Asora® Education Enterprises

Mapping Reported Tennessee Public School and District Proficiencies to the NAEP Scale in Math and Reading

David V. Anderson[i]
Asora® Education Enterprises
January 10, 2010

Executive Summary

Nearly every state in the United States administers achievement tests to public school children in the K-12 years to determine, among other things, who is proficient (at or above grade level) in reading and mathematics skills. The United States federal government also administers the National Assessment of Educational Progress (NAEP), also known as the Nation's Report Card, which likewise measures the percentages of children who are NAEP proficient in these same two areas. The criteria for proficiency vary widely among these tests.

Important characteristics of the NAEP include:

- A long track record of 39 years, establishing itself as a de facto national standard for achievement.
- Content standards tested by the NAEP are drawn from the National Assessment Governing Board.
- The NAEP achievement level of "proficient" arguably defines and measures what it is to be at "grade level."
- NAEP tests children in 4th and 8th grades. It reports proficiencies statewide, but not locally.
- In 12th grade NAEP tests only at the national level.
- The NAEP's use of statistical sampling techniques prevents reporting at district and school levels.

State-administered achievement tests, including Tennessee's, are characterized by:

- Testing sufficient numbers of children to report scores and proficiencies at the school and district levels.
- Testing in a variety of subject areas, including those of mathematics and reading.
- With few exceptions, setting the achievement proficiency standards markedly lower than those of the NAEP.
- Tennessee, on average, inflates♥ its 4th and 8th grade proficiencies by a factor of 3.38.
- For its Disadvantaged students, this inflation factor is measured to be 5.67.

As we noted, the NAEP exam scores are not available for individual school districts or schools. Thus, local stakeholders are left in a quandary, not knowing their proficiency percentages in terms of NAEP standards.

♥ Inflation is defined to be the ratio of the state reported proficiency percentage to that of the NAEP.

In this report we:

- Apply a method to convert state-reported proficiencies to more realistic NAEP-aligned estimates.
- Discuss why the method's accuracy is sufficient to ensure reliable results.
- Use the method on demographic groups to assess error levels and examination irregularities.
- Make available NAEP proficiency estimates for all public schools and districts within Tennessee.

These NAEP scale estimates may prove useful to stakeholders in Tennessee public education. It is outside the scope of this report to focus on remedies for some of the problems it illuminates, such as social promotion. Yet a fundamental prerequisite to reform is good information. One obvious step in that direction is relatively simple to accomplish: namely, remove the inflation from the Tennessee Comprehensive Assessment Program (TCAP) tests or otherwise inform stakeholders of its effects. In the interim, the NAEP estimates provided here can reasonably substitute for that "good information."

Mapping Reported Tennessee Public School and District Proficiencies to the NAEP Scale in Math and Reading

Table of Contents

Introduction to Tennessee Mappings 4
Most states report highly inflated proficiencies 4
How we can obtain NAEP-estimated proficiencies locally 6
Report preview 6

Results for Tennessee Public Schools and Districts 6
The maps for grades 4, 8, and 12 6
Using the maps 13
Tabular results: NAEP scale estimates in Tennessee 14
What do these estimates say about student proficiency and testing regimes? 17

Summary and Conclusions 18
What we have done 18
A note on private schools 18
A clearer picture of Tennessee public schools emerges 19
Better testing is foundational to instructional and other reforms 20
A word of encouragement 21

Appendix A: Proficiency Mapping Methodologies 21
Context of the mappings 21
Assumptions leading to ELQ mappings 24

Appendix B: Generating High School Mappings 27
The high school maps require additional approximations 27

Appendix C: Accounting for Errors 28
Measured errors 28
Expected errors 29
Error profiles in tabular format 31
Six kinds of errors 37

Endnotes 38

Introduction to Tennessee Mappings

Most states report highly inflated proficiencies

NAEP Proficient means performing at grade level

The examinations of the National Assessment of Educational Progress (NAEP) are achievement tests that, among their other purposes, measure if students are performing at grade level in mathematics and reading. Generally, students are tested in 4th, 8th, and 12th grades. In each examination there are four possible performance designations:

1. **Below Basic:** The student performs more than one year below grade level
2. **Basic:** The student performs essentially one year below grade level
3. **Proficient:** The student performs approximately at grade level♠
4. **Advanced:** The student performs beyond grade level

The more detailed definitions used by NAEP officials are given in a glossary maintained on the NAEP website.[ii] In recent years, educators who favor differing college and workforce preparation standards have argued that the Basic level should be equated with grade level performance and that the Proficient level should represent a college preparatory minimum.

According to the NAEP glossary, however, the Basic level of performance is "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade level."

While there is no quantitative information in this definition, the use of the term "partial mastery" surely signifies a level of performance below grade level. One could argue how many levels below, but we think at least one grade level below would accord with the concept of "partial mastery."

Thus, there are two NAEP designations that conform to being at or above grade level: Proficient and Advanced. And in the remainder of this report, we shall use the term "proficient" to mean Proficient or Advanced (proficient or better).

Statewide proficiencies and inflation in 26 states

The Standard & Poor's organization manages the website www.SchoolDataDirect.Org where one can find considerable information about achievement test proficiencies in the American public education sector.[iii] Of particular prominence in their data archives are the results of state-administered achievement tests as well as the federal NAEP test results. Generally, the state-administered results are available for individual schools and for districts as well as statewide. In comparison, the NAEP results are only available statewide and nationwide.♣

♠ To help resolve the ongoing dispute as to whether NAEP proficient defines grade level performance, we shall assert that it is a measure of "rigorously" defined grade-level skills. So when we speak of grade level in this report we are referring to a level of mastery equal to or greater than NAEP proficiency. Also, when we think of the oft-mentioned concept of "21st century skills," we think that should suggest higher skill levels that might be equated with the higher standard of NAEP proficiency.

♣ The reason for this restriction relates to the fact that the NAEP tests are administered to sparsely located samples of students. The statistical sampling is very reliable to produce statewide and national results, but would be unreliable for any smaller subdivisions such as school districts or individual schools. A further restriction applies to the 12th grade test where the results are reported only for a nationwide sample.

Most states also use the "Proficient" label to indicate student skills. However, by comparing the proficiencies at the statewide level, we know that the state exams "pass," or deem proficient, higher fractions of their students as compared to the NAEP - and oftentimes more than twice the number. Such discrepancies can be confusing to stakeholders.

To give these descriptions some perspective, we show in Table 1 below some information about achievement test inflation in 25 states and the District of Columbia.[iv] [v] The proficiencies shown are the statewide averages of 4th grade and 8th grade results in the areas of reading and mathematics. The Table also shows the inflation factor by which these states "augment" their mathematics and reading testing results above NAEP-measured proficiencies.

