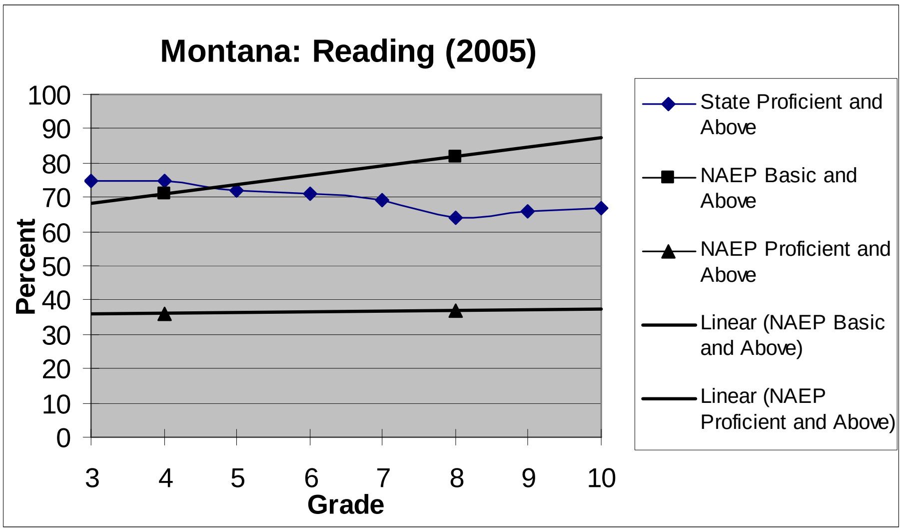
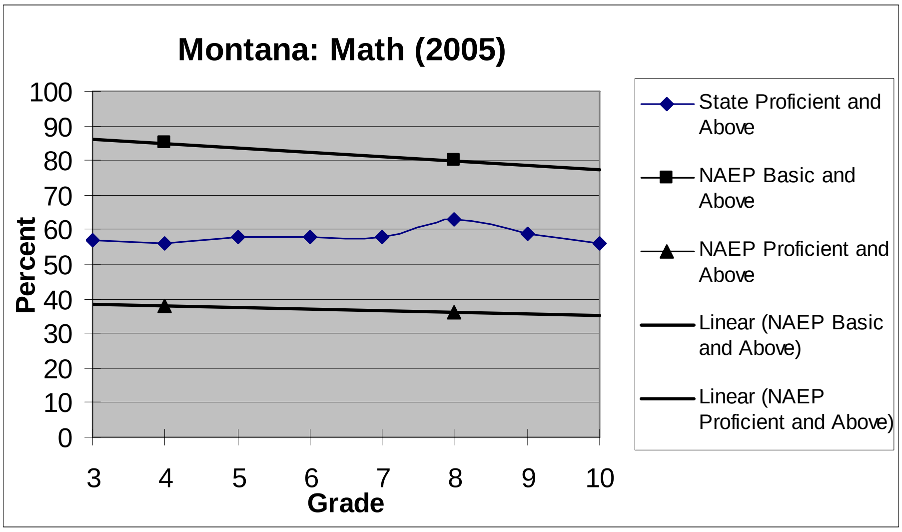


Figure 24. MTm



Data source:
http://www.opi.state.mt.us/
The Criterion-Referenced Test (CRT) assessment results report student progress in four performance levels: novice, nearing proficient, proficient, and advanced. Students performing at proficient and advanced levels met or exceeded the standards. Additionally, all Montana public school students in grades 4, 8 and 10 were assessed on state-approved mathematics and reading standards in 2005. However, data for grades 3, 5, 6, 7, and 9 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades.

Figure 25. NCr.

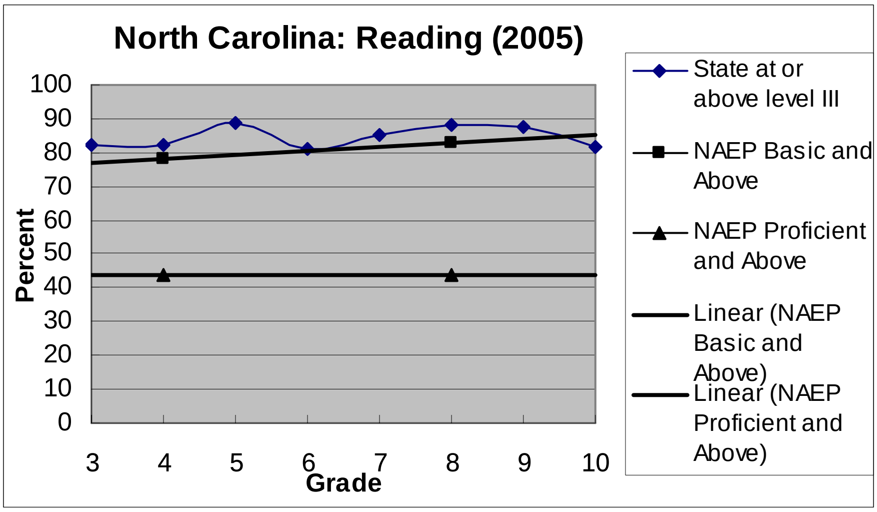
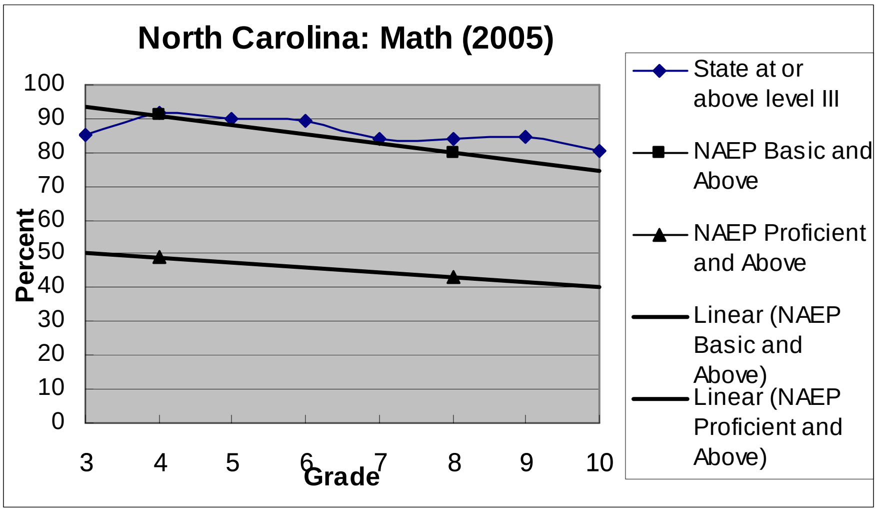


Figure 25. NCm



Data source:
http://www.ncpublicschools.org/accountability/reporting/leaperformancearchive/
This site reports student performance data by student subgroup for two years. The report includes the number of students at or above grade level (Level III), the number of valid scores, and the percent at or above Level III on end-of-grade (EOG), end-of-course (EOC), and alternate assessments. Districts reported student performance in four levels of performance: level I, II, III, and IV. Students performing at level III and IV met the state's standards. In addition, data for grade levels 3 to 8 and 10 were available for 2005. However, data for grade 9 were unavailable, so for all these series, linear interpolation was used to generate the results for grade 9. Here we need also point out that NC use Algebra I and English I test result from EOC as AYP indicator for grade 10 instead of reading and math data. For this reason, we use data from Algebra I and English I for grade 10 while those from reading and math are used for grade 3-8.

Figure 26. NDr

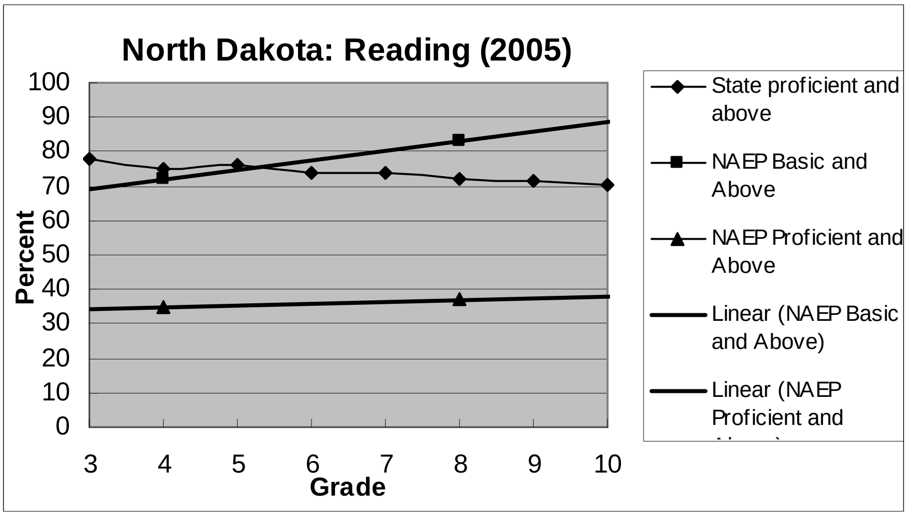
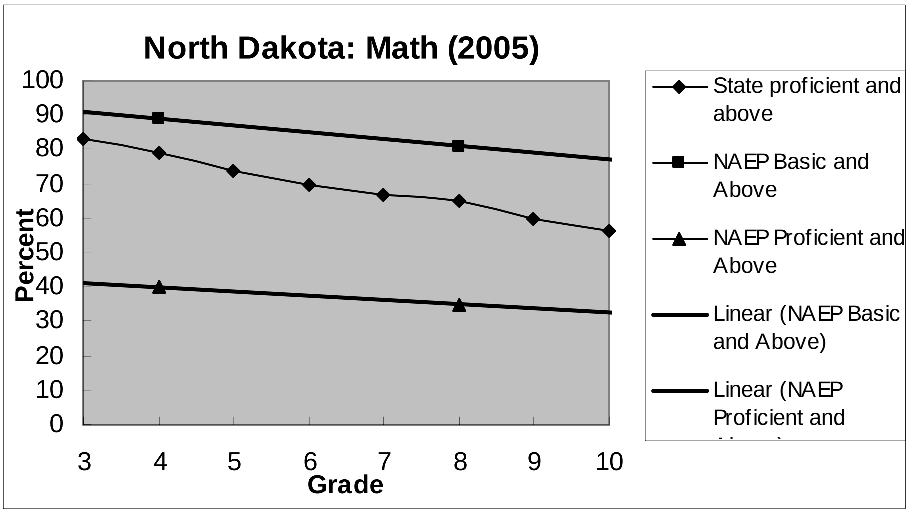


Figure 26. NDm



Data source:
http://www.dpi.state.nd.us/resource/biennial.PDF
Data come from the 2003-2005 Biennial Report which gives North Dakota State Testing results. There are four Proficiency Categories for the state test: Novice, Partly proficient, Proficient, and Advanced. Here, we extrapolated grades 9 and 10 from grades 3 to 8.for both graphs.

Figure 27. NEr

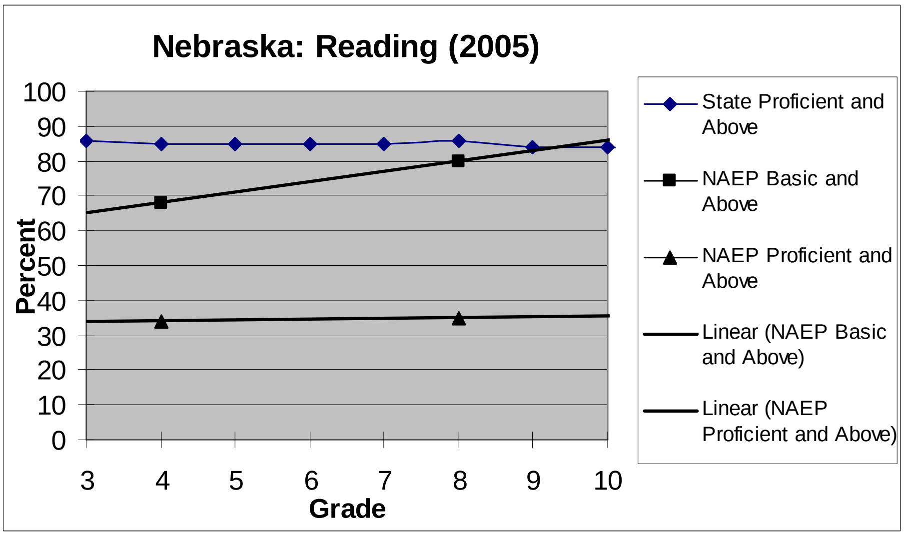
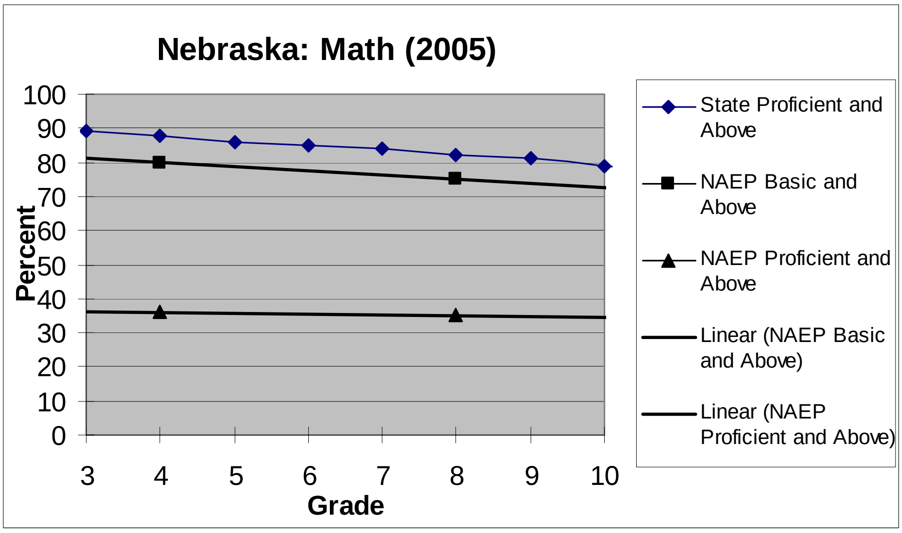


Figure 27. NEm



Data source:
http://reportcard.nde.state.ne.us/Page/PerfStandardOverall.asp?PerfStandardsCategory=1&Level=st&Subject=2

All Nebraska public school students in grades 4, 8 and 11 were assessed on state-approved mathematics and reading standards. Districts reported student performance in four levels of performance: beginning, progressing, proficient, and advanced. Students performing at the advanced and proficient levels met or exceeded the standards. Data for grades 3,5,6,7,9,10 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades.

Figure 28. NHr

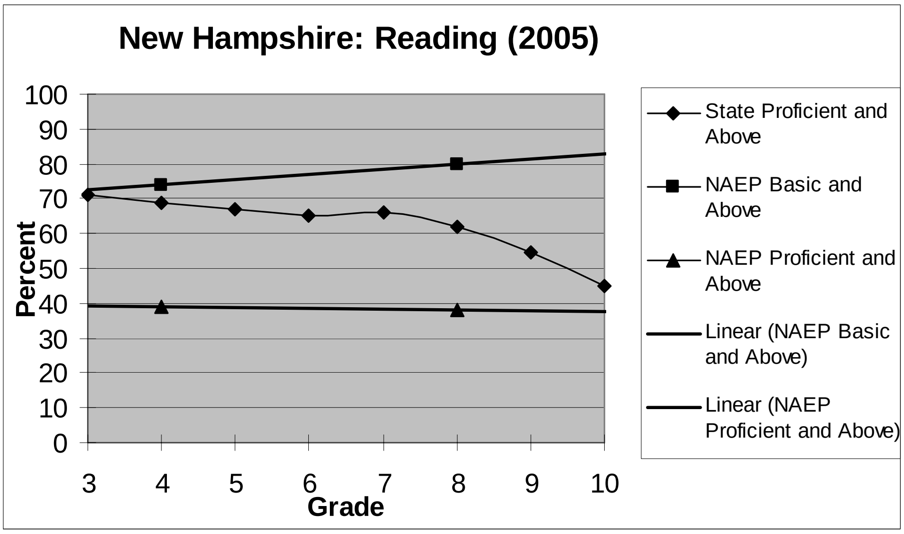
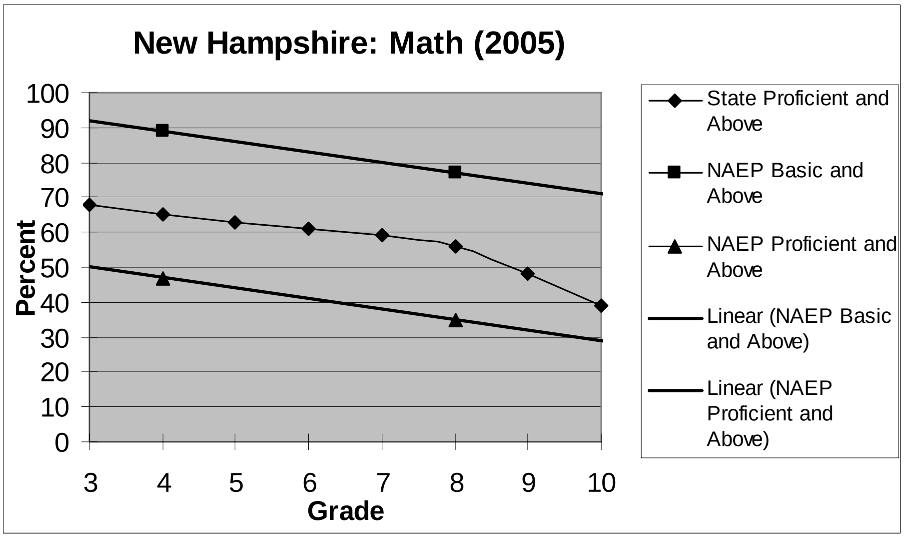


Figure 28. NHm



Data source:
http://www.ed.state.nh.us/education/doe/organization/curriculum/NECAP/NECAP_results.htm (For grade 3-8)
http://www.ed.state.nh.us/education/doe/organization/curriculum/Assessment/2005/Sta.pdf (For grade 10)

The New England Common Assessment Program (NECAP) is a program resulting from the collaboration among New Hampshire, Rhode Island and Vermont. In New Hampshire, NECAP tests were administered in grades 3 though 8. Students' achievement has four levels according to NECAP, "Proficient with Distinction", "Proficient", "Partially Proficient" and "Substantially below Proficient". For grade 10, data come from the grade 10 assessment administered in May 2005 by New Hampshire Educational Improvement and Assessment Program (NHEIAP). In this test, students' achievement levels consist of "Advanced", "Proficient", "Basic" and "Novice". Given data available in grades 3 through 8 plus 10, data for Grade 9 were obtained by interpolation.

Figure 29. NJr

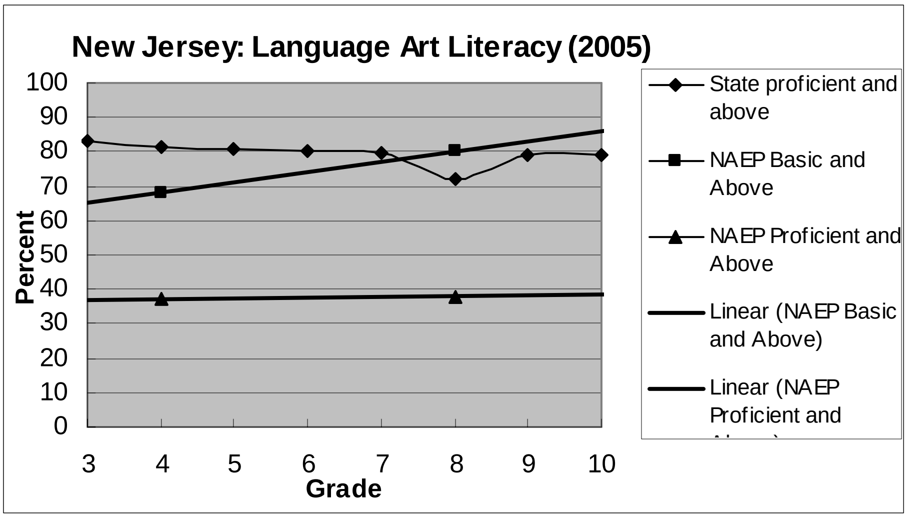
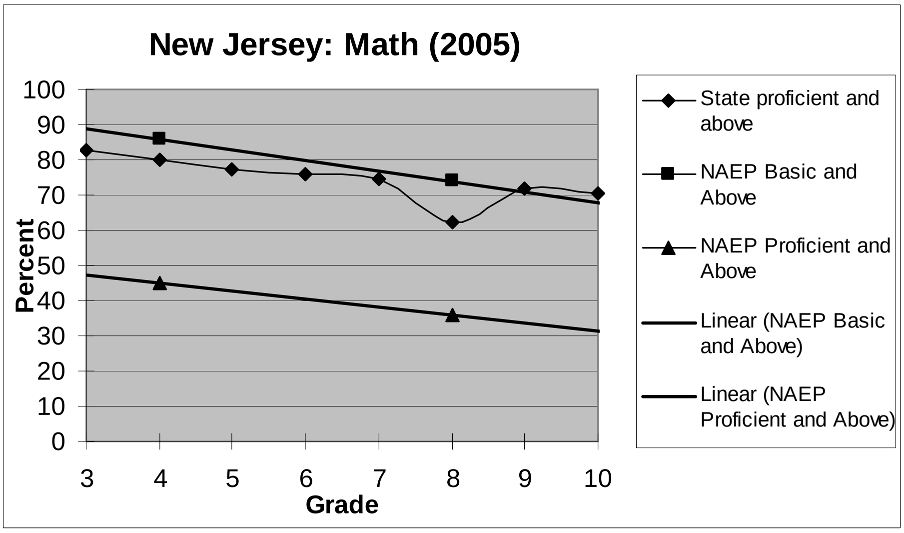


Figure 29. NJm



Data source:
http://www.nj.gov/njded/schools/achievement/2006/njask3/summary.pdf (For grade 3, 4)
http://www.nj.gov/njded/schools/achievement/2006/gepa/summary.pdf (For grade 8)
http://www.nj.gov/njded/schools/achievement/2006/hspa/summary.pdf (For grade 11)
Data for grades 3 and 4 comes from New Jersey Assessment of Skills and Knowledge (NJ ASK), a comprehensive, multi-grade assessment program. The results of this elementary-level assessment are intended to be used to identify students who need additional instructional support in order to reach the Core Curriculum Content Standards. Data for grade 8 came from the 2005 New Jersey Grade Eight Proficiency Assessment (GEPA), which is used as a primary indicator for identifying those students who may need instructional intervention. Data for grade 11 come from HSPA. All these three assessments have the same three achievement levels: Partially

Proficient, Proficient, and Advanced Proficient. We interpolated grades 5, 6, 7, 9 and 10 from grades 3, 4, 8 and 11 for both graphs.

Figure 30. NMr

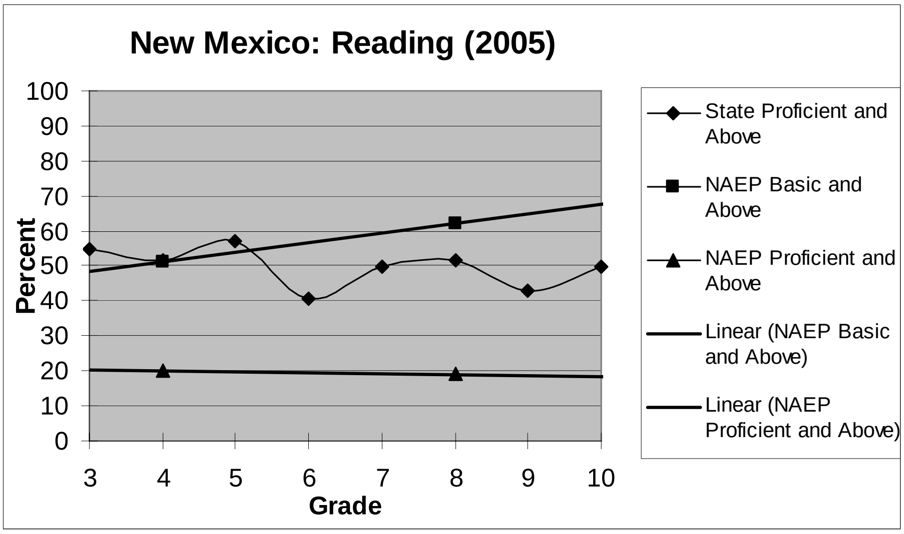
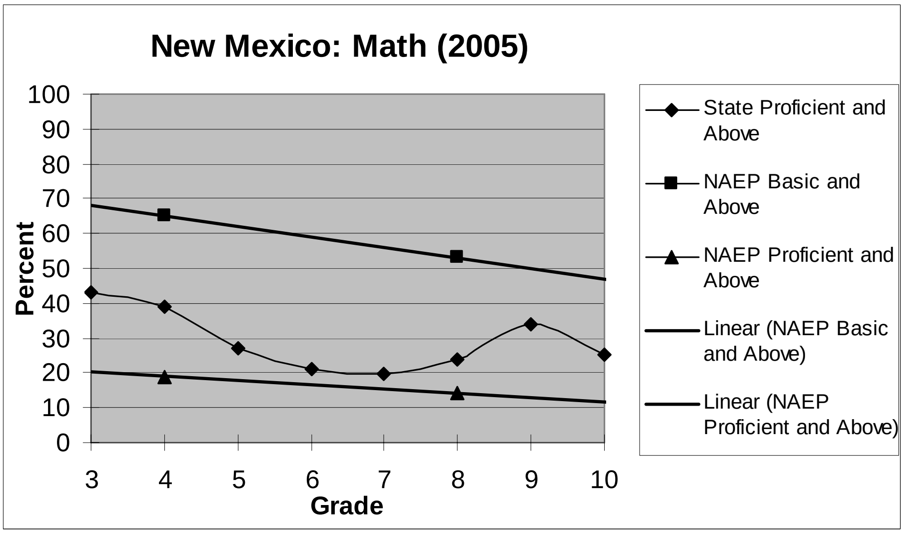


Figure 30. NMm



Data source:
http://www.ped.state.nm.us/press/2005/august/assesment_samples/Test%20Results%20081805.pdf

New Mexico Standard Based Assessments (NMSBAs) were for the first time administered in March 2005 in grades 3, 5, 6, 7, and 9. Tests in Grades 4, 8, 11 were substantially redesigned to improve alignment to the New Mexico Content Standards. Student Achievement levels are: “beginning step”, “Nearing Proficiency”, “Proficient” and “Advanced”. Data for grade 10 were obtained through interpolation based on Grade 3-9 and Grade 11.

Figure 31. OHr

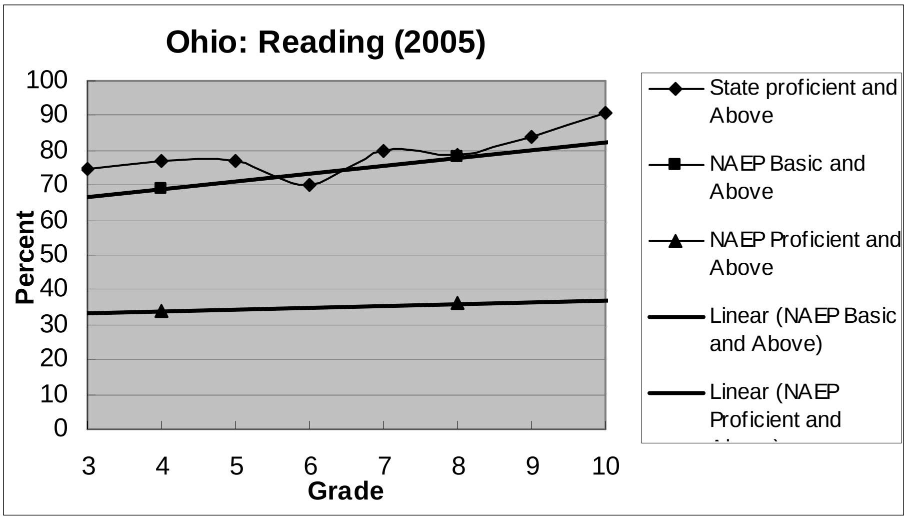
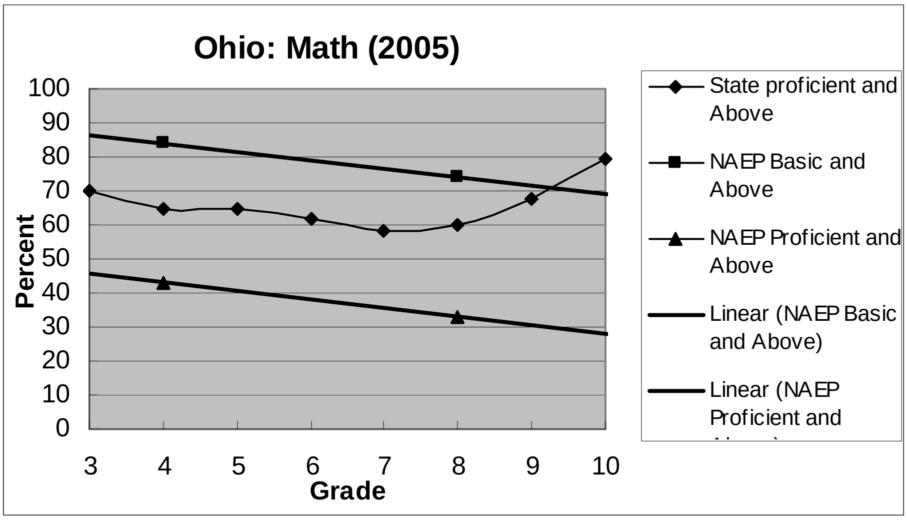


Figure 31. OHm



Data source:
http://www.ode.state.oh.us/proficiency/results.asp
Data for grades 3-8 in math and reading come from State Achievement Tests while data for Grade 10 are from the Ohio Graduation Tests (OGT). The OGT was first administered in March 2004. For each test, there are five achievement levels: Advanced, Accelerated, Proficient, Basic, and Limited. For the OGT, Students at the proficient level or above are considered to "meet the graduation standard". Here, we interpolated grades 7 and 9 from Grade 3-6, 8 and 10 for Reading and extrapolated grades 5 and 9 from grades 3-4, 6-8 and 10 for Math.

Figure 32. OKr

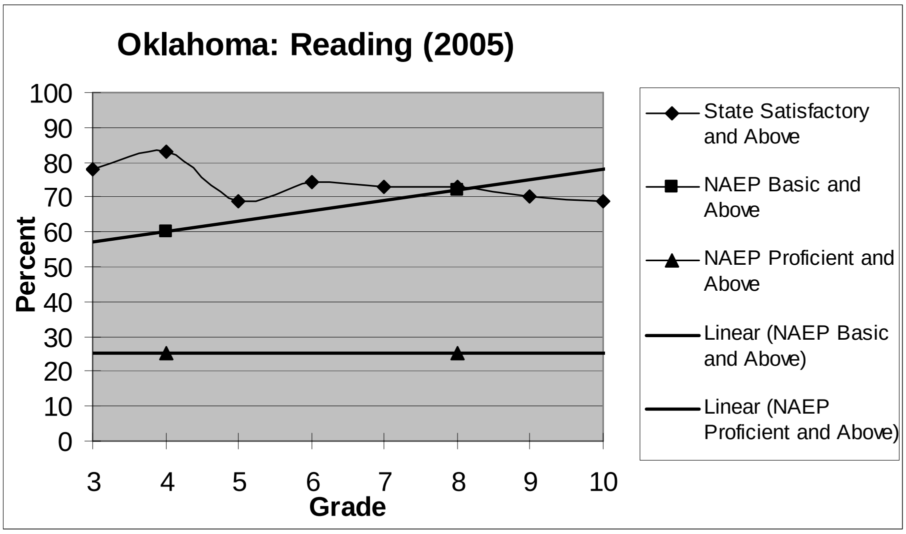
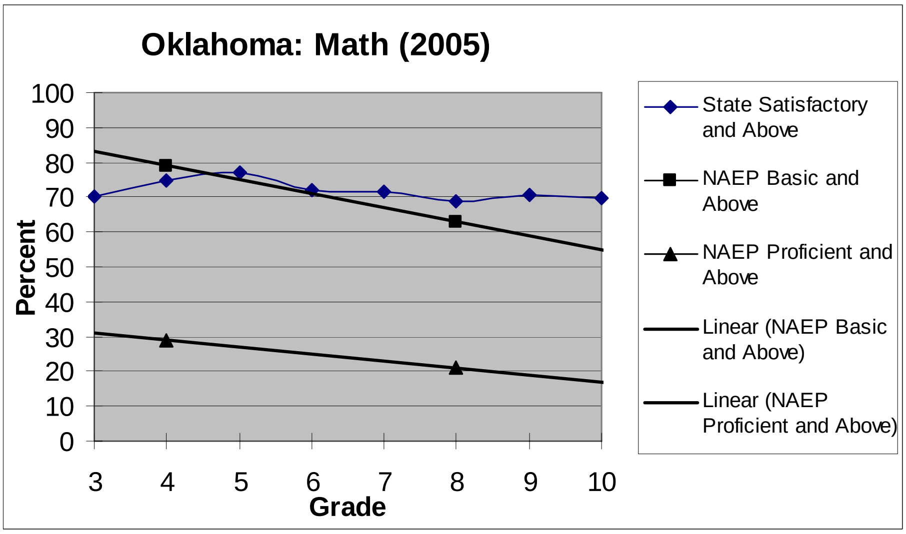


Figure 32. OKm



Data source:
http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf
Our data come from the state Annual Report Card of 2004-2005. The performance results are classified in four levels**:** Advanced Performance Level; Satisfactory Performance Level; Limited Knowledge Performance Level and Unsatisfactory Performance Level. The Satisfactory level is considered proficient. Therefore, a student scoring at Satisfactory Level or Advanced Level meets the proficiency standard required of NCLB. Here, we interpolated grades 6, 7, 9 and 10 from grades 3, 4, 5 and 8 for both reading and math.

Figure 33. ORr

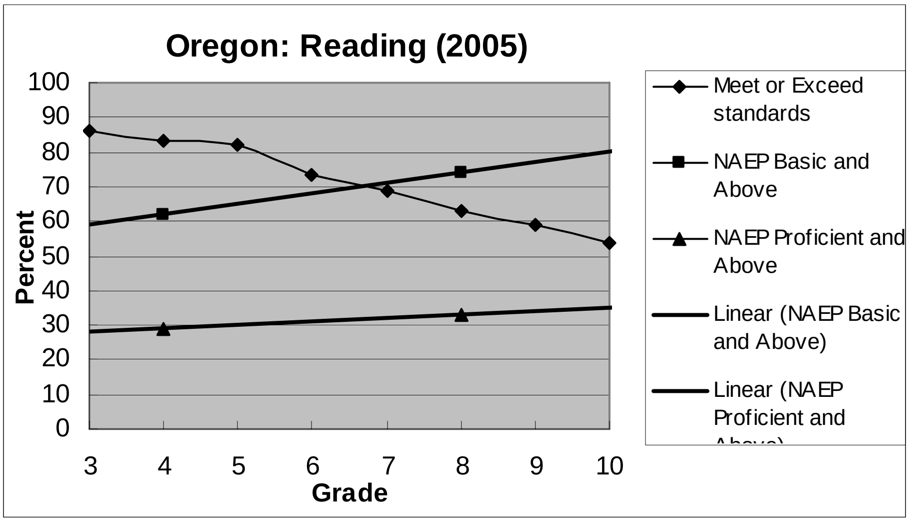
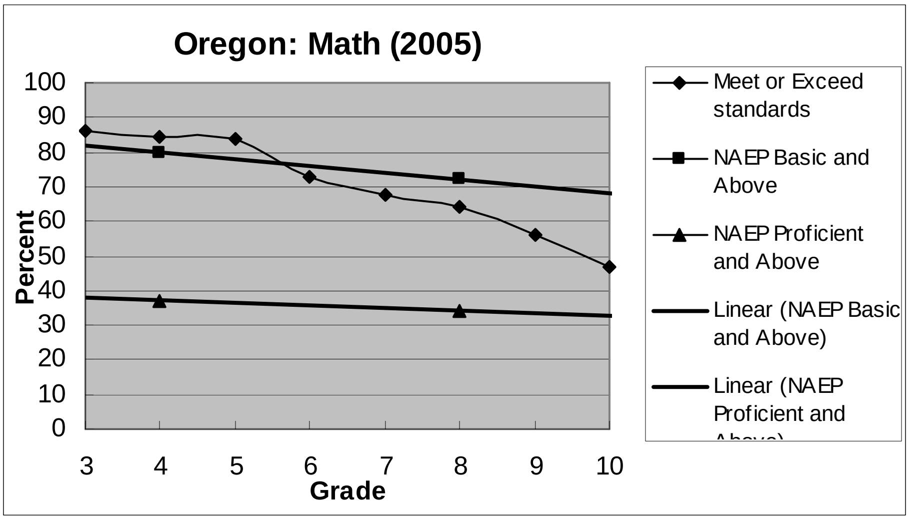


Figure 33. ORm



Data source:
http://www.ode.state.or.us/data/annreportcard/rptcard2005.pdf
Data come from 2004-2005 Oregon Statewide Report Card. Statewide assessments are administered at grades 3, 5, 8 and 10 in reading, writing, mathematics and science. Oregon students are categorized into three levels: Exceeds standards, Meets standards, and Not meets standards. We interpolated grades 4, 6, 7 and 9 from grades 3, 5, 8 and 10 for both reading and math.