State Assessed In 2007	State Exam: Average 4th - 8th Math Proficiency Percentage	NAEP Exam: Average 4th - 8th Math Proficiency Percentage	Average Math Inflation	State Exam: Average 4th - 8th Reading Proficiency Percentage	NAEP Exam: Average 4th - 8th Reading Proficiency Percentage	Average Reading Inflation	Overall Inflation
SC	30.5%	34.0%	0.90	33.5%	25.5%	1.31	1.11
MA	46.5%	54.5%	0.94	65.5%	46.0%	1.42	1.18
VT	60.5%	45.0%	1.34	68.5%	41.5%	1.65	1.50
NH	61.5%	45.0%	1.37	69.0%	39.0%	1.77	1.57
ME	56.0%	38.0%	1.47	66.0%	36.5%	1.81	1.64
NJ	76.5%	46.0%	1.66	77.5%	41.0%	1.89	1.78
PA	73.0%	42.5%	1.72	72.5%	38.0%	1.91	1.81
RI	50.5%	31.0%	1.63	61.0%	29.0%	2.10	1.87
CA	44.5%	27.0%	1.65	46.0%	22.0%	2.09	1.87
NY	69.5%	36.5%	1.90	62.5%	34.0%	1.84	1.87
FL	66.0%	33.5%	1.97	58.5%	31.0%	1.89	1.93
CT	81.0%	40.0%	2.03	73.5%	39.0%	1.88	1.95
IN	73.0%	40.5%	1.80	71.0%	32.0%	2.22	2.01
MD	75.5%	38.5%	1.96	77.5%	34.5%	2.25	2.10
WI	76.0%	42.0%	1.81	83.0%	34.5%	2.41	2.11
VA	78.0%	39.5%	1.97	83.0%	36.0%	2.31	2.14
DE	68.0%	35.5%	1.92	78.0%	32.5%	2.40	2.16
NM	38.0%	20.5%	1.85	55.5%	20.5%	2.71	2.28
MI	76.5%	33.0%	2.32	80.5%	30.0%	2.68	2.50
IL	83.5%	33.5%	2.49	78.0%	31.0%	2.52	2.50
TX	78.5%	37.5%	2.09	86.5%	29.0%	2.98	2.54
GA	80.5%	28.5%	2.82	87.0%	27.0%	3.22	3.02
OK	80.0%	27.0%	2.96	84.5%	26.5%	3.19	3.08
TN	88.0%	26.0%	3.38	89.5%	26.5%	3.38	3.38
MS	67.5%	17.5%	3.86	71.0%	18.0%	3.94	3.90
DC	53.5%	11.0%	4.86	44.0%	13.0%	3.38	4.12

Table 1. Showing in 26 jurisdictions how 4th and 8th grade state-reported proficiency percentages are inflated by comparing against the de facto national standard: The NAEP benchmark. Tennessee public school students' skills, indicated by NAEP math and reading proficiencies of 26.0% and 26.5% respectively, are significantly below the national averages of 29.5% and 31.5%. The table is ordered by the average inflation factor (last column), with the nation's highest inflation state shown last and the lowest inflation state shown first.

Evidently, South Carolina and Massachusetts are the only states that do not significantly inflate student proficiencies. After ranking all the states for which test results are available, we find that the median state inflates its proficiencies about 100%. For Tennessee it is more than 200% - making it one of the top three "inflators" nationally.[vi]

It is worthy of comment to note that higher inflation factors generally correspond to lower NAEP proficiencies. A more complete review of an expanded Table 1 showing all tested states shows a significant inverse correlation between NAEP measured proficiencies and the amount of state "introduced" inflation.

The analysis upon which our mappings are based assumes that there are two primary causal factors that help determine the level of inflation seen: The cut score and the difficulty of the examination content. Other practices commonly used in the administration of these testing regimes, such as special accommodations, can also affect the inflation observed.

Looking at proficiencies and inflation in Tennessee

The highly inflated proficiency percentages reported by the Tennessee Comprehensive Assessment Program (TCAP) effectively hide the poor performance in many Tennessee schools. As we make our NAEP estimates for the various public schools and districts in Tennessee, we find very large majorities of the students who are sub-proficient in terms of the NAEP criteria.

How we obtain NAEP-estimated proficiencies locally

Techniques of applied mathematics were used to generate the formulas we use to convert state-reported proficiencies to ones consistent with the NAEP. Many of the details are provided in Appendix A. Our preferred method is called the Piecewise ELQ mapping, where ELQ is an acronym for “ellipse-quartic,” and is suggestive of the fact that our mapping curves (or formulas) are pieced together from a combination of curve segments, some of which are ELQ curves while others are quadratic polynomials. An earlier method, the Simple ELQ mapping, while not used in our Tennessee school and district level estimates, is the foundation on which the newer method has been built.

Report preview

This report is structured as follows: In the next section, on results, we present graphical representations of the mappings and a tabulation of some selected results. There we also direct readers to the spreadsheets in **TN-NAEP-Estimates.xls**, which display in tabular format the NAEP proficiency estimates for math and reading for all Tennessee public schools and districts in grade levels 4, 8, and 12. Then, in the last part of the main report, we offer conclusions and discuss the applicability as well as some limitations of this analysis.

Three appendices provide additional information. Appendices A and B present some details of how the mappings used in primary schools and secondary schools, respectively, were generated. Appendix C provides a quantitative review of the kinds of errors encountered in this study.

Results for Tennessee Public Schools and Districts

In what follows, we present some of our results by showing graphs of the actual interpolation formulas used in analyzing Tennessee’s reported proficiency percentages. Specifically we study grade levels 4, 8, and 12, which are among the ones tested in Tennessee. Given the extreme inflation of Tennessee’s TCAP-reported proficiencies, all of the mapping curves shown below in Figs. 1a – 3b have quartic terms (Lamé curve segments) in their formulas.

The maps for grades 4, 8, and 12

Tennessee’s achievement test results from the Tennessee Comprehensive Assessment Program (TCAP) are available from the previously mentioned website www.SchoolDataDirect.Org and from the website[vii] of the Tennessee Department of Public Education (TDOE). To resolve a problem with the acquisition of this data from TDOE sources we decided to take the data from SchoolDataDirect.[♣] All of our NAEP data is from NAEP’s website.[viii]

♣ We had earlier received spreadsheet files of TCAP proficiencies directly from TDOE. The proficiency numbers on the TDOE website and those in the files sent to us, when compared, had many small discrepancies - typically of less than one-percent. We sought guidance from TDOE regarding which set was “official” but as of this writing have had no response.

To make plausible our mapping estimation procedure, in the next few sections we show the various graphical representations of the mapping formulas actually used with respect to the 4th, 8th, and 12th grade assessment regimes in Tennessee. In each case, the NAEP scale estimate is found by first finding the TCAP proficiency percentage on the horizontal coordinate and then using that position to "read off" from the curve the vertical measurement or NAEP scale proficiency on the left-hand scale. Our actual NAEP estimates are evaluated from the corresponding formulas.

The 4th and 8th grade maps

We begin by displaying the 4th grade interpolation, or mapping functions.