Figure 34. PAr

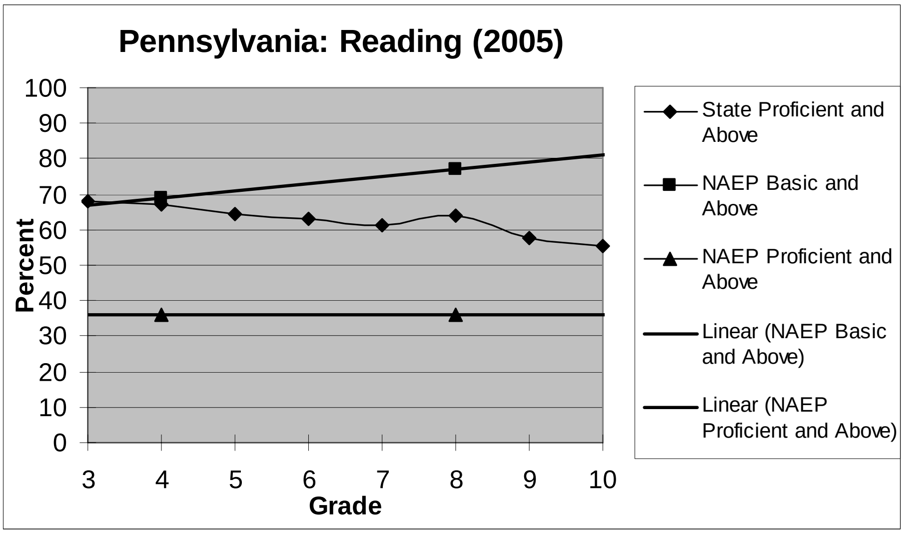
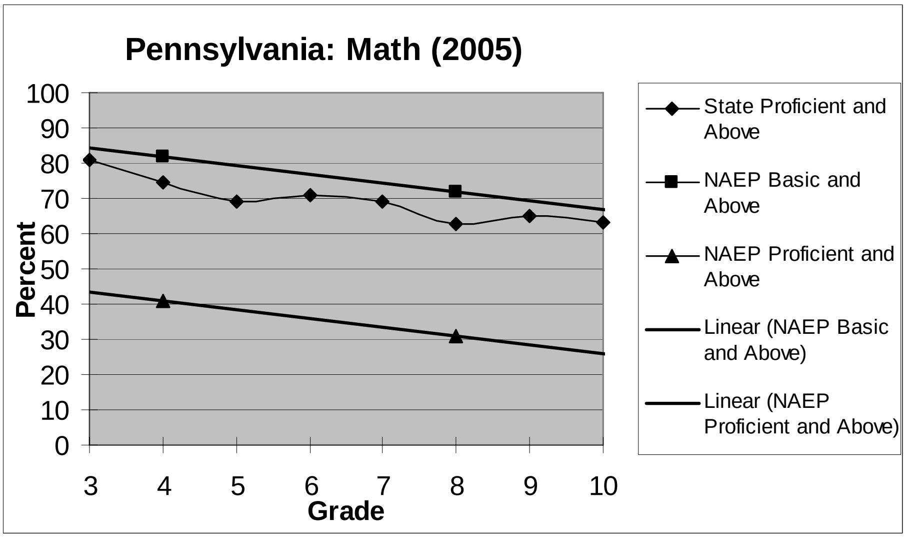


Figure 34. PAm



Data source:
http://www.pde.state.pa.us/a_and_t/cwp/view.asp?a=3&q=115272
Data comes from the Pennsylvania System of School Assessment (PSSA),,a standardized test given to students in 3rd, 5th, 8th and 11th grades throughout the state. Students are classified into four achievement levels: advanced, proficient, basic or below basic. We interpolated and extrapolated grades 4, 6, 7, 9 and 10 from the other previous grades we had for both graphs.

Figure 35. SCr

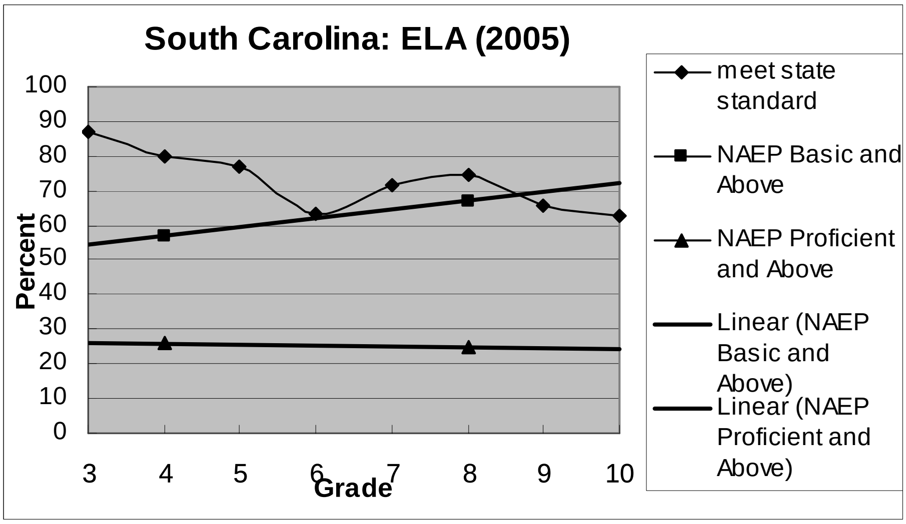
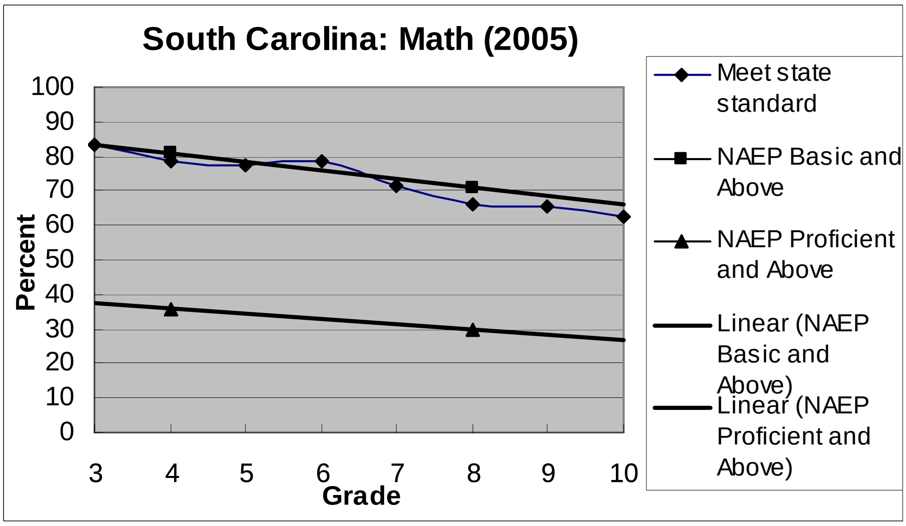


Figure 35. SCm



Data source:
http://www.myscschools.com/tracks/testscores/pact/2005/
In South Carolina, the Palmetto Achievement Challenge Test (PACT) is administered in grades 3-8 and includes English Language Arts, Mathematics, Science and Social Studies. Students are categorized into four achievement levels: Below Basic, Basic, Proficient, or Advanced. Students who are at basic level or above are considered “meet state standard”. Data for grades 9 and 10 were extrapolated from grades 3 to Grade 8 for both graphs.

Figure 36. SDr

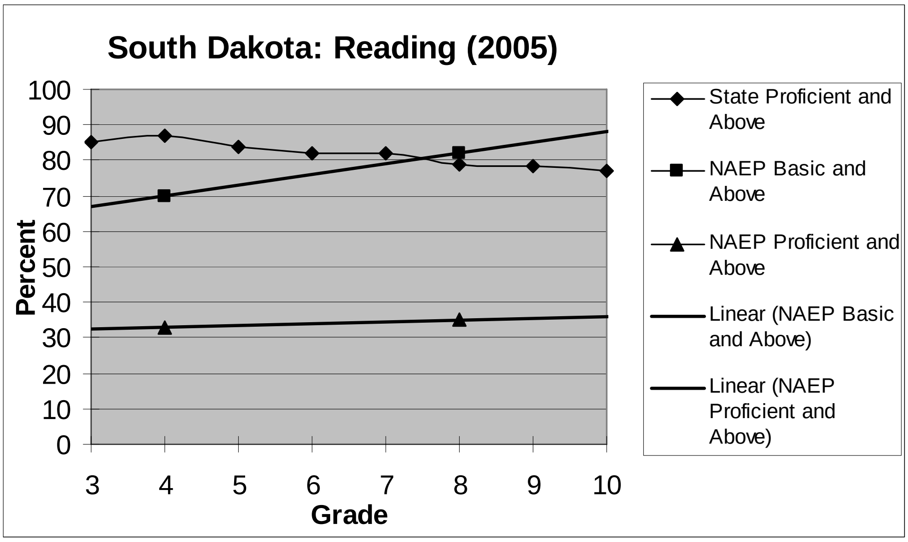
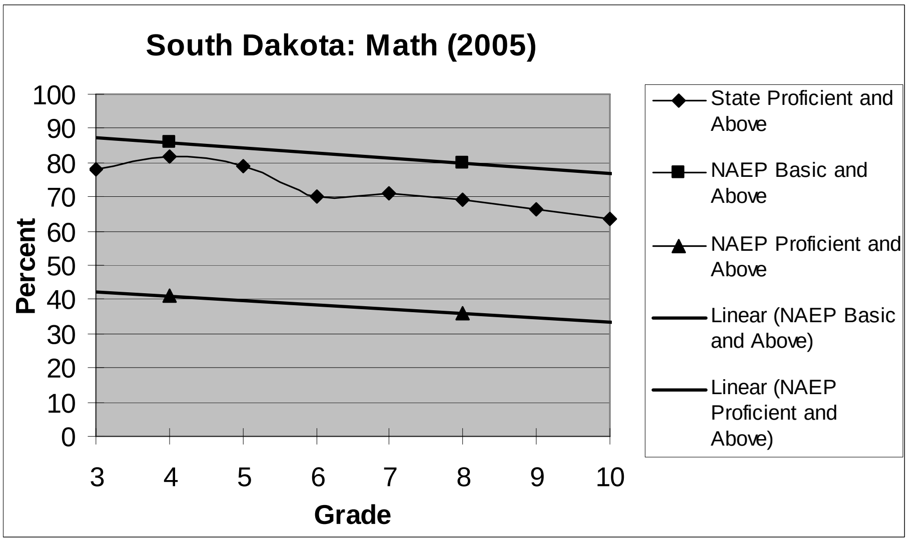


Figure 36. SDm



Data source:
https://sis.ddncampus.net:8081/nclb/portal/portal.xsl?&extractID=7
Our data come from 2005 Report Card for No Child Left Behind. Students in grades 3 through 8 and grade 11 completed the Dakota STEP assessment in the spring of 2005. Students are categorized into four achievement levels: Advanced, Proficient, Basic, and Below Basic. We extrapolated grades 9 and 10 from grades 3 to 8 for both reading and math.

Figure 37. TNr

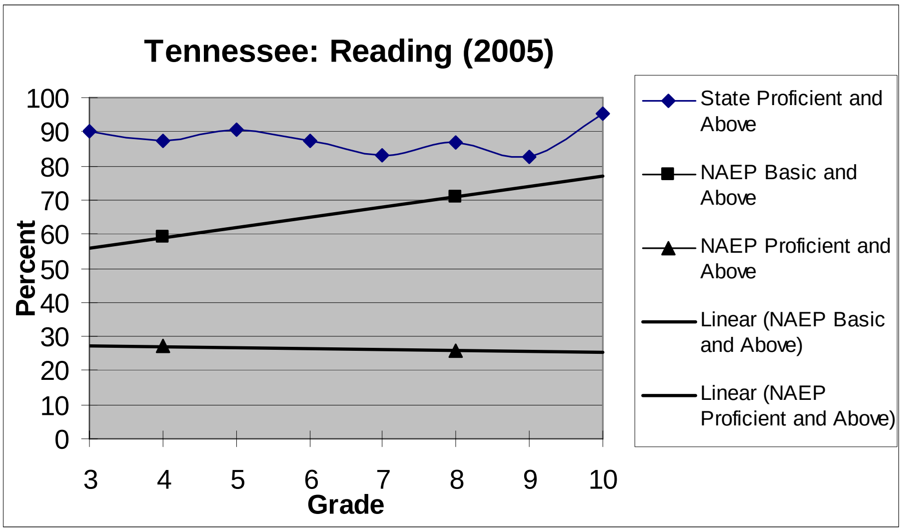
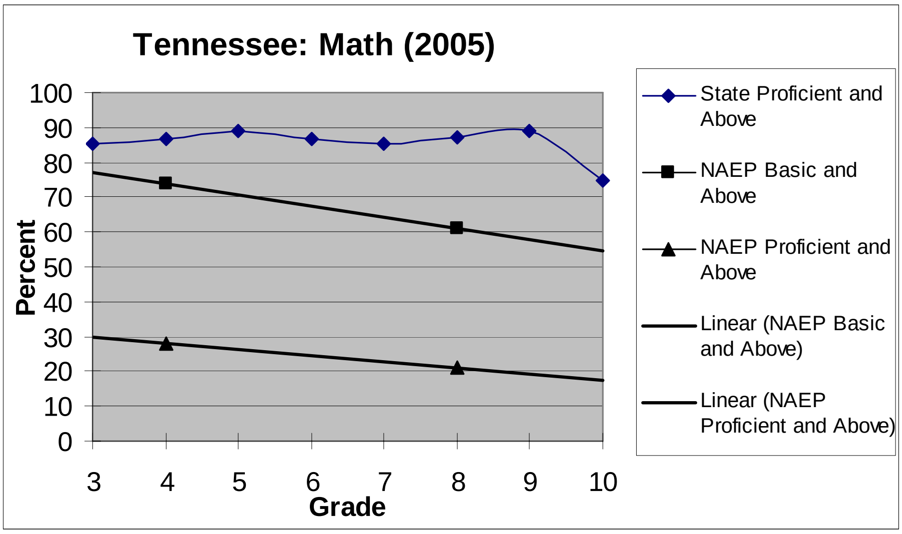


Figure 37. TNm



Data source:
http://www.k-12.state.tn.us/rptcrd05/state2.asp
The data are from "State of Tennessee report card 2005." Districts reported student performance in three levels of performance: below proficient, proficient and advanced. Students performing at the advanced and proficient levels met or exceeded the state's standards. In addition, data for all grade levels, 3 through 10, were available for 2005, so linear interpolation and extrapolation were not used to generate results.

Figure 38. TXr

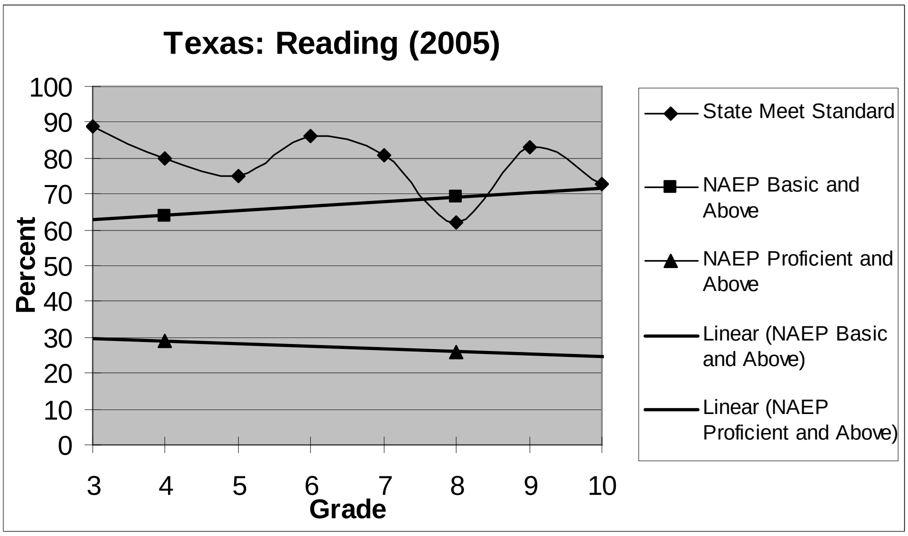
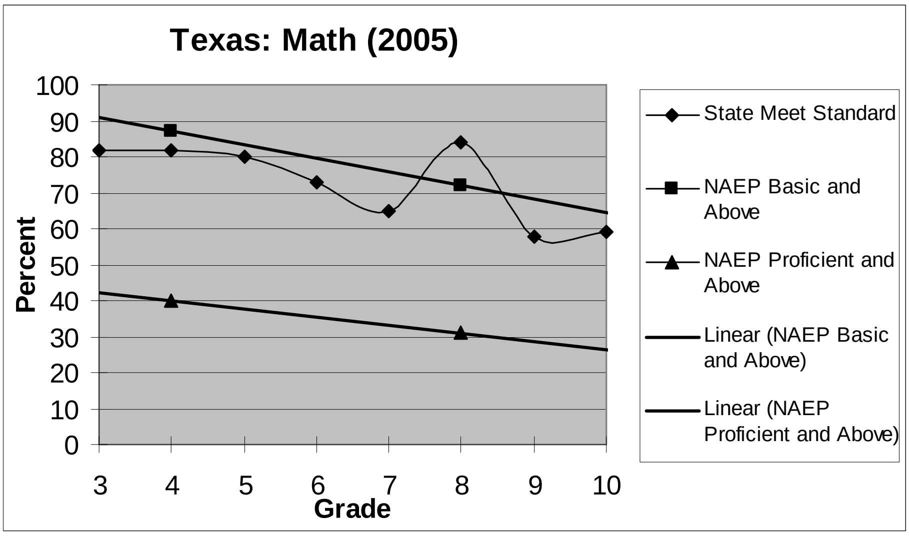


Figure 38. TXm



Data source:
http://www.tea.state.tx.us/perfreport/aeis/2005/state.html
Our data comes from Texas Assessment of Knowledge and skills (TAKS). The TAKS measures the statewide curriculum in reading at grades 3-9; in writing at grades 4 and 7; in English Language Arts at grades 10 and 11; in mathematics at grades 3-11; in science at grades 5,10, and 11; and social studies at grades 8, 10, and 11. Based on students' performance, they are classified into three performance levels: Did Not Meet Standard, Met Standard, and Commended Performance. Students in the Met Standard or the Commended Performance categories have passed the TAKS test and also reached "proficient" level as NCLB required. Here, we extrapolated grade 10 from grades 3 through 9 for the reading graph.

Figure 39. VAr

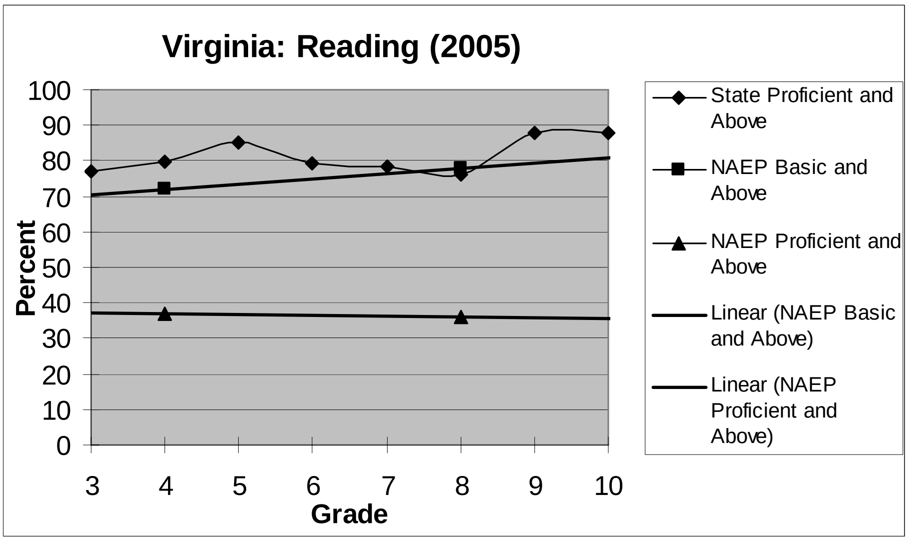
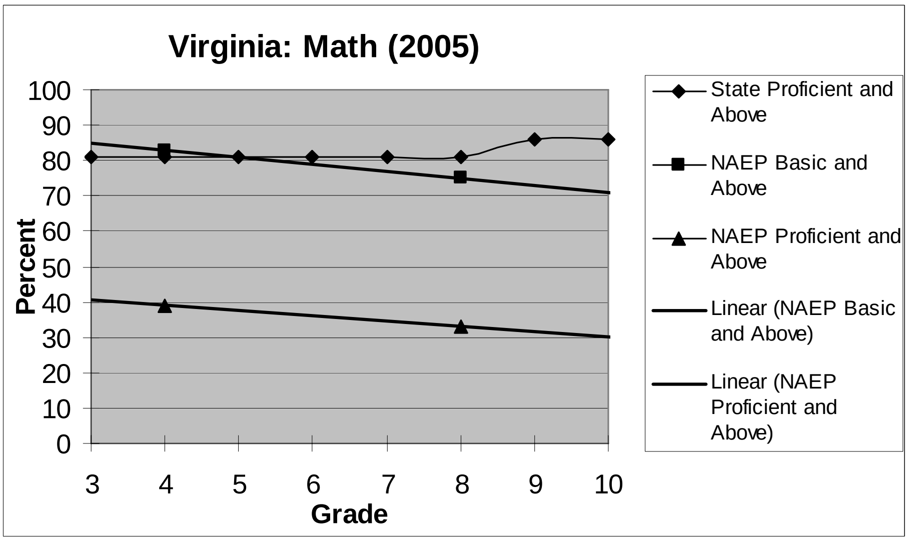


Figure 39. VAm



Data source:
http://pen2.vak12ed.edu/cgi-bin/broker?_service=doe_prod&_program=prodcode.doerp101rcdp001.sas#gr11

Our data came from the Virginia School Report Card at the state level. The state report includes data on the achievement of students on all state standardized tests, including SOL (Standards of Learning Assessment) tests, substitute assessments approved by the Board of Education, and tests taken by students with disabilities and students of limited English proficiency. There are three achievement levels: advanced, proficient, and fail. Students in advanced or proficient level are considered as "have passed" the tests. We interpolated grades 4, 6 and 7 from grades 3, 5 and 8 for both reading and math. Grade 9 and 10 are from the combined results of grade 9 to grade 12.

Figure 40. WAr

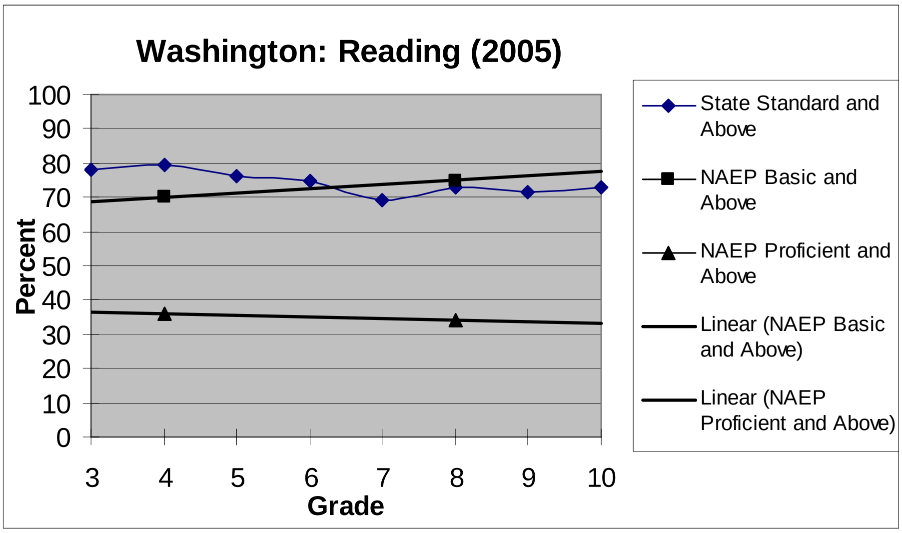
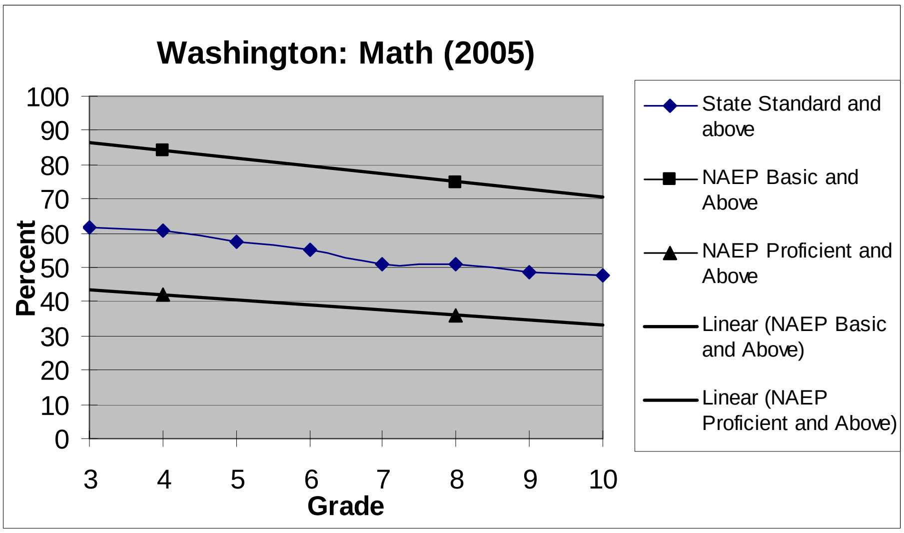


Figure 40. WAm



Data source:
http://reportcard.ospi.k12.wa.us/
The Washington Assessment of Student Learning (WASL) is based on the state's Essential Academic Learning Requirements (EALRs). The WASL reports student progress in four performance levels: 1, 2, 3, and 4. Students performing at level 3 and 4 met or exceeded the standards and are considered at "proficient" level as NCLB required. Additionally, all public school students in grades 4, 7 and 10 were assessed on state-approved mathematics and reading standards in 2005. However, data for grades 3, 5, 6, 8, and 9 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades.

Figure 41. WIr

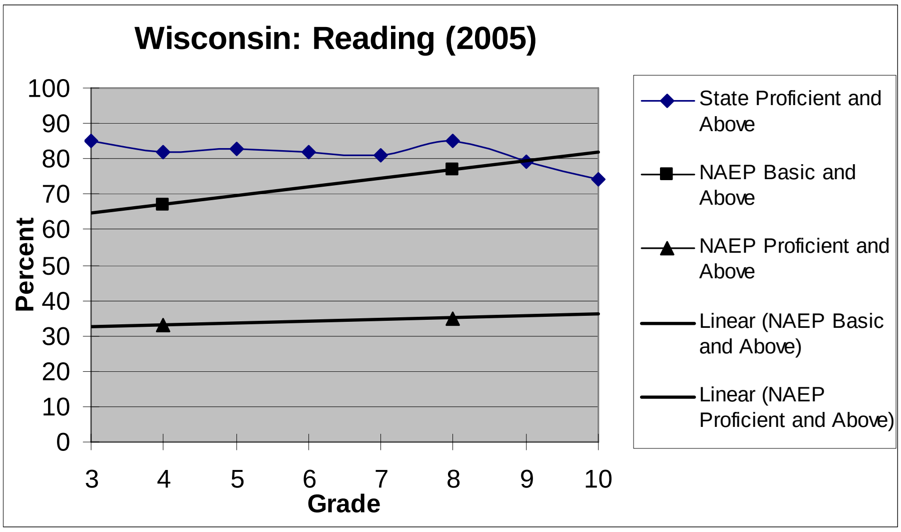
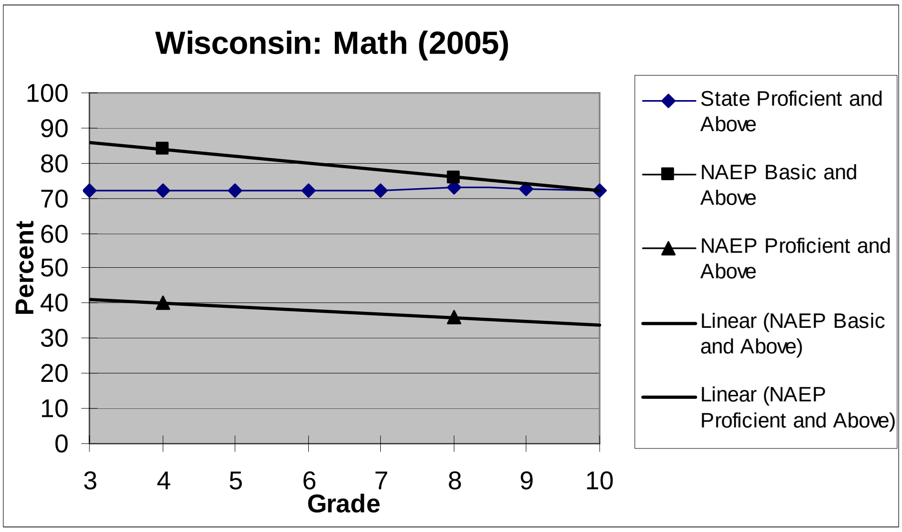


Figure 41. WIm



Data source:
http://data.dpi.state.wi.us/data/graphshell.asp?SubjectID=1RE&GraphFile=GEDISA&DETAIL=YES&Grade=ALL&Group=AllStudentsFAY&EligibleOnly=NO&Level=ALL&WOW=WSAS&ORGLEVEL=ST&FULLKEY=ZZZZZZZZZZZZ&DN=None+Chosen&SN=None+Chosen
Results are from the statewide achievement tests known as the Wisconsin Knowledge and Concepts Examinations (WKCE). Wisconsin uses four proficiency categories: advanced, proficient, basic, and minimal performance. Students performing at advanced and proficient levels met or exceeded the state's standards. In 2005, all Wisconsin public school students in grades 4, 8 and 10 were assessed on state-approved mathematics and reading standards. However, data for grades 3, 5, 6, 7, and 9 were unavailable, so for all these series, linear interpolation and extrapolation were used to generate the results for those grades.

Figure 42. WVr

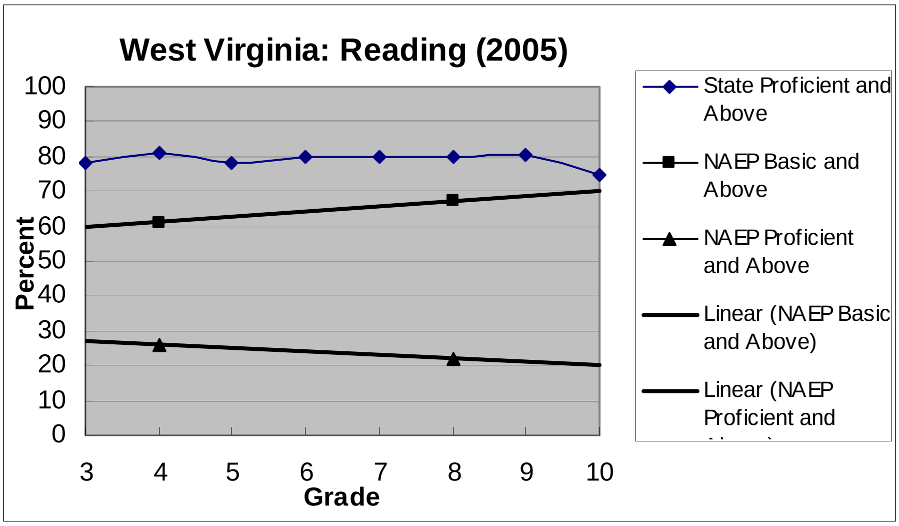
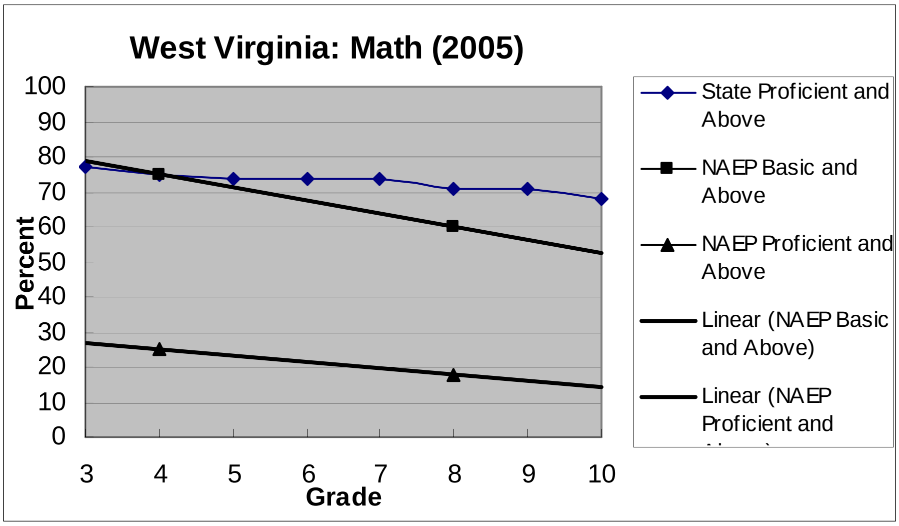


Figure 42. WVm



Data source:
https://wveis.k12.wv.us/nclb/pub/rpt0405/chartdata05rcs.cfm?year=05&county=999&school=999&coname=&rptnum=1&rpage=pickreportcard.cfm

Districts reported student performance in five levels of performance: novice, below mastery, mastery, above mastery, and distinguished. Students performing at or above the mastery level met or exceeded the state's standards and are regarded as proficient. Data for grade levels 3 to 8 and 10, were available for 2005. Linear interpolation was used to generate the results for grade 9.

Figure 43. WYr

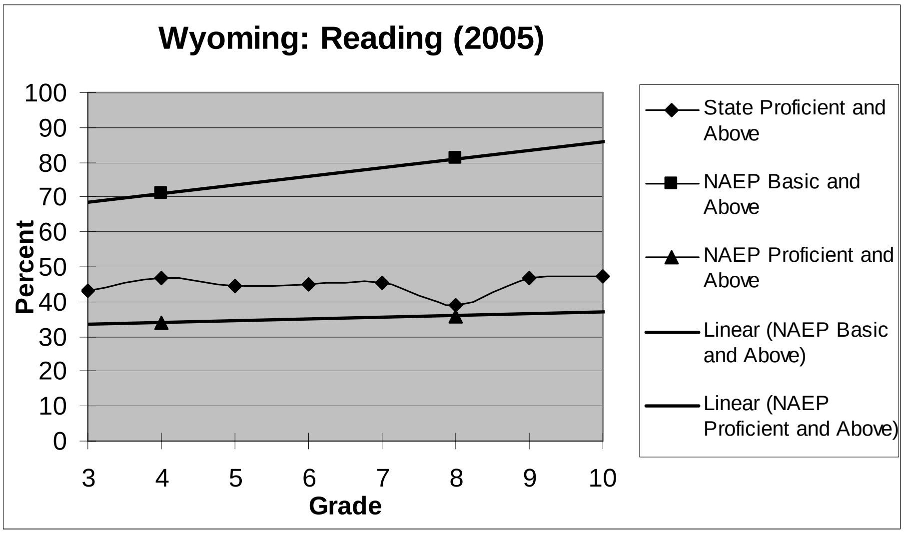
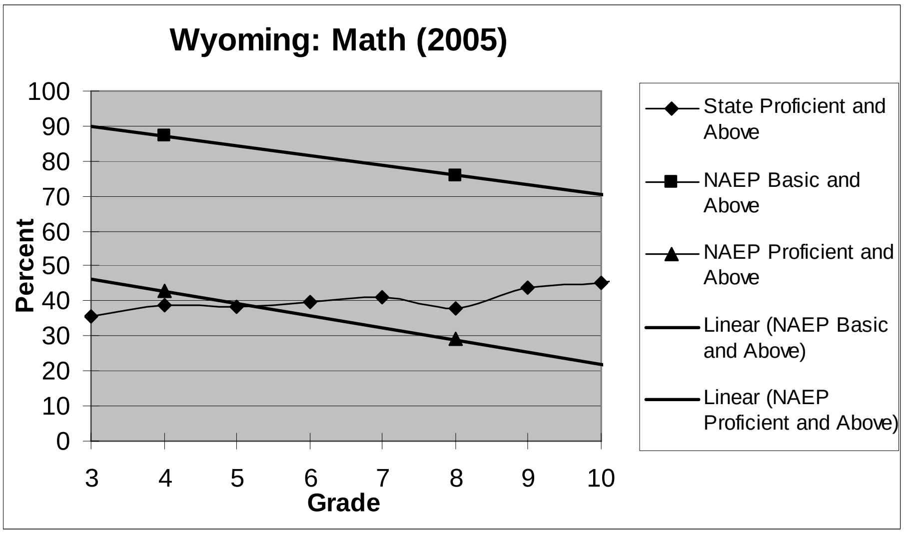


Figure 43. WYm



Data source:
http://www.k12.wy.us/SA/WyCAS/wycas05/state05.asp
Wyoming Comprehensive Assessment System (WyCAS) is the assessment system developed by the Wyoming Department of Education. The 2005 WyCAS provides information in reading, writing and math. Based on students' performance, they are categorized into four levels: Advanced, Proficient, Partially Proficient, and Novice. Here, we interpolated grades 3, 5, 6, 7, 9 and 10 from grades 4, 8 and 11 for both graphs.