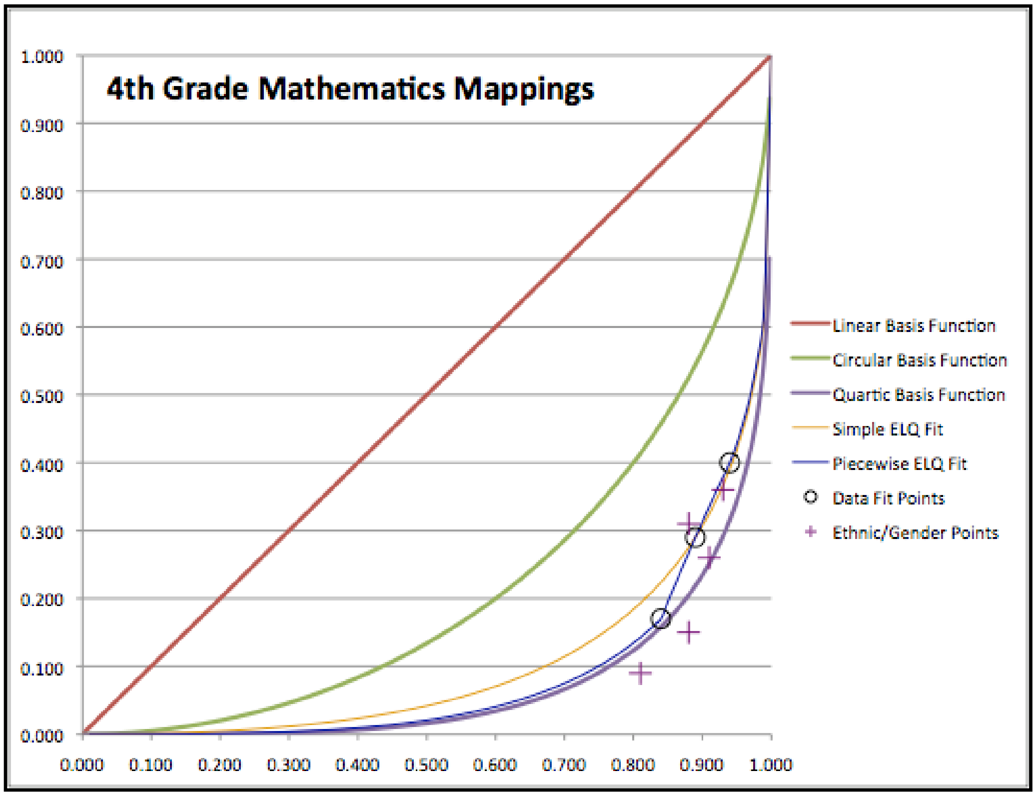

Figure 1a. Map for Tennessee's 4th grade public schools shows the R(S) mapping curves for mathematics. S is the horizontal coordinate and R the vertical. The orange and blue curves show the Simple ELQ and Piecewise ELQ mappings, respectively. The data for demographic groups are shown by the + and o symbols where the latter show the demographic data pairs we fit. The vertical displacements of the + symbols from the curves are measures of the errors in this analysis. See Appendix C for a more detailed analysis of the errors. The data pairs indicated by the + symbols, going left to right, pertain to the statewide proficiencies of Blacks, Hispanics, Males, Females, and Whites (B, H, M, F, W), respectively. And those indicated by the o symbols, proceeding left to right, correspond to the statewide proficiencies of Disadvantaged students, All students, and Advantaged students, respectively.

Likewise, we show the mapping curves for 4th grade reading skills.

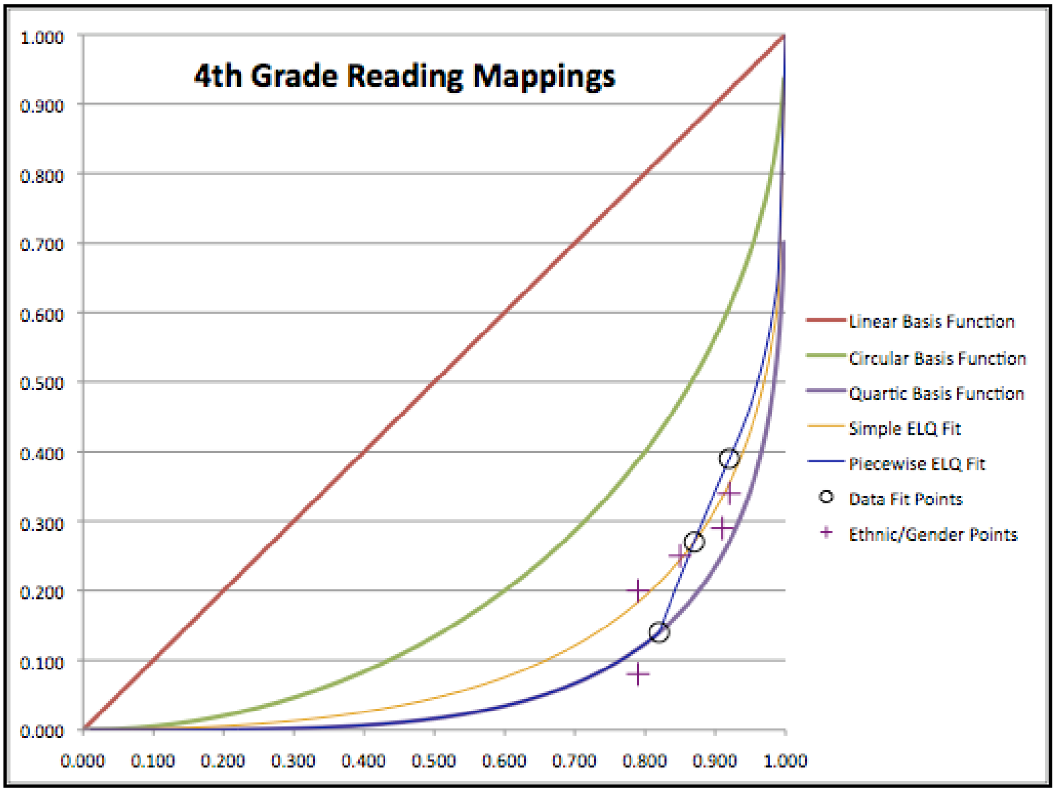

Figure 1b. Map for Tennessee's 4th grade public school students shows the mapping curves R(S) for reading. S is the horizontal coordinate, R the vertical. The orange and blue curves show the Simple ELQ and Piecewise ELQ mappings, respectively. Here the + symbols, going left to right, pertain to (B=H, M, F, W) where the "=" sign indicates a tie.

The very high level of inflation that we saw for Tennessee in Table 1 is evident in these two figures. Graphically, the inflation is the ratio of the height of the red curve over that of the data point (= *S/R*). The overall statewide average inflation is that definition applied to the data point shown by the middle circle marker. Likewise, the statewide inflation for disadvantaged students uses the circle marker to the left. Table 2 shows the relevant inflation factors pertaining to both 4th and 8th grade Tennessee public schools.

	Inflation for 4th Grade Math	Inflation for 4th Grade Reading	Inflation for 8th Grade Math	Inflation for 8th Grade Reading
All	3.07	3.22	3.78	3.54
Disadvantaged	4.94	5.86	6.83	6.29
Advantaged	2.35	2.36	2.88	2.74

Table 2. The TCAP inflation factors are shown for three demographics of the statewide population of students. Overall, the TCAP results indicate that at least three times as many children are proficient as is the case according to the NAEP. For disadvantaged students, the inflation ratios are much larger.