II Analysis of the State Results

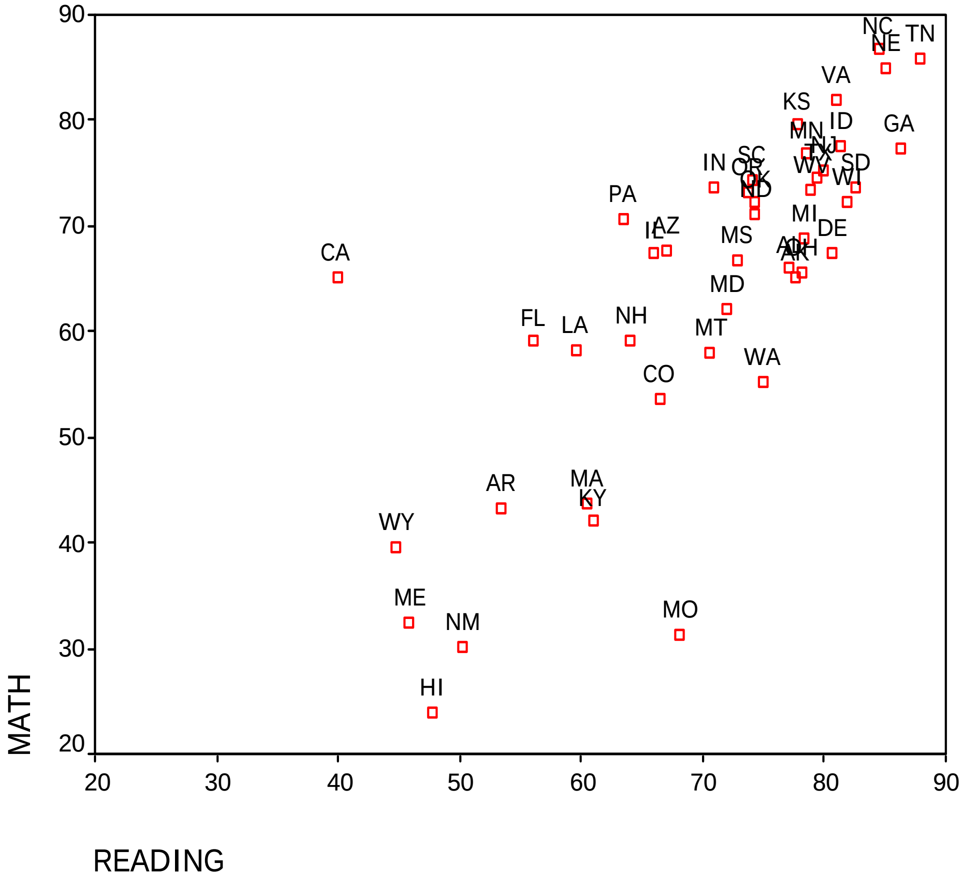
Two features of the state graph seem particularly interesting: 1) the nature of the state's "proficient" line; 2) consistency between the state's proficient line and NAEP.

1. Analysis for State Proficient Line

A. Horizontal and Vertical Moderation

Lissitz & Huynh (2003) introduced the term "vertical moderation" to refer to the degree of trend in what might be called "idealism vs. realism" in a state's proficiency standards across grade levels. Schafer (2005) extended the concept to "horizontal moderation," which represents the degree of consistency in idealism vs. realism among content areas. In order to study vertical and horizontal moderation, we estimated the regression line, regressing percent proficient on grade level. The slope and sixth-grade intercept of each state proficient trend line as calculated from raw data are summarized as following two scatter plots:

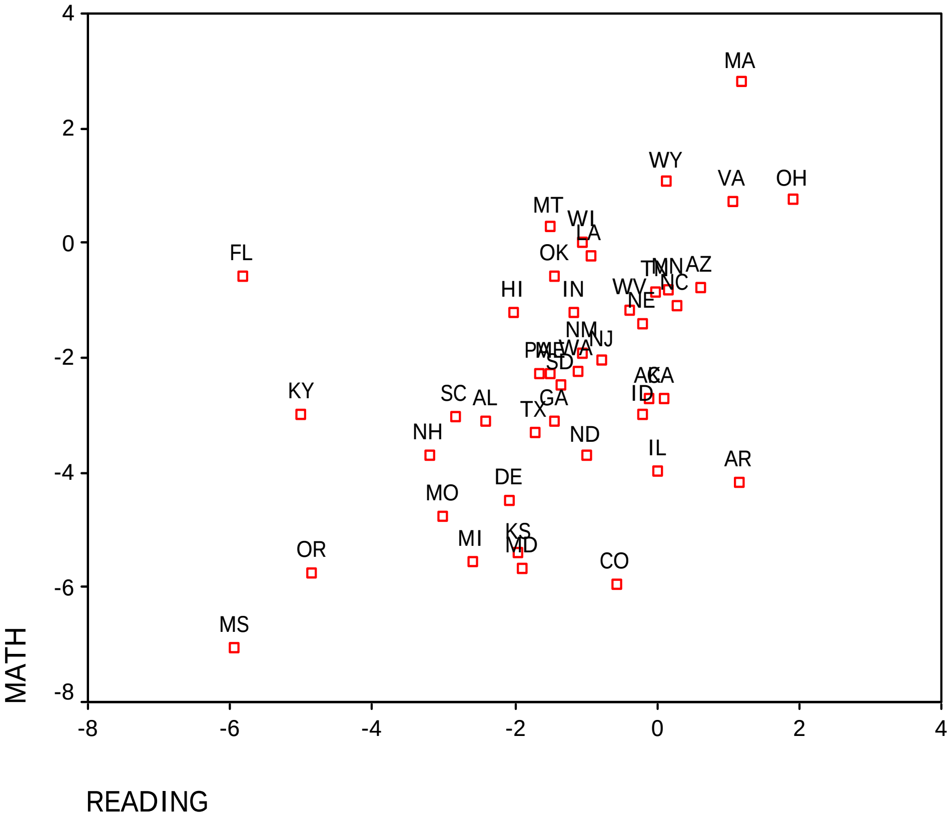
Figure 44.I: **Distribution of predicted value for grade 6 based on regression on the state proficient trend line across 3-10 grades in reading and math for 43 states**



** Correlation =.784

The sixth-grade intercept can represent a typical status (at grade 6) of the state student performance trend across grade 3 to grade 10. From the graph, it is clear that states vary greatly on the typical percentage of "state proficient and above" line. For example, NC, NE and TN are over 80% in both reading and math while HI is lower than 30% in math. We can also infer that the typical statuses of the trend lines are consistent across contents (reading and math). That is, states with higher typical values of trend line for reading tend also to have higher values for math. The Pearson correlation coefficient (r=.784, p<.01) between reading and math documents this consistency.

Figure 44.S: Slopes of state proficient trend line across 3-10 grades in reading and math for 43 states



** Correlation =.512

The slope can represent the trend of state student performance line across grade 3 to grade 10. A positive slope means that percents proficient increase with increasing grade level and vice versa. From the graph, we can infer that states vary greatly on the trend of "state proficient and above" line. For instance, MA is over 1 for both reading and math while MS and OR are both lower than -4 for both subjects. That is, for some states the degree of idealism in performance standards increases over grade levels (those with negative slopes), while for others the degree of realism in performance standards increases over grade levels (those with positive slopes).

However, most states appear to be in the negative-negative quadrant, meaning that their performance standards become more idealistic as grade level increases in both content areas.

We can also conclude that the slope of trend line is typically consistent across the two contents (reading and math), suggesting horizontal moderation. States with higher slopes for reading tend to have higher slopes for math, but not as strongly as for the intercept comparison. The Pearson correlation coefficient was .512 ($p<.01$). However, there are some states that show clear lack of horizontal moderation in trend; AR, FL, and KY stand out as possible outliers in the graph. In these states, there appears to be interaction between horizontal and vertical moderation; that is, the trend across grades in one subject is different from the trend in the other subject.

B. Degree of articulation (vertical moderation)

To study the degree of vertical moderation that states show, we calculated the average distance of the grade-level points (since high schools seem often to have divergent results from the other grades, only the 6 points from grade 3 to grade 8 were used) from the regression line. In other words, we calculated the sum of squared difference between observed proficient percent and predicted proficient percent from the regression line (the sum of squares for residuals) and then found the square root of the average sum of squares to get the standard deviation, which represents the degree of articulation. We generated this root mean square for regression residuals across grades in both reading and math for each state. This is a measure of the "smoothness" of the vertical moderation for a given content-state combination.

Figure 45: Root mean square for regression residuals across grades 3-8 in reading and math for 43 states.

** Correlation = .428

Most states seem to cluster in a range where there were small deviations in percents proficient from regression line, having a root mean square error less than 3. There are also some clear outliers, such as TX. It should be noted that for some cases, the degree of vertical moderation is artificially high (and the corresponding measure near zero) since some states have missing data on two or more grades, such as MO and NE, and we estimated intermediate points linearly.

States that are articulated in reading also tend to be articulated in math; r=.428 (p<.01). However, some states such as GA and CA appear as outliers in the figure.

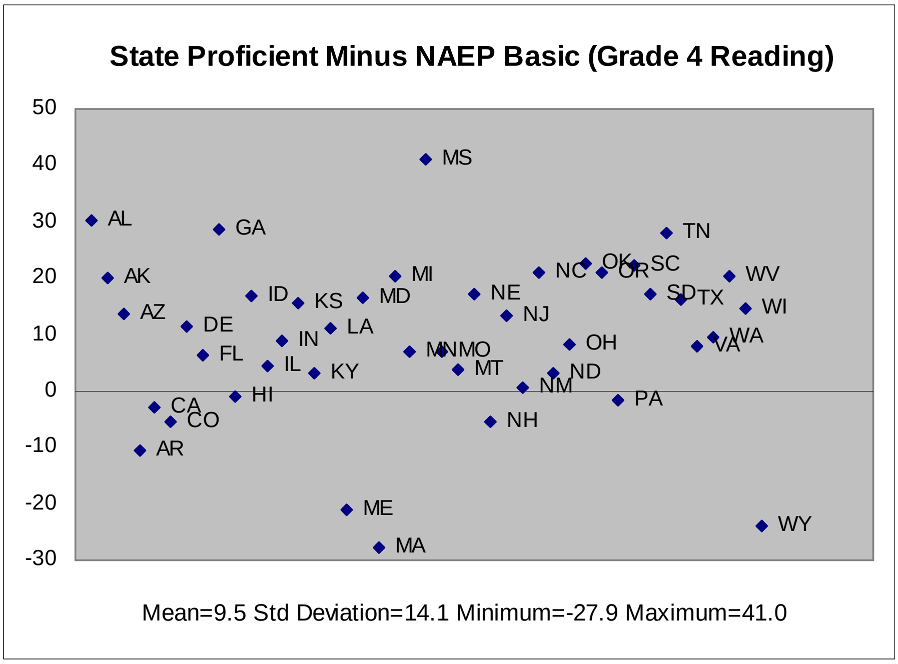
2. Consistency with NAEP Basic and Proficient

This section contains two parts. One is a series of one-dimensional analyses in which the differences between state proficiency and the two NAEP standards (basic and proficient) were analyzed in each subject and each grade, respectively. The second part is a series of two dimensional analyses in which we study the relationship between grade levels in one subject and the relationship between subjects in one grade level.

A. One-Dimensional Analyses

Differences between state percent at and above proficient vs. NAEP percent at and above standard were calculated by subtracting the NAEP percent from the state percent (so that a positive number indicates that the state test showed more students above the cut point than NAEP). These differences are displayed in scatter plots with state codes and in histograms. In the scatter plots, we use 0 as a reference line to show the direction and the distance between state and NAEP standards. In addition, the histograms show the shape of the distribution of those differences, with a normal curve used as a reference.

Figure 46B4r: Consistency with NAEP Basic for Grade Four Reading



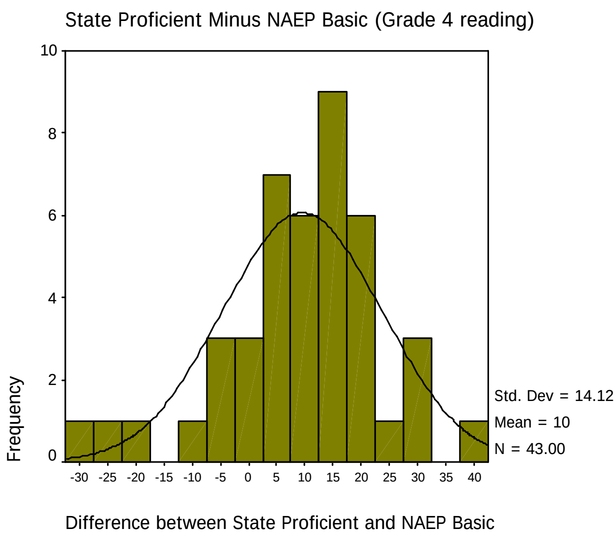
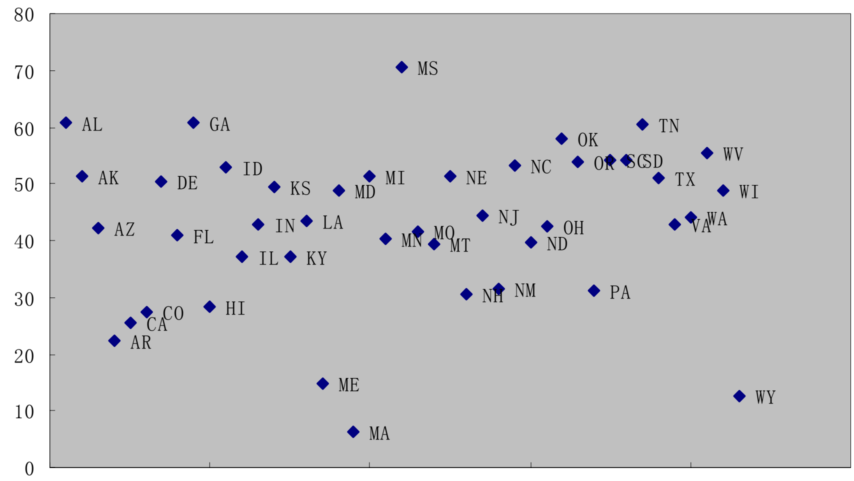


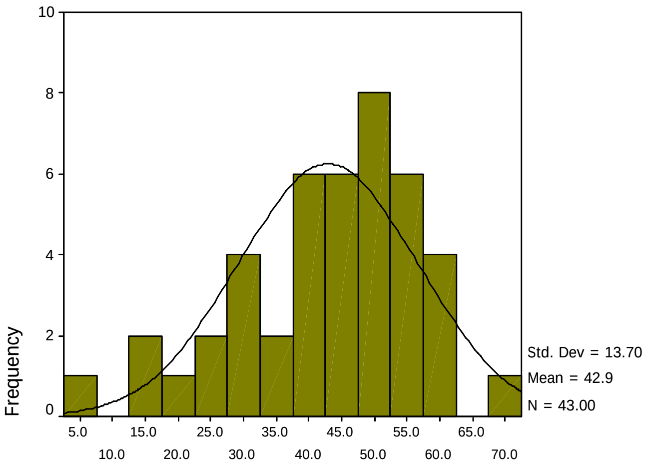
Figure 46P4r: Consistency with NAEP Proficient for Grade Four Reading

State Proficient Minus NAEP Proficient (Grade 4 Reading)



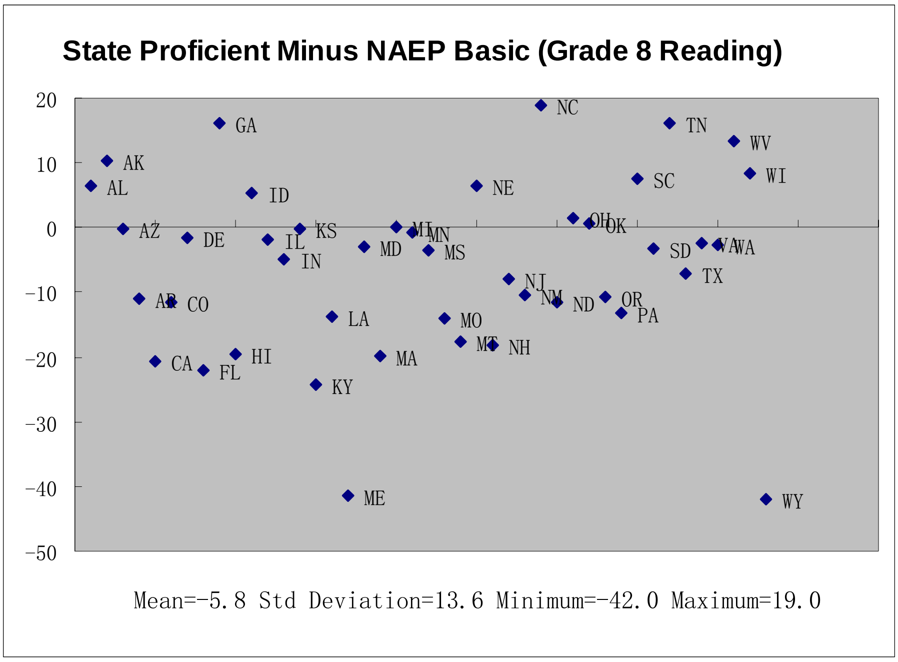
Mean=42.9 Std Deviation=13.7 Minimum=6.3 Maximum=70.6

State Proficient Minus NAEP Proficient (Grade 4 reading)