The preceding figures also show the basis function curves: a straight line, a circle, and a quartic- also known as the Lamé curve. Given the very high inflation, characterized by the data points being near the quartic curve, leads us to employ a mapping curve that combines elements of the quartic and circular basis functions, but not the linear one.

The orange and blue interpolation, or mapping curves, shown in Figs. 1a and 1b correspond to the Simple ELQ mapping and the Piecewise Continuous ELQ mapping respectively. The simple mapping fits the middle data point (indicated by the circular marker), which is the proficiency percentage of all 4th grade students statewide. The Piecewise Continuous ELQ mapping fits all three data points marked by the circular markers. Appendix A provides considerably more information on the mapping methods and their justification. The actual derivations are contained in a separate report available from the author's website.[ix]

The ethnic and gender data pairs indicated by the + symbol give us a means of measuring the errors of the mapping methods. The vertical distance between these markers and the mapping curve measures the error.

The 8th grade maps

As is evident below, the Tennessee maps in 8th grade are qualitatively similar to the ones we showed for 4th grade. What is different is the concavity of each mapping curve being higher than in the preceding figures.

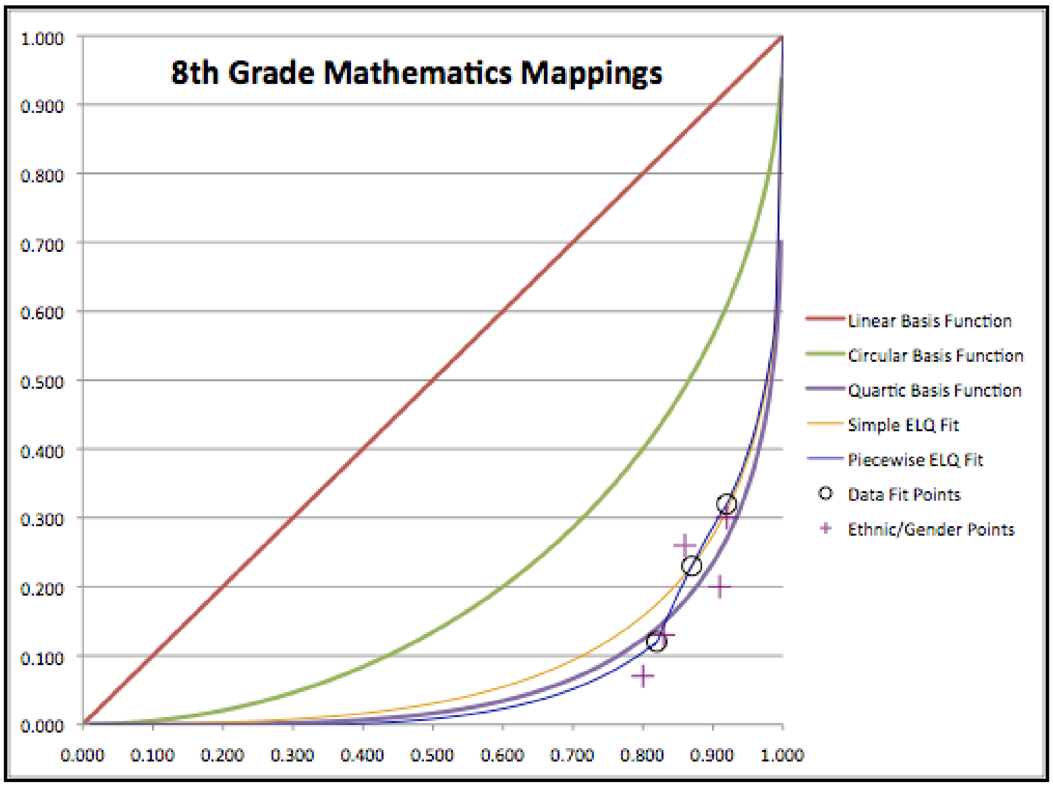

Figure 2a. Map for Tennessee's 8th grade public school students shows R(S) for the Simple ELQ mapping (orange curve) and the Piecewise ELQ mapping curve (blue) for mathematics. S is the horizontal coordinate and R the vertical. The descriptions of the markers given above also pertain here. Like the situations found in the 4th grade testing, here there is also a very large amount of inflation (as Table 1 attests). Here the + symbols, going left to right, pertain (B, H, M, F, W), respectively, which perhaps not coincidentally is the same ordering seen for 4th grade math in Figure 1a.

The mappings for 8th grade reading proficiencies are next, in Figure 2b.

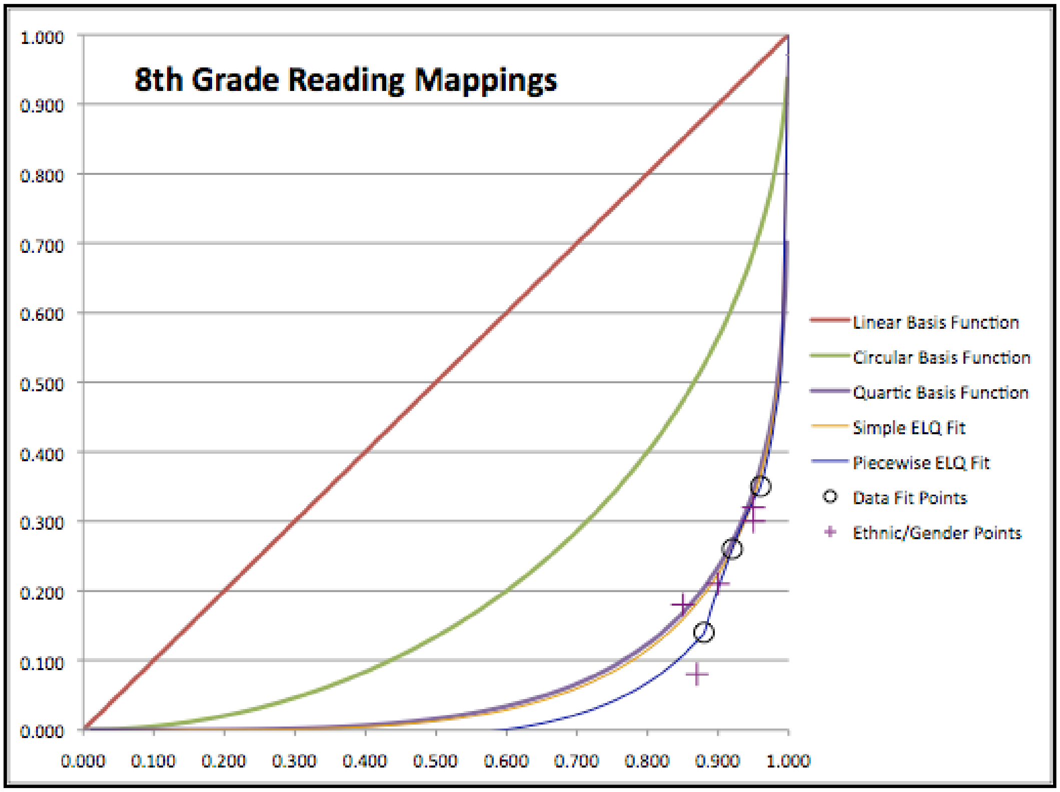

Figure 2b. Map for Tennessee's 8th grade public school students shows R(S) for the Simple ELQ mapping (orange curve) and the Piecewise ELQ mapping (blue curve) for reading. As before, S is the horizontal coordinate and R the vertical. Here the + symbols, going left to right, pertain (H, B, M, F=W), respectively, which is similar to the ordering for 4th grade reading shown in Figure 1b.

As will be shown more definitively in some of the tabular results presented farther along, Tennessee has more inflation in its 8th grade testing than at the 4th grade level. This has the effect of muting or even reversing the 8th grade proficiency drops that are seen by the NAEP. For example, Tennessee reports significantly higher reading proficiencies in 8th grade than in 4th grade when, in fact, the NAEP shows them approximately the same (within statistical errors). Likewise for math, where the NAEP sees a large drop in math proficiencies, the Tennessee results, inconsistently, show only a small decline.

The 12th grade maps

To obtain mapping formulas for the conversion of TCAP 12th grade proficiencies to the NAEP scale requires a more complicated analysis than we used in the lower grade levels. Please see Appendix B for details.

While the 12th grade mapping formulas appear reasonable (shown on the next page as the orange and blue curves), we regard them as less accurate than those generated at the 4th and 8th grades. Given that the 12th grade analysis requires additional assumptions and approximations, our confidence in the NAEP scale estimates is less certain than in the estimates we provide for 4th and 8th grade students.

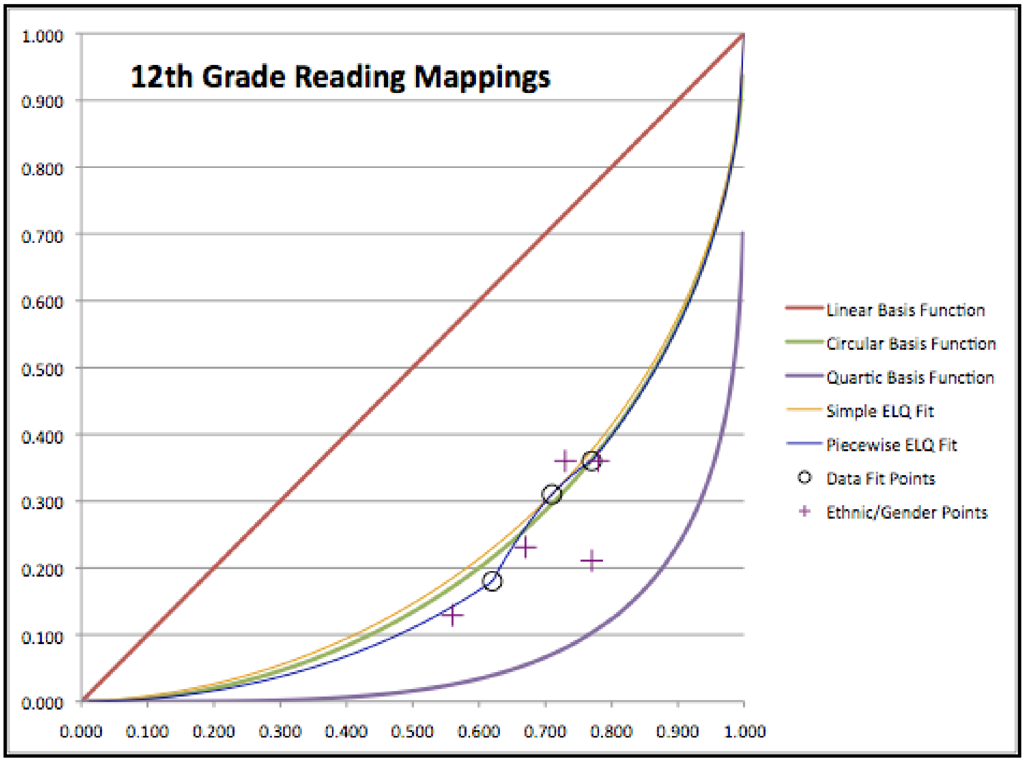

Figure 3a. The R(S) map for Tennessee 12th grade mathematics NAEP proficiency percentages is the result of two approximations. The first extrapolates, by means of a least square error analysis, 4th and 8th grade statewide NAEP proficiencies to obtain 12th grade approximate values of R, shown by the o and + markers' vertical coordinates. The three o markers, left to right, correspond to the demographic groups: Disadvantaged, All, and Advantaged. The ELQ formulas are used for both the Simple ELQ (orange) and Piecewise ELQ (blue) except for the mid-region blue, which is given by a quadratic polynomial fit. Here the + symbols, going left to right, pertain to (B, M, F, W, H), respectively.

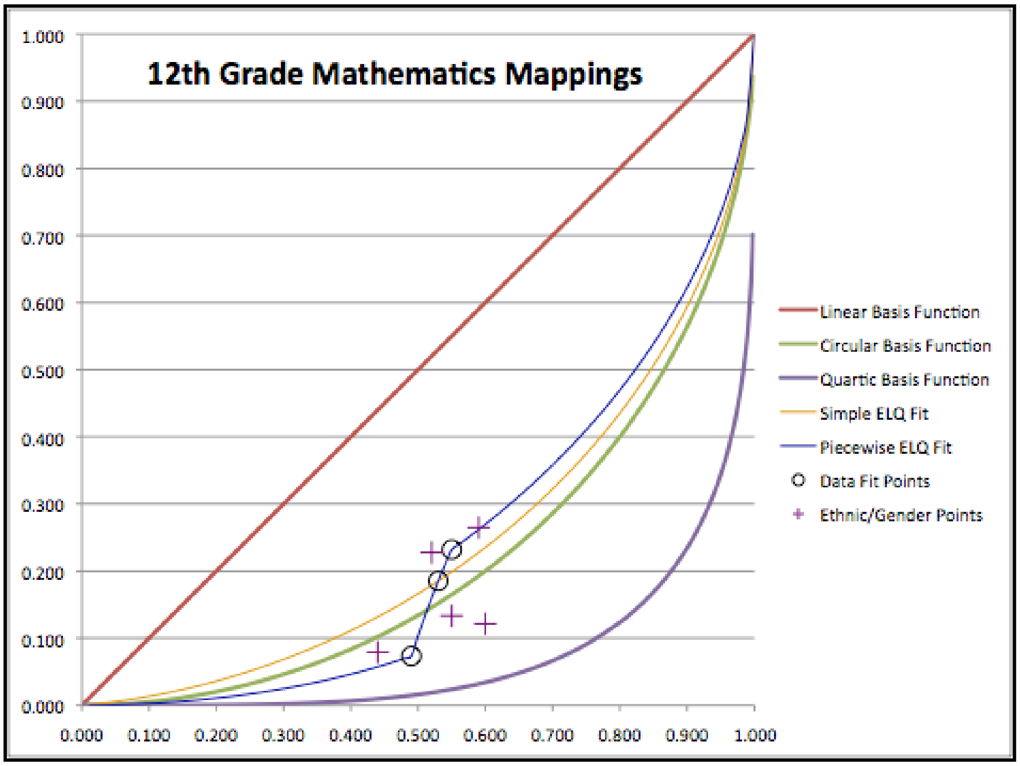

Figure 3b. The R(S) maps for Tennessee 12th grade NAEP reading proficiency percentages are shown. The nomenclature is much the same as in Fig. 3a. Here the + symbols, going left to right, pertain to (B, M, F, H, W), respectively. In this and the previous Figure, the Hispanic TCAP proficiencies approximate those of Whites, while the NAEP estimate proficiencies for Hispanic students are relatively much lower.