Difference between State Proficient and NAEP Proficient

Figure 46B8r: Consistency with NAEP Basic for Grade Eight Reading



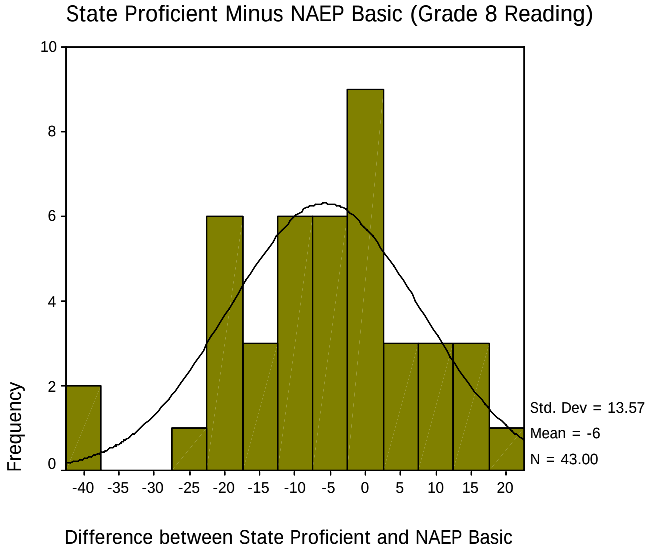
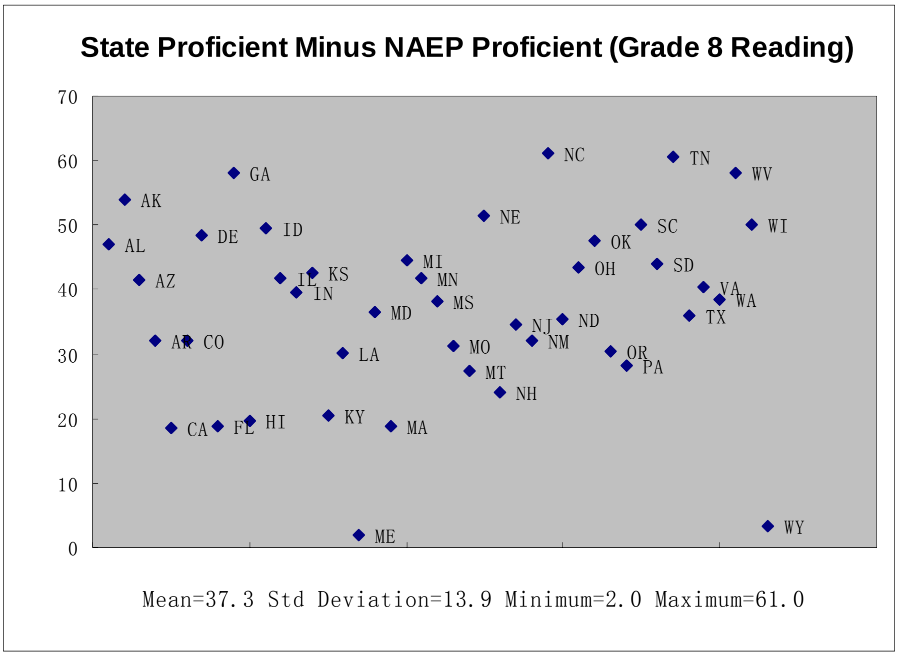


Figure 46P8r: Consistency with NAEP Proficient for Grade Eight Reading



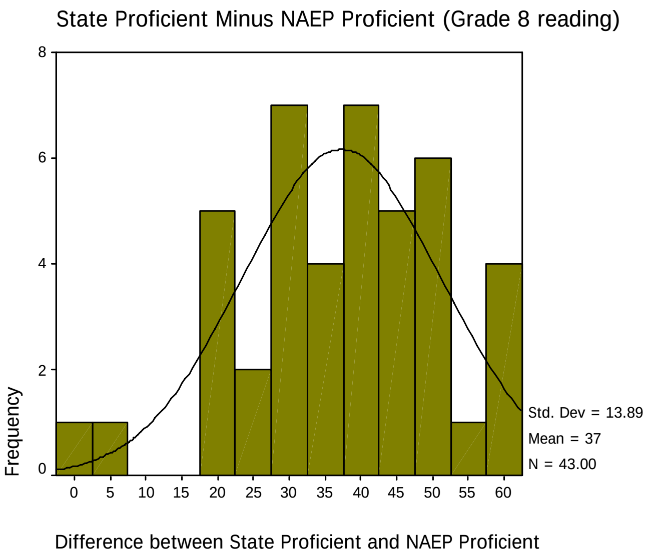
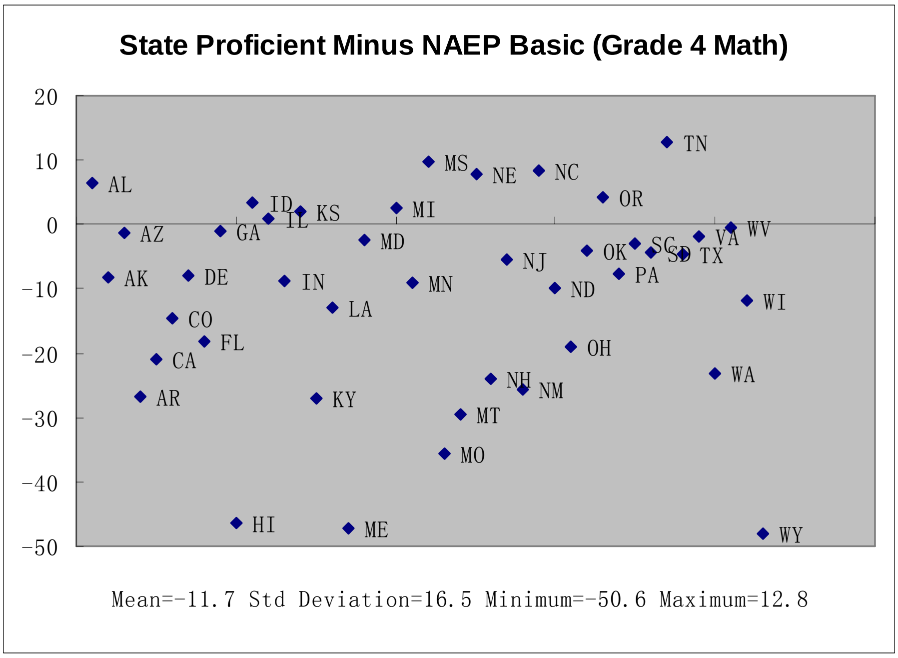


Figure 46B4m: Consistency with NAEP Basic for Grade Four Math



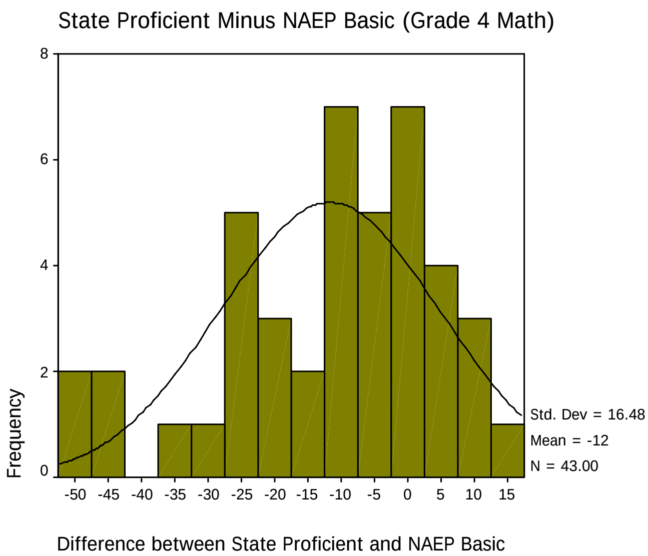
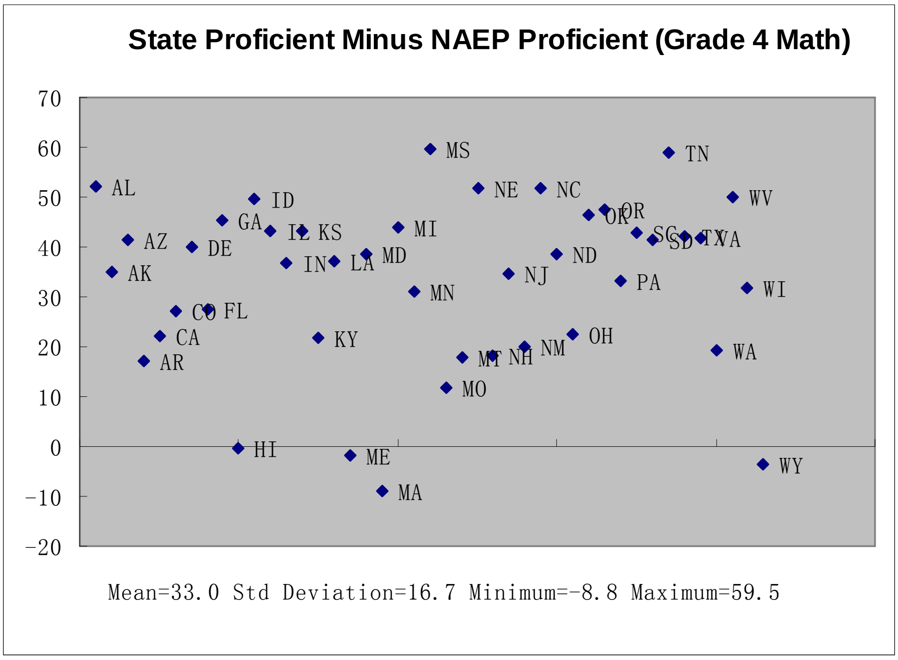


Figure 46P4m: Consistency with NAEP Proficient for Grade Four Math



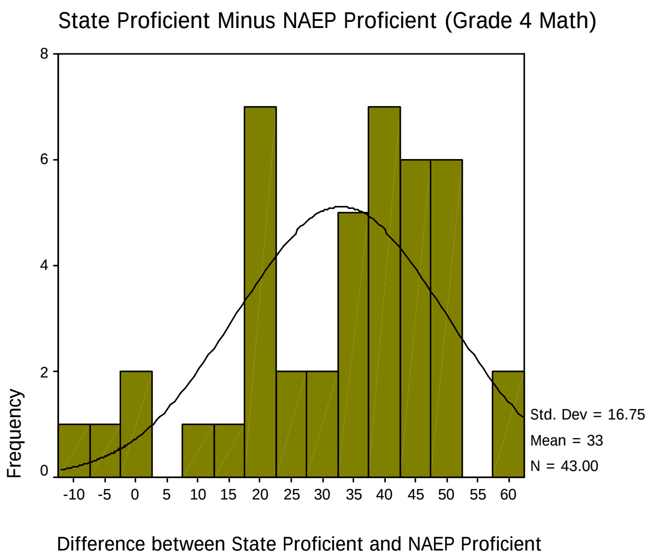
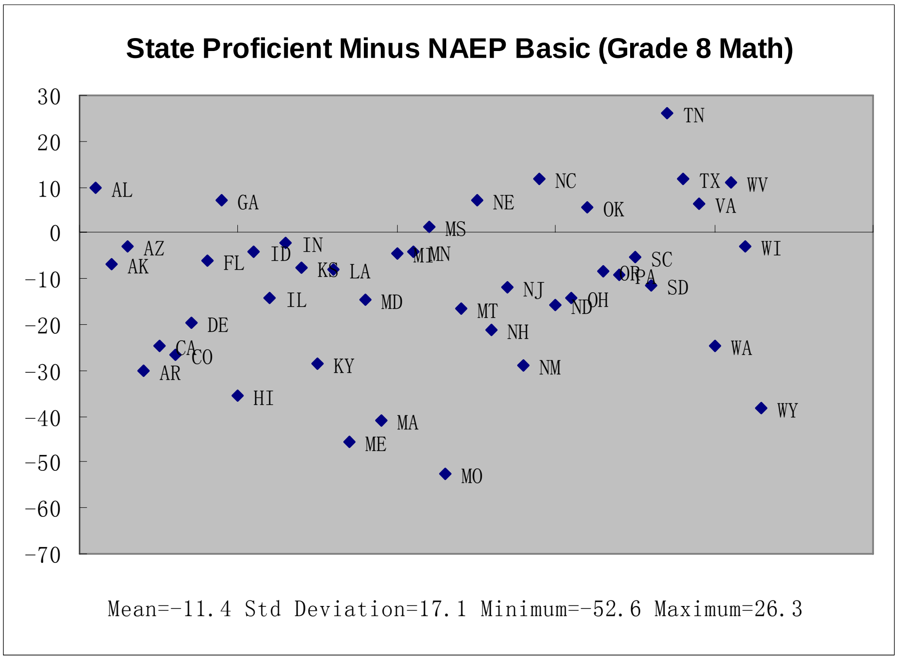


Figure 46B8m: Consistency with NAEP Basic for Grade Eight Math



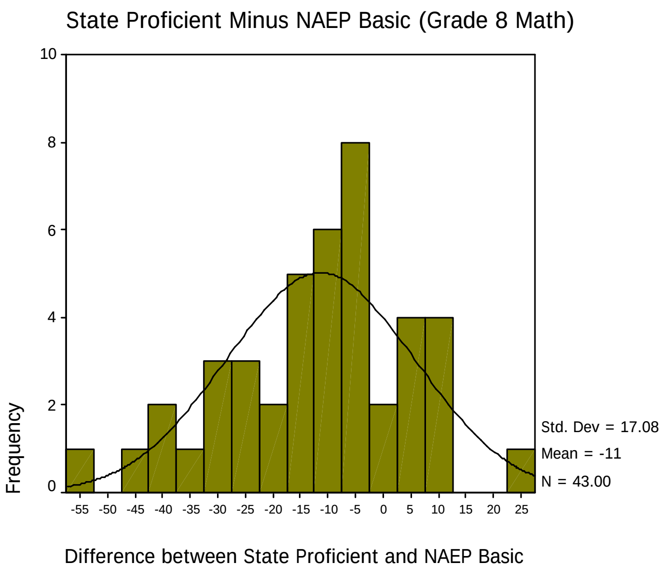
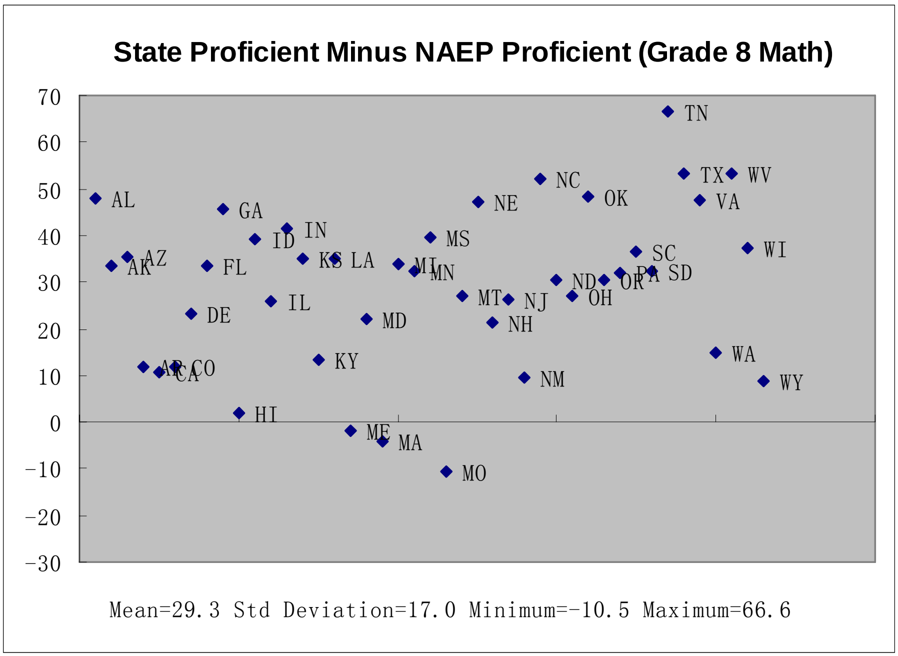


Figure 46P8m: Consistency with NAEP Proficient for Grade Eight Math



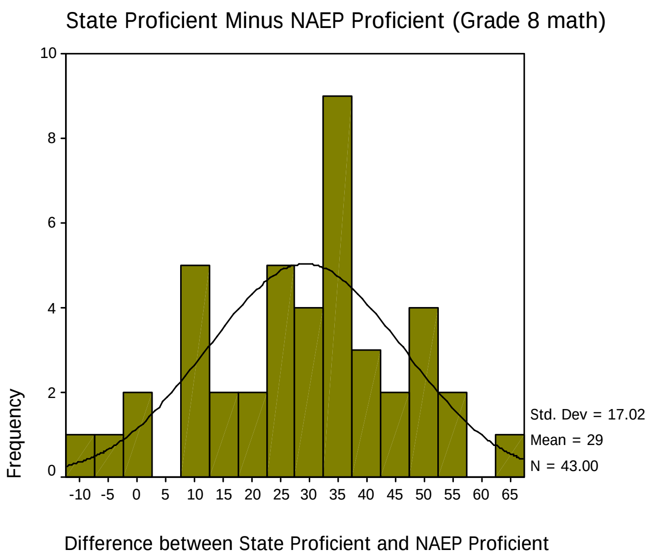


Table 1 summarizes the descriptive statistics for the differences appearing in Figure 46 between the percentages at and above the state proficient level minus the NAEP basic or proficient level. In reading, the average state percent approximates the average NAEP percent for basic; the average differences are within ten percentage points. Differences between state percentages and percentages for NAEP proficient are far larger. Thus, average state proficient levels, on average, are approximately equally realistic (or idealistic) as the NAEP basic standard for reading and more realistic (less idealistic) than the NAEP proficient standard.

Average differences between states and NAEP in math show a somewhat different pattern. State percentages appear to be smaller on average than NAEP basic and larger than NAEP proficient. Thus, state proficiency levels appear to be more idealistic than NAEP basic but more realistic than NAEP proficient. Also, generally speaking, the magnitude of standard deviation in math is larger than that in reading.