Measured errors for demographic groups

Given that the Piecewise ELQ mapping formula is generated by fitting the three demographics - All, Disadvantaged, and Advantaged - it follows that the errors for those groups will be zero. As indicated above, the ethnic and gender groups' errors are used to gauge the accuracy of the mapping formulas.

As a general rule (in 80% or more of the cases studied prior to our analysis for Tennessee), we have found these errors to be less than 5% for our 4th and 8th grade mappings. At the high school level, to be described later, the mapping approximations are more complex and errors have been found to be somewhat larger, but generally bound by 10%. The measured errors for 4th and 8th grades in Tennessee sometime violate this "general rule," which may be due to irregularities in the TCAP data.[♥] Nevertheless, we believe that the errors are still sufficiently small so as to not invalidate the general conclusions we shall reach. Appendix C provides more details concerning these errors.

The errors for the demographic groups are quantitatively the vertical displacements of the + marks in the Figures from the blue Piecewise ELQ curve. Most of the large errors - those exceeding 10% - affect the Hispanic students' proficiencies. In 12th grade their proficiency errors are 14.8% and 14.9% for math and reading, respectively, while in 4th grade their math proficiency error is 12%. The only other error exceeding 10% concerns 8th grade Female math proficiencies where the error is 10.4%.

[♥] The largest error at the 4th and 8th grade levels is for 4th grade Hispanic math proficiencies, which is 12% off the mapping curve - evident in Figure 1a.

As we previously mentioned, our 4th and 8th grade NAEP estimates are usually accurate within 5% on the proficiency scale, while the error bounds at the 12th grade are more in the range of 10%. The theory behind our mapping methods also suggests such error levels. In Tennessee, however, there may be additional sources of errors, which can help explain the quite large horizontal displacements of the Hispanic math and reading data pairs seen in Figs. 3a and 3b above.

What we do find interesting about Figures 1a – 3b is the visual evidence that both ELQ interpolation methods provide sufficient accuracy for most of the various demographic groups, subjects and grade levels. The one remarkable exception being the 12th grade TCAP proficiencies reported for Hispanic students, which are improbably given as nearly equal to the White demographic group- quite unlike the NAEP measurements for these two groups. More information on these errors is presented in Appendix C.

Tennessee's TCAP proficiencies are significantly inflated

Table 3, below, extends the information presented in Table 2 to show the inflation factors for the 12th grade analysis.

	Inflation for 12th Grade Math	Inflation for 12th Grade Reading
All	2.86	2.29
Disadvantaged	6.68	3.44
Advantaged	2.37	2.14

Table 3 simply shows the ratio of the TCAP-reported proficiency to that of the NAEP-reported value for the three demographic groups we used for data fitting. As in Table 2, the inflation factor for the Disadvantaged 12th grade mathematics proficiency is remarkably high, while that for reading is considerably lower than the high levels seen for the 4th and 8th grades.

The high concavity of the mapping curves shown in the preceding figures, as well as the inflation factors seen in the accompanying tables, confirms that the TCAP testing regime grossly exaggerates the skills of Tennessee's public school students. Nearly all states report inflated proficiencies but few are as exaggerated as those seen in Tennessee. In contrast, states such as South Carolina and Massachusetts make an effort to align their testing standards with the NAEP. In doing so they obtain mapping curves that more closely coincide with the straight $R=S$ line (plotted in red in the foregoing figures).

Using the maps

Why we use the Piecewise Continuous ELQ Mapping. In the preceding figures, each showed the mapping curves of the Simple ELQ formula and that of the Piecewise Continuous ELQ formula. The latter formula, in studies conducted in a simulated assessment environment and in other studies conducted with data pairs from 26 different jurisdictions, is almost always more accurate than the Simple version when applied to 4th and 8th grade data. The newer Piecewise formula has not been applied at the high school level previously. We have compared both formulas using data and estimates for Tennessee 12th graders' skills and find again that the Piecewise version produces smaller errors with the exception of the Hispanic proficiencies, whose large errors are almost certainly driven by "irregularities" in the testing process. Appendix C provides further information about error sources such as this.

How we apply the Piecewise Continuous ELQ Mapping. The process is simple: We enter the TCAP proficiency into the Piecewise ELQ formula which then produces the estimated NAEP result. It yields an estimate of what percentage of the children in that school or district would have been found proficient on the NAEP.

Tabular results: NAEP scale estimates in Tennessee

Before presenting the results, it is important to reiterate that they depend on a number of assumptions. We know that four of our assumptions, of the five listed in Appendix A, are valid for the Tennessee data and they appear to be applicable to many other states.[♣]

The Piecewise ELQ method obviously produces some level of error when it is applied, but we find these are small compared to the errors of inflation that it removes. Elsewhere we have demonstrated its advantages and small errors in the report[x] that presents its derivation. Additionally, we have tested its accuracy against the known results of the various demographic groups. This gives us more confidence in the following results.

School- and District-level results

The large volume of results precludes presenting them in the main body of this report. Instead we have presented them in a spreadsheet workbook, **TN-NAEP-Estimates.xls**, which accompanies this report and/or is available from the author.

The workbook consists of six worksheets as follows:

- **TN Districts 4&8 Alpha** presents the 4th and 8th grade results of the school districts in alphabetical order.
- **TN Districts 4&8 Ordered** presents the 4th and 8th grade results of the school districts in proficiency order.
- **TN Schools 4&8 Alpha** presents the 4th and 8th grade results of Tennessee schools in alphabetical order.
- **TN Schools 4&8 Ordered** presents the 4th and 8th grade results of Tennessee schools in proficiency order.
- **TN High Schools Alpha** presents the 12th grade results for Tennessee high schools & districts in alphabetical order.
- **TN High Schools Ordered** presents the 12th grade results for Tennessee high schools & districts in proficiency order.

To provide a glimpse of the information and formats found in the spreadsheets, we next display some small portions of three of them corresponding to the least proficient and most proficient schools in each of the three tested grade levels.

Arrangement and description of the spreadsheets

The spreadsheets are fairly well labeled and need little further description. However, it may be helpful to elucidate the top section and the last column as they are particularly important.

The top eight or nine rows of each spreadsheet pertain to the testing of the statewide populations of students according to seven or eight demographic groups and the total population of tested pupils.

The first row, labeled "statewide," contains the two data points that the Piecewise ELQ formula "fits" for each subject. Thus there are four in all (usually shown in pink). There is the (S_0, R_0) pair for mathematics and the (S_0, R_0) pair for reading (or English language arts). For each subject, the (S_0, R_0) pair determines the Piecewise ELQ interpolation curve, which is used to estimate R-values for each S value given. The (S_0, R_0) pair is marked on the figures by the circular icon. Finally, in the last three blue shaded columns of the first

♣ It has been applied to estimate math and reading proficiencies (on the NAEP scale) for all public schools and districts in Bristol County in Massachusetts, Ventura County in California, Hudson County in New Jersey, and in the states of Oklahoma, New Hampshire, New Mexico, Rhode Island, Pennsylvania and Vermont.

row, we compute the Piecewise ELQ estimates for the given (S_0, R_0) pairs. Since each Piecewise ELQ curve is, by design, constructed to fit the data points, you'll notice that the entries in the first two of the last three columns simply replicate the data values given to the left. We shall describe the last column shortly.

All of the remaining rows of the top eight or nine pertain to the seven or eight demographic groups for which proficiencies are available from NAEP and from the SBA assessments. In each of these cases, the Piecewise ELQ formula predicts a NAEP scale proficiency as shown in the first two blue shaded columns. Since NAEP has separately reported these demographic group proficiencies, this gives us a means of calculating the errors in the Piecewise ELQ formula for these six tested groups.

In the second part of the table, below the top eight or nine rows, we present the school-by-school or district-by-district information. Again, the first two blue columns contain our NAEP estimates for math and reading respectively. The yellow shaded cells are marked NA. This relates to the fact that the NAEP results are not available for individual schools or districts. It is this deficit of information that motivated the ELQ analysis, which provides estimates for these missing columns in the first two blue shaded columns.

The last blue column contains what we designate as the *overall* NAEP scale proficiency. It is simply the minimum of the two preceding columns. In our terminology, a student is accorded the label *overall* proficient only if he or she is proficient in both math and reading. When looking at the proficiency percentages within any tested group (school or district) it is clear that the overall proficiency, at best, will be the minimum of the math and reading proficiencies for that group.

We provide the information in two different arrangements:

- Ordered alphabetically **-** to facilitate the lookup of schools and districts of interest.
- Ordered by overall proficiency **-** to see where tested groups stand among others.

While the bulk of the results are in the spreadsheets, we present here results for some of the most and least proficient tested groups. Accordingly, in the next three subsections, six "snapshots" from the Excel workbook **TN-NAEP-Estimates.xlsx** are shown.

Appearance & configuration of the tables for 4th, 8th and 12th grades

To help readers understand the arrangement of the tables presented in the associated spreadsheets, we here display small excerpts of the tables for the three grade levels in question. The segments presented here in each grade level show the median schools with respect to being ordered by the NAEP estimated overall proficiency, which is shown in the rightmost column of these tables.

Number	District Number	Site Number	District Name	School or Tested Group	State Test: 4th Math Proficiency Percentage	NAEP Test: Statewide 4th Math Proficiency Percentage	State Test: 4th Reading Proficiency Percentage	NAEP Test: Statewide 4th Reading Proficiency Percentage	NAEP Consistent ELQ 4th Math Inflation PP	NAEP Consistent ELQ 4th Reading Inflation PP	NAEP 4th Grade Consistent ELQ Proficient in Both
				Year of testing=	2007	2007	2007	2007			
1	NA	NA	TN	StateWide All	89.0%	29.0%	87.0%	27.0%	29.0%	27.0%	27.0%
2	NA	NA	TN	Disadvantaged	84.0%	17.0%	82.0%	14.0%	17.0%	14.0%	14.0%
3	NA	NA	TN	Advantaged	94.0%	40.0%	92.0%	39.0%	40.0%	39.0%	39.0%
4	NA	NA	TN	White Demographic Group	93.0%	36.0%	92.0%	34.0%	37.9%	39.0%	37.9%
5	NA	NA	TN	Black Demographic Group	81.0%	9.0%	79.0%	8.0%	14.2%	11.6%	11.6%
6	NA	NA	TN	Hispanic Demographic Group	88.0%	15.0%	79.0%	20.0%	26.7%	11.6%	11.6%
7	NA	NA	TN	Asian	95.0%		94.0%		42.8%	43.7%	42.8%
8	NA	NA	TN	Female	91.0%	26.0%	91.0%	29.0%	33.5%	36.7%	33.5%
9	NA	NA	TN	Male	88.0%	31.0%	85.0%	25.0%	26.7%	21.9%	21.9%
485	570	56	MADISON COUN	Lincoln Magnet Elementary	91.3%	NA	86.8%	NA	33.5%	27.0%	27.0%
486	470	42	KNOX COUNTY	Cedar Bluff Intermediate	91.9%	NA	87.4%	NA	35.7%	27.0%	27.0%
487	750	20	RUTHERFORD C	Christiana Elementary	91.9%	NA	86.7%	NA	35.7%	27.0%	27.0%
488	470	260	KNOX COUNTY	Shannondale Elementary	93.7%	NA	87.4%	NA	40.0%	27.0%	27.0%
489	710	50	PUTNAM COUNT	Jere Whitson Elementary	93.7%	NA	87.0%	NA	40.0%	27.0%	27.0%

Table 4. We show the top header segment and a small portion from the middle of the spreadsheet table [TN Schools 4&8 Ordered] that displays Tennessee 4th grade proficiency percentages as reported from the TCAP tests and as estimated for the NAEP. The schools are presented in the order of their overall estimated NAEP proficiencies (of the rightmost column). Here we show just a few rows adjoining those for the school with median performance. Pink and blue-backgrounded cells indicate input data and results from the application of the Piecewise ELQ formulas, respectively. The cells labeled "NA" have no data because actual NAEP proficiencies are not reported locally. The top nine rows display statewide data and estimates for nine demographic groups. Of those, the top three rows' estimates agree precisely because they represent the data "fit" points.

The tables for 8th grade testing retain the same arrangements and formats as seen now in Table 5:

Number	District Number	SitNo	District Name	School or Tested Group	State Test: 8th Math Proficiency Percentage	NAEP Test: Statewide 8th Math Proficiency Percentage	State Test: 8th Reading Proficiency Percentage	NAEP Test: Statewide 8th Reading Proficiency Percentage	Consistent ELQ 8th Math Inflation PP	Consistent ELQ 8th Reading Inflation PP	NAEP Consistent ELQ 8th Proficient in Both
				Year of testing=	2007	2007	2007	2007			
1	NA	NA	TN	StateWide All	87.0%	23.0%	92.0%	26.0%	23.0%	26.0%	23.0%
2	NA	NA	TN	Disadvantaged	82.0%	12.0%	88.0%	14.0%	12.0%	14.0%	12.0%
3	NA	NA	TN	Advantaged	92.0%	32.0%	96.0%	35.0%	32.0%	35.0%	32.0%
4	NA	NA	TN	White Demographic Group	92.0%	30.0%	95.0%	32.0%	32.0%	33.0%	32.0%
5	NA	NA	TN	Black Demographic Group	80.0%	7.0%	87.0%	8.0%	10.5%	12.8%	10.5%
6	NA	NA	TN	Hispanic Demographic Group	83.0%	13.0%	85.0%	18.0%	14.4%	10.7%	10.7%
7	NA	NA	TN	USA Indian Demographic Group	95.0%		95.0%		39.4%	33.0%	33.0%
8	NA	NA	TN	Female	91.0%	20.0%	95.0%	30.0%	30.4%	33.0%	30.4%
9	NA	NA	TN	Male	86.0%	26.0%	90.0%	21.0%	21.0%	20.4%	20.4%
275	330	157	HAMILTON COUNTY	Ooltewah Middle School	91.7%	NA	91.4%	NA	32.0%	23.3%	23.3%
276	670	35	OVERTON COUNTY	Livingston Middle School	92.3%	NA	91.1%	NA	32.0%	23.3%	23.3%
277	750	75	RUTHERFORD COU	Rockvale Elementary	92.6%	NA	90.6%	NA	34.2%	23.3%	23.3%
278	330	165	Hamilton County	LOOKOUT VALL MI	93.5%	NA	90.6%	NA	34.2%	23.3%	23.3%
279	400	12	HENRY COUNTY	Dorothy And Noble Harrelson School	94.2%	NA	90.7%	NA	36.6%	23.3%	23.3%