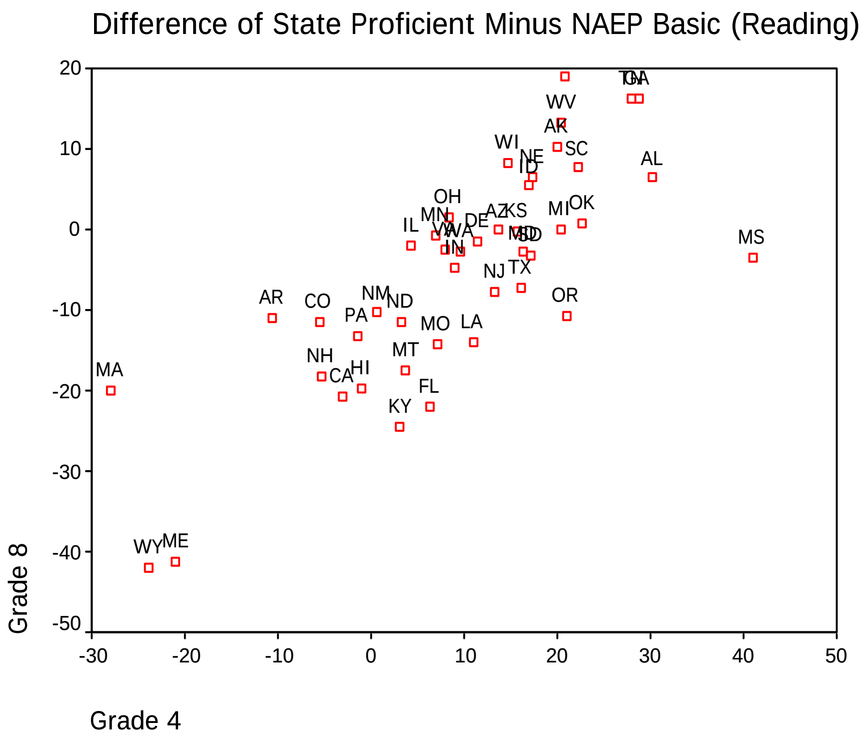
Table 1: Descriptive Statistics for the Differences Between State and NAEP Standards

Descriptive Statistics					
State Minus NAEP(%)	N	Minimum	Maximum	Mean	Std. Deviation
Basic Grad4 reading	43	-27.91	40.96	9.52	14.12
Profict Grad4 reading	43	6.33	70.62	42.92	13.70
Basic Grad8 reading	43	-42.00	18.97	-5.81	13.57
Profict Grad8 reading	43	1.99	61.02	37.29	13.89
Basic Grad4 math	43	-50.59	12.77	-11.69	16.48
Profict Grad4 math	43	-8.78	59.47	33.03	16.75
Basic Grad8 math	43	-52.57	26.25	-11.36	17.08
Profict Grad8 math	43	-10.52	66.64	29.27	17.02

B. Two-dimensional Analyses

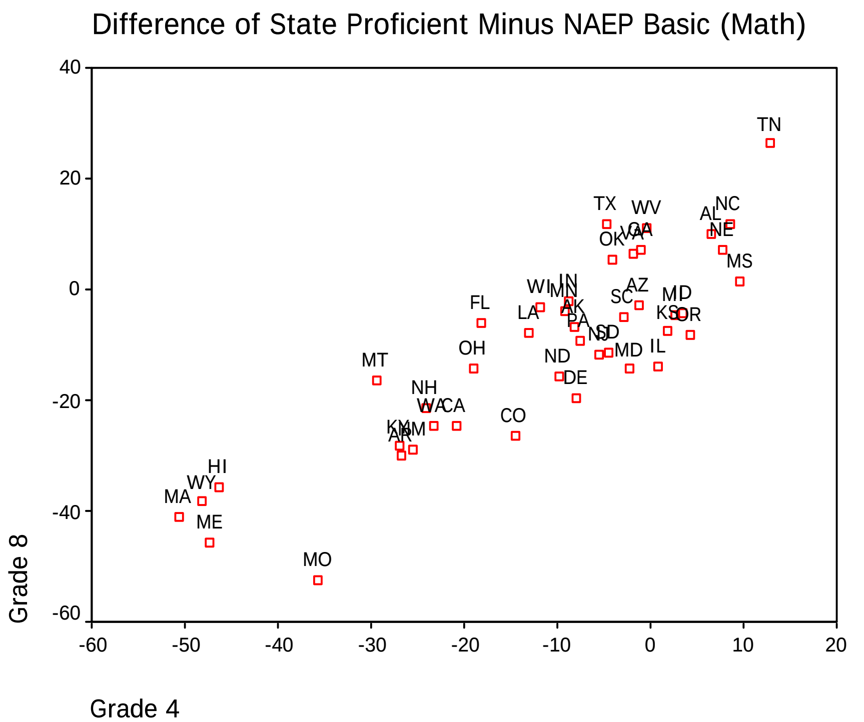
1). NAEP Basic

Figure 47Br: Grade Consistency of State vs. NAEP Basic in Reading



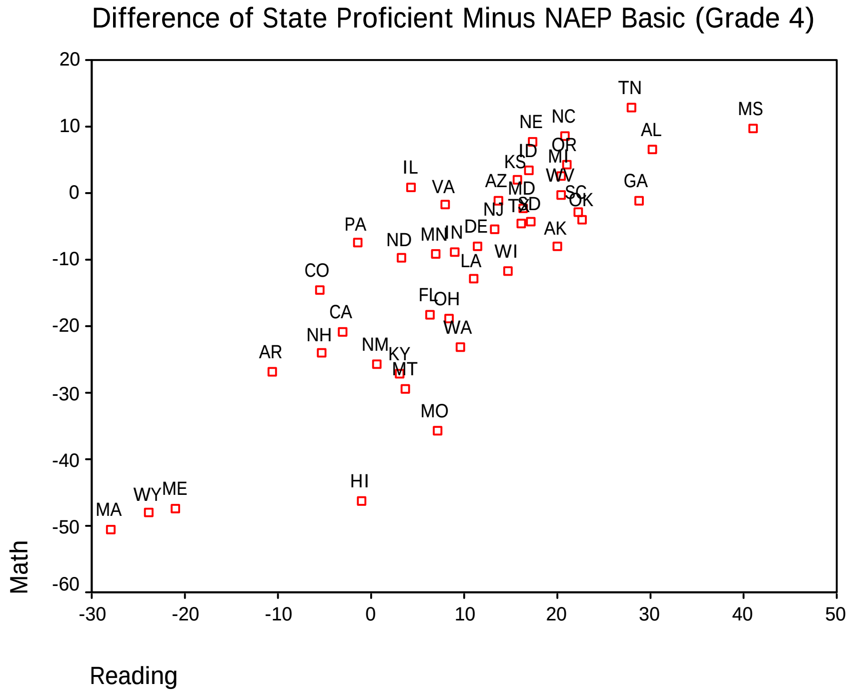
** Correlation = 0.784

Figure 47Bm: Grade Consistency of State vs. NAEP Basic in Math



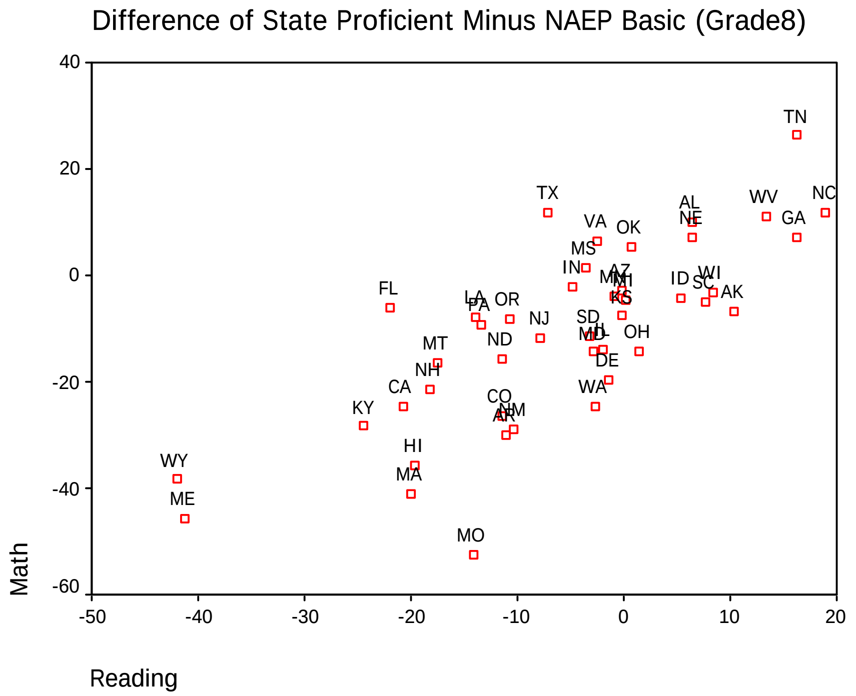
* Correlation = 0.866

Figure 47B4: Content Consistency of State vs. NAEP Basic at Grade 4



** Correlation = 0.846

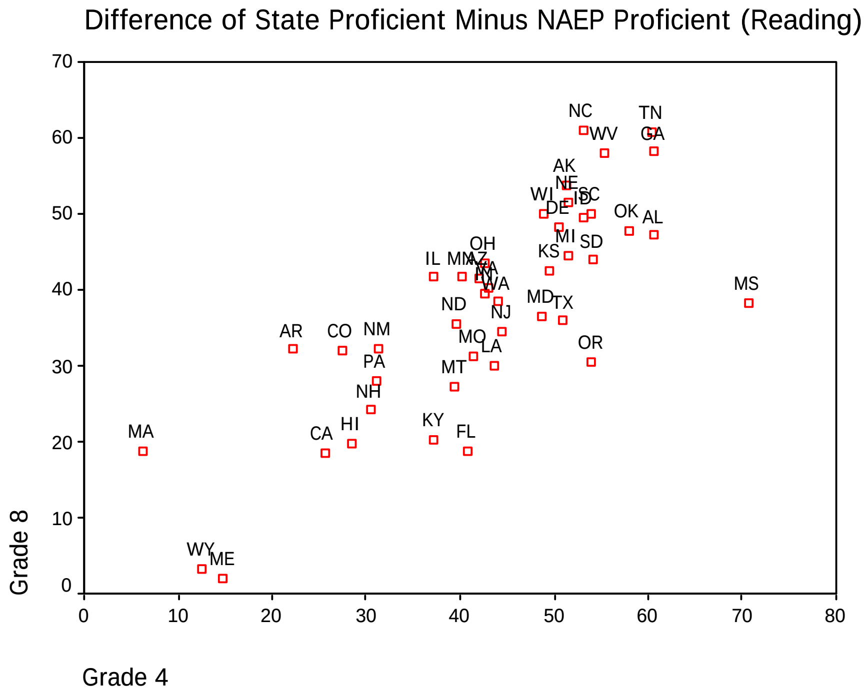
Figure 47B8: Content Consistency of State vs. NAEP Basic at Grade 8



** Correlation = 0.763

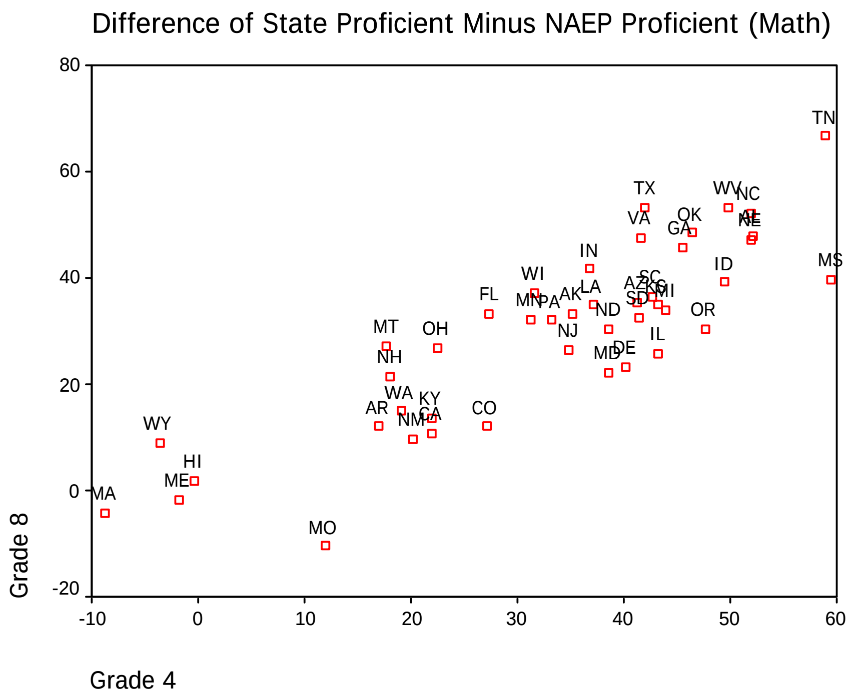
2). NEAP Proficient

Figure 47Pr: Grade Consistency of State vs. NAEP Proficient in Reading



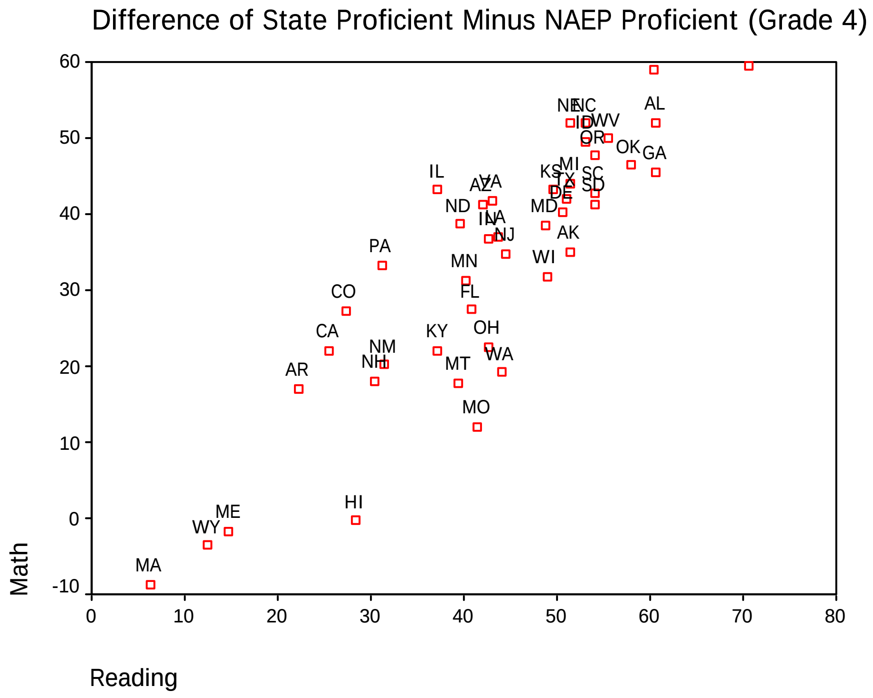
** Correlation = 0.793

Figure 47Pm: Grade Consistency of State vs. NAEP Proficient in Math



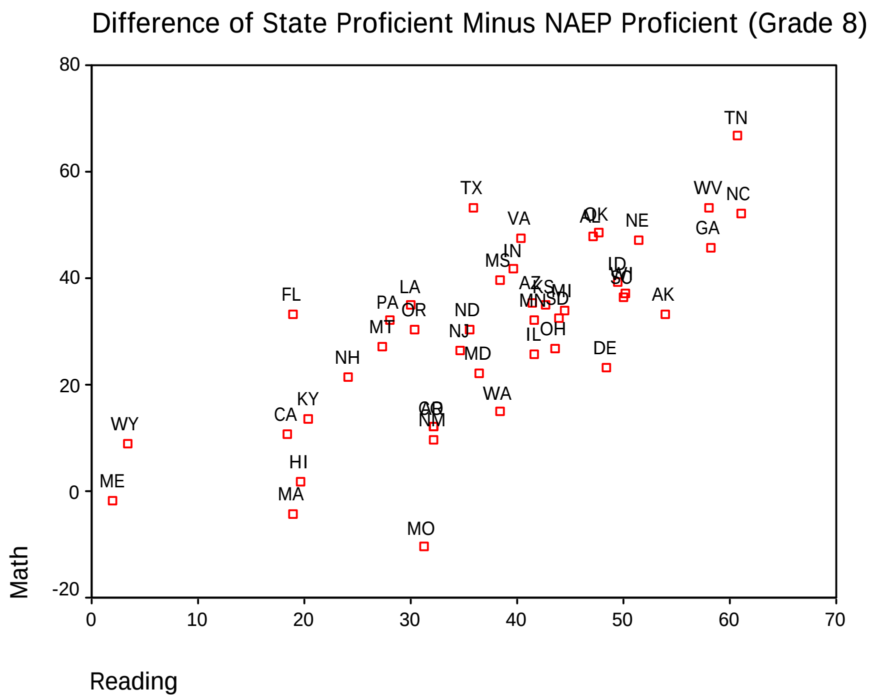
** Correlation = 0.863

Figure 47P4: Content Consistency of State vs. NAEP Proficient at Grade 4



** Correlation = 0.882

Figure 47P8: Content Consistency of State vs. NAEP Proficient at Grade 8



** Correlation = 0.741

Discussion and Conclusions

We used the typical (sixth-grade) predictions of the regression lines of percent proficient on grade level to evaluate the consistency of state expectations for reading and math and compared them in a graphic (see Figure 44I). When expectations are equivalent for reading and math, the state would appear on the diagonal in Figure 44I. Those that are well away from the diagonal run a risk of over-identifying schools as in need of improvement in one content area and a corresponding risk of under-identifying schools in the other content. This may have implications for asymmetric resource allocations.

We used the slopes of the regression lines of percent proficient on grade level to evaluate one aspect (direction) of vertical moderation. If the slopes are flat, then grade levels are approximately equally likely to be the cause of identification of a school as in need of improvement. Those states that show flat regression lines in both contents are near the origin (0,0) point in Figure 44S. But if the slopes are not flat, then we can evaluate whether the same increasing or decreasing pattern is or is not in common between the two contents. If the pattern is the same, then the state should appear on the diagonal in Figure 44S.

Another aspect of vertical moderation (smoothness) can be evaluated by whether the percents proficient are predictable from grade-to-grade, whether the overall pattern is increasing, decreasing, or flat. In Figure 45, we evaluated deviations from regressions of percents proficient on grade three to grade eight. States that are near the origin show greater smoothness in vertical moderation and most states appear to cluster there.

There appears to be a trend toward using NAEP Basic as opposed to NAEP Proficient as the benchmark for states' Proficient categories. Although it could be argued that, in the light of the striking variation along the idealistic-realistic dimension among states, there is little value in answering this question for states in general; Figure 46 and Table 1, which show deviations from the two NAEP achievement level percentages, suggest that NAEP Basic is more consistent with states' Proficient than is NAEP Proficient.

It could be argued that states may not show much consistency between contents and between grades because there are real differences in instructional effectiveness in one or both these directions. If so, then perhaps using NAEP percentages as benchmarks and comparing states' differences with them could become interesting. Figure 47 displays those results. Vertical moderation is shown in Figures 47Br, 47Bm, 47Pr, and 47Pm; states close to the diagonal show vertical moderation when NAEP is taken into account. Horizontal moderation is shown in Figures 47B4, 47P4, 47B8, and 47P8; states close to the diagonal show horizontal moderation when NAEP is taken into account.

Recommendations and Limitations

We have determined that states differ significantly in the various indices that we have developed for evaluating states' vertical and horizontal moderation and in their relationship with NAEP results. The value of these analyses in evaluating the standards set in states will in part depend on how they are related with the results of educational reform efforts that states undertake. Are idealistic expectations motivating or discouraging; are realistic expectations motivating or encouraging? There do not appear to be clear ways to answer these questions at this time. Perhaps this paper can suggest ways to compare states on the characteristics we have described. Researchers are encouraged to quantify and study differences like those we have found as predictors of improvement as measures using states' results and/or NAEP results from future administration as correlates of change.

The current importance of these results to states depends on their idiosyncratic policy goals. Thus, we will not comment on states' individual results. Each state can evaluate its own position in the graphs and determine if any implications for action exist. It is neither our purpose to argue the value of any particular configuration of percents proficient nor any particular comparison between a state's results and those from NAEP.

The comparisons we made between state and NAEP results should be interpreted in the light of several limitations. Among these are:

Possible instability of NAEP results at grades four and eight. Sampling errors in percents proficient and basic were not incorporated into our analyses.

Possible instability of state results at all grades. Even though all students in the state are included and therefore statistics are relatively free of student sampling error, sampling error in content coverage (alignment) for different test forms can affect student scores.

Changes in the enacted curriculum. As teachers and their supervisors react to tests that are more and more important to them, their curricular, and therefore instructional goals may become more consistent with those expressed by the state.

Content equivalence between state and NAEP content. While states commonly use NAEP (along with other) frameworks to validate their own, it is possible that NAEP does not represent the content of all states in a way that is fair to all, which can affect the validity of the comparisons we have made.

Cut score equivalence. Some states are still in the process of determining cut scores for their proficient achievement levels. Therefore, the data for some states may be out-of-date.

Data errors. Public web sites were used to generate all the data used in this report. Each of these sites used its own means of expression of its data. While care was taken to ensure accuracy, it is possible that errors inadvertently were introduced into the results. We apologize in advance if this may have occurred.

References

Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). *Is the No Child Left Behind Act working?* Policy Analysis for California Education, Berkeley.

Jones, L. V., Olkin, I., & American Educational Research Association. (2004). *The nation's report card: : evolution and perspectives*. Bloomington, IN: Published by the Phi Delta Kappa Educational Foundation in cooperation with the American Educational Research Association.

Linn, R. L. (2000). Assessments and Accountability. *Educational Researcher, 29*(2), 4-16.

Linn, R. L. (2003). Performance standards: Utility for different uses of assessments. . *Education Policy Analysis Archives*, *11*(31).

Linn, R. L. (2005a). *Adjusting for Difference in Tests*. Paper presented at the Symposium on the Use of School-Level Data for Evaluating Federal EducationPrograms. Retrieved from http://www7.nationalacademies.org/bota/School-Level%20Data_Robert%20Linn-Paper.pdf.

Linn, R. L. (2005b). *Issues in the design of accountability systems CSE Technical Report 650*. Paper presented at the 2005 National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Retrieved. from http://www.cse.ucla.edu/reports/r650.pdf.

Lissitz, Robert W. & Huynh Huynh (2003). Vertical equating for state assessments: issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation*, 8(10). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=10 .

McLaughlin, D. (2005).*Considerations in using the longitudinal school-level state assessment score database*. Paper presented at the Symposium on the Use of School-Level Data for Evaluating Federal Education Programs. Retrieved from http://www7.nationalacademies.org/bota/School-Level%20Data_McLaughlin-Final.pdf.

McLaughlin, D., & Mello, V. B. d. (2002). *Comparison of State Elementary School Mathematics Achievement Standards, Using NAEP 2000*. Paper presented at the American Educational Research Association (AERA). Retrieved from http://208.253.216.16/assessment/Reports/AERA_2002_state_standards.pdf.

Mosquin, P. & Chromy, J. (2004). *Federal sample sizes for confirmation of state tests in the No Child Left Behind Act*. Washington, DC: American Institutes for Research. Retrieved from http://www.air.org/publications/documents/MosquinChromy_AIR1.pdf.

National Center for Education Statistics (2003). *The nation's report card: Reading highlights 2003*. Washington, DC: Author. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2003/2004452.pdf.

National Center for Education Statistics (2005). *The nation's report card: Reading highlights 2005*. Washington, DC: Author. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2005/2006451.pdf.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (1999). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*: Washington, D.C.: National Academy Press.

Schafer, W. D. (2005). Criteria for standard setting from the sponsor's perspective. *Applied Measurement in Education, 18*(1), 61-81.

Shepard, L. A., Glaser, R., Linn, R. L., & Bohrnstedt, G. W. (1993). *Setting Performance Standards for Students Achievement: A Report of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement levels.*: Stanford, Calif.: National Academy of Education.

Appendices

Appendix 1. Data sources:

#		State	URLs for state assessment
1	AK	Alaska	http://www.educ.state.ak.us/tls/assessment/results/2005/2005StatewideSBA.pdf
2	AL	Alabama	http://www.alsde.edu/Accountability/2005Reports/AL2005ARMT_0624051.pdf?lstSchoolYear=3&lstReport=2005Reports%2FAL2005ARMT_0624051.pdf
3	AR	Arkansas	http://130.184.43.9/reportcards/state05.php
4	AZ	Arizona	http://www.ade.state.az.us/srcs/statereportcards/statereportcard04-05.pdf
5	CA	California	http://star.cde.ca.gov/star2005/viewreport.asp?ps=true&lstTestYear=2004&lstTestType=C&lstCounty=&lstDistrict=&lstSchool=&lstGroup=1&lstSubGroup=1
6	CO	Colorado	http://www.cde.state.co.us/cdeassess/documents/csap/csap_summary.html
7	CT*	Connecticut	http://www.captreports.com/web2005/Summary/ERG/OS.html,(for grade 10 only); State report for NCLB could not be opened: http://www.csde.state.ct.us/public/cedar/nclb/dist_school_nclb_results/index.htm
8	DC*	District of Columbia	There is only combined data at elementary level and secondary level.
9	DE	Delaware	http://www.doe.state.de.us/files/pdf/de_edreportcard200405.pdf
10	FL	Florida	http://fcat.fldoe.org/index.asp#reports
11	GA	Georgia	http://reportcard2005.gaosa.org/k12/reports.aspX?ID=ALL:ALL&TestKey=C*4&TestType=qcc
12	HI	Hawaii	http://arch.k12.hi.us/, then choose "Hawaii State Assessment: 2005 State Results"
13	IA*	Iowa	There is only biennium data for 2003-2005: http://www.state.ia.us/educate/ecese/nclb/doc/reportcard05.pdf
14	ID	Idaho	http://www.sde.state.id.us/ipd/aypassessment05/default.asp
15	IL	Illinois	http://webprod1.isbe.net/ereportcard/publicsite/getSearchCriteria.aspx
16	IN	Indiana	http://www.doe.state.in.us/reed/newsr/2005/12-December/051214prOverview.pdf http://nces.ed.gov/programs/stateprofiles/sresult.asp?mode=full&displaycat=7&s1=18
17	KS	Kansas	http://www.ksde.org/assessment/state_tables_2005.pdf

18	KY	Kentucky	http://apps.kde.state.ky.us/secure_cats_reports_05/index.cfm?fuseaction=main.display_regionstate
19	LA	Louisiana	http://www.doe.state.la.us/lde/uploads/7714.pdf
20	MA	Massachusetts	http://www.doe.mass.edu/mcas/2005/results/summary.pdf
21	MD	Maryland	http://msp.msde.state.md.us/downloadindex.aspx?K=99AAAA
22	ME	Maine	http://www.maine.gov/education/mea/edmea.htm
23	MI	Michigan	http://www.michigan.gov/documents/FALL_2005_STATEWIDE_MEAP_RESULTS_151913_7.pdf
24	MN	Minnesota	http://education.state.mn.us/mde/static/MCA2005Public_Filter9.TAB
25	MO	Missouri	http://dese.mo.gov/divimprove/assess/stateresults.html.pdf; http://dese.mo.gov/divimprove/assess/State_MAP2005_Reading.pdf
26	MS	Mississippi	http://orsap.mde.k12.ms.us:8080/MAARS/indexProcessor.jsp
27	MT	Montana	http://www.opi.state.mt.us/
28	NC	North Carolina	http://disag.ncpublicschools.org/2005/
29	ND	North Dakota	http://www.dpi.state.nd.us/resource/biennial.PDF
30	NE	Nebraska	http://reportcard.nde.state.ne.us/Page/PerfStandardOverall.asp?PerfStandard&Level=st&Subject=2 http://reportcard.nde.state.ne.us/Page/PerfStandardOverall.asp?PerfStandard&Level=st&Subject=1
31	NH	New Hampshire	http://www.ed.state.nh.us/education/doe/organization/curriculum/NECAP/NECAP_results.htm (For grade 3-8) http://www.ed.state.nh.us/education/doe/organization/curriculum/Assessment/2005/Sta.pdf (For grade 10)
32	NJ	New Jersey	http://www.nj.gov/njded/schools/achievement/2006/njask3/summary.pdf (For grade 3, 4); http://www.nj.gov/njded/schools/achievement/2006/gepa/summary.pdf (For grade 8); http://www.nj.gov/njded/schools/achievement/2006/hspa/summary.pdf (For grade 11).
33	NM	New Mexico	http://www.ped.state.nm.us/press/2005/august/assesment_samples/Test%20Results%2081805.pdf
34	NV*	Nevada	The most recent data we can find is in 2004.
35	NY*	New York	No data for 2005
36	OH	Ohio	http://www.ode.state.oh.us/proficiency/results.asp
37	OK	Oklahoma	http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf
38	OR	Oregon	http://www.ode.state.or.us/data/annreportcard/rptcard200

			5.pdf
39	PA	Pennsylvania	http://www.pde.state.pa.us/a_and_t/cwp/view.asp?a=3&q=115272
40	RI*	Rhode Island	http://131.109.26.252/reportcard/04/
41	SC	South Carolina	http://www.myscschools.com/tracks/testscores/pact/2005/
42	SD	South Dakota	https://sis.ddncampus.net:8081/nclb/portal/portal.xsl?&extractID=7
43	TN	Tennessee	http://www.k-12.state.tn.us/rptcrd05/state2.asp
44	TX	Texas	http://www.tea.state.tx.us/perfreport/aeis/2005/state.html
45	UT*	Utah	http://www.usoe.k12.ut.us/eval/Results.asp
46	VA	Virginia	http://pen2.vak12ed.edu/cgi-bin/broker?_service=doe_prod&_program=prodcode.doerp101rcdp001.sas#gr11
47	VT*	Vermont	http://www.state.vt.us/educ/new/pdfdoc/pgm_assessment/cas/cas_99_05_statewide.pdf
48	WA	Washionton	http://reportcard.ospi.k12.wa.us/
49	WI	Wisconsin	http://data.dpi.state.wi.us/data/graphshell.asp?SubjectID=3MA&GraphFile= ETAIL=YES&Grade=10&Group=AllStudentsFAY&EligibleOnly=NO&Le WOW=WSAS&ORGLEVEL=ST&FULLKEY=ZZZZZZZZZZZZ&DN=N &SN=None+Chosen http://dpi.wi.gov/oea/spr_kce.html
50	WV	West Virginia	https://wveis.k12.wv.us/nclb/pub/rpt0405/chartdata05rcs.cfm?year=05&county=999&school=999&coname=&rptnum=1&rpage=pickreportcard.cfm
51	WY	Wyoming	http://www.k12.wy.us/SA/WyCAS/wycas05/state05.asp
52	PR*	Puerto Rico	The link could not be opened: http://eduportal.de.gobierno.pr/EDUPortal/default.htm

* The state without sufficient data for 2005

Appendix 2. The Comparison of Different Methods to Measure the Degree of Articulation.

Purpose:

For the vertical moderation assessments in our paper, we faced a choice of how to model predictions around which we would calculate residuals that we would interpret as showing lack of moderation. Recall that these residuals were squared and averaged in finding the root mean square residual that we used as our index of lack of moderation. We chose to use a linear regression line for the data from grades 3-8 in the paper. In this appendix, we compare 2 methods, namely simple linear (straight line) and linear regression, which we use to predict (smooth) state proficient percents based on two different upper series end-points, the data of grade 3-8 and grade 3-10, respectively, and thus to decide which method can measure degree of articulation (vertical moderation) more appropriately.

Methodology:

Based on theory of Vertical Moderation Standard (VMS), two conditions are necessary to make a VMS system justifiable and interpretable: first, there is a common set of policy definitions for achievement levels (such as basic, proficient, advanced in NAEP); second, the percentages of students at different achievement levels must achieve some degree of consistency across grades (Huynh & Schneider, 2005). Such consistency considers that instructional efforts are relatively uniform across grade levels, and consequently, the percentages of students at different achievement levels might have increasing or decreasing trend, but should not fluctuate too much across the grades. For this reason, by using a straight trend line as a reference, we could measure the degree of consistency of the state proficient line across grades.

We compared the 2 kinds of straight lines following 2 steps:

Step1. Generate trend lines.

4 trend lines are generated in reading and math, respectively, by using simple straight line and linear regression line.

1. Simple straight Line 3 & 8: use grade 3 and grade 8 only to generate the straight line. In a state graph for a content area, this would appear as the straight line that connects the percents proficient and above at grade 3 and grade 8.
2. Simple straight Line 3 & 10: use grade 3 and grade 10 only to generate the straight line. Similar to (1) above, this straight line connects the percents proficient and above at grades 3 and 10.
3. Linear regression line 3-8: use grade 3- 8 to generate the regression line. In a state graph for a content area, this would be the least-squares regression line that is fit to the six data points for the state percents proficient and above.
4. Linear regression line 3-10: use grade 3-10 to generate the regression line. Similar to (3) above, this regression line is fit to the data for all data points from grades 3-10.

Step2. Calculate RMSE

RMSE (root mean square error) is the square root of the squared average differences between the predicted value from trend line and the observed value for reading and math, respectively. RMSE in our research is used to represent the degree of articulation.

The formulas of RMSE are as follows:

RMSE3-8 = $\sqrt{\sum_{i=3}^{8} (predictions_i - observation_i)^2 \;/\; 6}$

RMSE3-10 = $\sqrt{\sum_{i=3}^{10} (predictions_i - observation_i)^2 \;/\; 8}$

RMSE in our research is used to represent the degree of articulation in order to assess the consistency of trend line across grades.

Results

Table 1 shows the summary statistics of RMSE. In total, 8 results of RMSE are calculated.

Table1. Description Statistics of RMSE for the 43 States

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation
SIM_8_R	43	7.9833	.0000	7.9833	132.2623	3.075867	1.9279340
SIM_8_M	43	8.7369	.0000	8.7369	121.7182	2.830656	2.0819859
SIM_10_R	43	9.7125	.0000	9.7125	166.9710	3.883047	2.3155114
SIM_10_M	43	12.5868	.0000	12.5868	166.5449	3.873137	2.8577804
REG_8_R	43	6.4508	.0000	6.4508	99.2163	2.307355	1.4226893
REG_8_M	43	6.2239	.0000	6.2239	88.0128	2.046809	1.5102377
REG_10_R	43	6.9949	.0000	6.9949	107.7485	2.505780	1.3730788
REG_10_M	43	6.7812	.0000	6.7812	107.1240	2.491256	1.7074980
Valid N (listwise)	43						

Note: SIM=simple straight line; REG= linear regression line; 8= grade 3-8, 10= grade 3-10; R= reading, M=math.

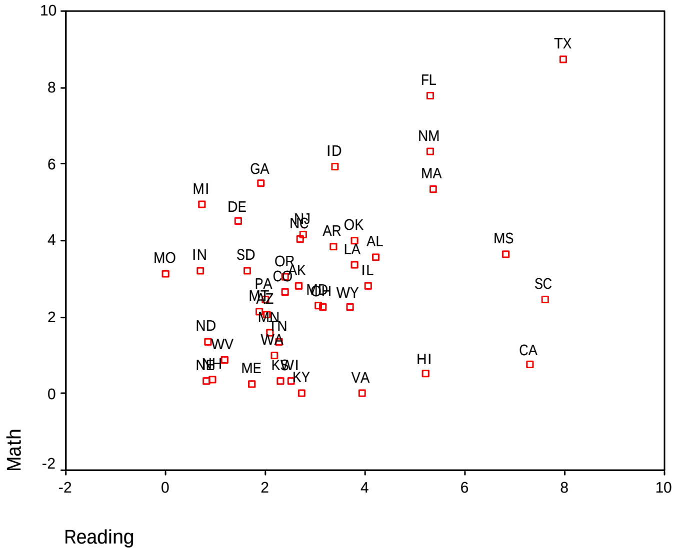
The results suggest two points. First, the regression lines have smaller RMSE than their associated simple straight lines and thus, generally speaking, regression lines work better than simple straight line in terms of prediction. This result is to be expected because the regression line is formed by using the least squares method, which means that the squared differences of the observed points away from the regression line will be minimized across all possible straight lines. Therefore, the magnitude of RMSE by using regression lines as prediction is smaller than that by using simple straight line.

Second, the means of RMSE for grade 3-8 are less than those for 3-10 in reading and math, respectively. This result verifies our concern that high school in some states might not be comparable with grades 3-8. Introducing high school data will increase the magnitude of inconsistency. However, it also exists in several other states and may result from possible multiple uses of high school data in states, such as for graduation decisions.

The following 4 scatter plots (Figures 1-4) display the two methods for the two ranges, grades 3-8 and grades 3-10. Although the correlations are larger for grades 3-10, this result seems to be due to the outlier states of North Carolina in grade 10 for both content areas. These results led us to use the regression line approach, based on only grades 3-8 to calculate the degree of articulation instead of grades 3-10 for all states.

Method1. Simple straight line

Figure 1: RMSE across grades 3-8 in reading and math for 43 states



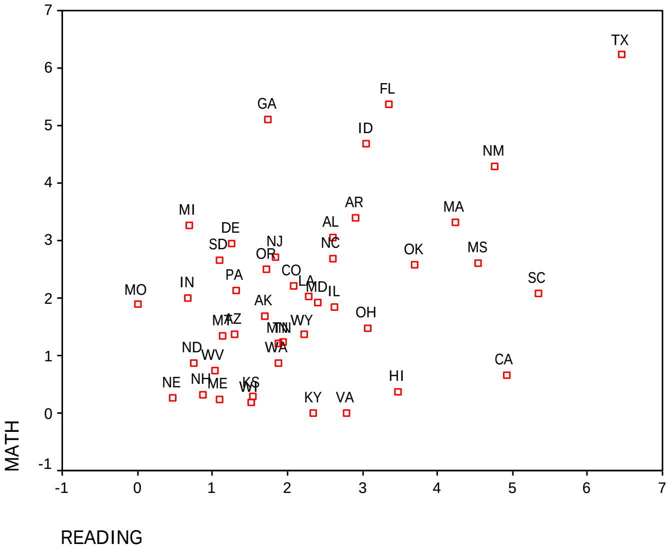
****Correlation = .352**

Figure 2: RMSE across grades 3-10 in reading and math for 43 states

****Correlation = 0.322**

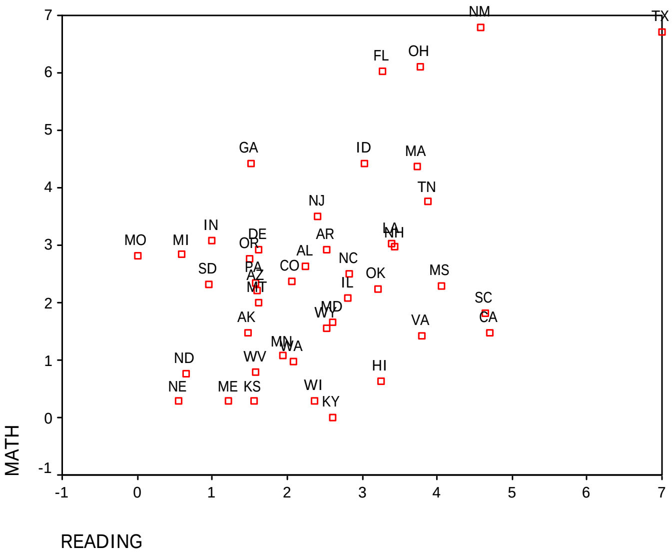
Method2. Linear regression line

Figure 3: RMSE across grades 3-8 in reading and math for 43 states



Correlation = 0.428

Figure 4: RMSE across grades 3-10 in reading and math for 43 states



Correlation = 0.474

The correlations between reading and math are all statistically significant (p<.01). For the 3-8 analyses, the correlation that resulted from the linear regression line was slightly larger than that for the simple straight line.