Table 5. Here are shown some representative rows from the 8th grade proficiency tables. As in the preceding table for 4th grade, we show a handful of rows surrounding the results for the median school.

The Tables for the 12th grade testing results are similar to the foregoing except that the NAEP proficiencies shown in the header are themselves the result of approximations that are further described in Appendix B.

					Source	Source	Source	Source			
					TCAP	NAEP TN	TCAP	NAEP TN			
				Year of testing >	2007	2007	2007	2007	PP designates proficiency %		
Number	District Number	School Number	District Name Or Region	School or Tested Group	State Test: 12th Grade Math Proficiency Percentage	NAEP Test: Estimate 12th Grade Math Proficiency Percentage	State Test: 12th Grade Reading Proficiency Percentage	NAEP Test: Estimated 12th Grade Reading Proficiency Percentage	NAEP Consistent ELQ 12th Grade Math Inflation PP	NAEP Consistent ELQ 12th Grade Reading Inflation PP	NAEP Consistent Proficent in Both
1	NA	NA	TN	StateWide	53.0%	18.5%	71.0%	31.0%	18.5%	31.0%	18.5%
2	NA	NA	TN	Disadvantaged	49.0%	7.3%	62.0%	18.0%	7.3%	18.0%	7.3%
3	NA	NA	TN	Advantaged	55.0%	23.2%	77.0%	36.0%	23.2%	36.0%	23.2%
4	NA	NA	TN	White Demographic Group	59.0%	26.5%	78.0%	36.0%	26.2%	37.2%	26.2%
5	NA	NA	TN	Black Demographic Group	44.0%	7.9%	56.0%	12.9%	5.7%	14.2%	5.7%
6	NA	NA	TN	Hispanic Demographic Group	60.0%	12.1%	77.0%	21.1%	27.0%	36.0%	27.0%
7	NA	NA	TN	Female Demographic Group	55.0%	13.3%	73.0%	36.0%	23.2%	33.0%	23.2%
8	NA	NA	TN	Male Demographic Group	52.0%	22.8%	67.0%	23.1%	16.0%	26.0%	16.0%
121	791	119	Memphis City	Cordova High School	52.3%	NA	76.7%	NA	16.0%	36.0%	16.0%
122	370	17	Hawkins County	Cherokee High School	52.4%	NA	84.6%	NA	16.0%	47.1%	16.0%
123	821	45	Bristol	Tennessee High School	52.5%	NA	85.5%	NA	16.0%	47.1%	16.0%
124	60	20	Bradley County	Bradley Central High School	80.1%	NA	61.2%	NA	47.0%	17.3%	17.3%
125	60	78	Bradley County	Walker Valley High School	52.6%	NA	79.0%	NA	18.5%	38.5%	18.5%

Table 6. The tables reporting 12th grade NAEP estimates are similar to the two previous tables for 4th and 8th grades except the "data" in the header showing the NAEP proficiencies are themselves approximations. We indicate this by showing those numbers against a blue background in the two columns labeled "source." The three other blue backgrounded columns to the right show the values calculated by the Piecewise ELQ formula.

What do these estimates say about student proficiency and testing regimes?

The results in terms of NAEP scale proficiency percentages are shown against the blue backgrounds in the several tables in the accompanying spreadsheets. Our discussion will focus on those students who are proficient in both mathematics and in reading - the latter sometimes called English language arts. Thus, for a student to be accorded the status of proficient, we hold that he or she must be proficient in both reading and mathematics. Likewise, a district or school's proficiency percentage of most importance in our analysis is the percentage of students who are proficient in both, which we optimistically assume to be the minimum of the reading and mathematics proficiency percentages.[♠] That number is shown in the rightmost column and is also the number used to establish their ranking and order of display in the table. While a more detailed analysis might focus separately on the proficiencies for mathematics or reading, our focus is on the overall proficiencies shown in the last column.

We first look at the national situation, but not because it represents a desirable goal. We look at it to understand that the problems seen in Tennessee are also seen nationally. One thing evident from the national NAEP proficiency percentages is their decline over the school careers of the tested cohorts. Table 7 compares the national trend and that seen in Tennessee.

♠ The maximum percentage of students who could possibly be proficient in both subjects is the minimum of the two proficiency percentages for math and reading. At the other extreme, it is possible that the number of students who are proficient in both areas could be as small as the sum of the two proficiency percentages minus 100%. This means that it is possible - though quite unlikely - that until the sum of the two proficiencies exceeds 100% one could have 0% proficient in the combination. Thus we are making a very optimistic assumption in choosing the maximum within this range, and therefore our reported proficiency estimates are probably overstating the actual situations. Thus the results shown in our tables, disappointing as they are, are probably still exaggerating the true levels of student competence.

Grade Level	National NAEP Proficiencies	Tennessee NAEP Proficiencies	Proficiency Gap
4th	32%	27%	5%
8th	29%	23%	6%
12th	22%	19%	3%

Table 7. Here we see that Tennessee's NAEP proficiencies lag those measured nationally by a relatively modest though significant gap. This proficiency "gap" generally widens as one looks up through the grade levels, but in Tennessee it is seen to narrow after grade 8.

We caution readers that the results we present depend on the input data. The reliability of our results depends on the consistency and accuracy of the data from the TCAP examinations. Beyond those considerations, the utility of our results also depends on the various approximations made in developing the ELQ mapping methodologies.

Summary and Conclusions

What we have done

Our goal in developing our methods for mapping TCAP reported student proficiencies onto the NAEP scale has been to provide stakeholders in Tennessee public schools a better metric for understanding student skills and particularly to understand how they match up to national standards. The Piecewise Continuous ELQ formulas we derived have given us reasonably accurate quantitative estimates of children's grade level skills within every public school in Tennessee that reported TCAP numbers. We did this for students in the 4th, 8th and 12th grades. The estimates we provide have successfully removed most of the "inflation" that characterized the proficiencies reported by the TCAP and give interested readers a more realistic accounting of where each school stands with respect to student skills in the fundamentally important subject areas of mathematics and English language arts.

A note on private schools

We have not addressed private schools in this report. Nationwide, NAEP reports approximately 45% of such students to be proficient. This suggests that while better than most public schools, America's private schools also suffer from having a majority of their students below grade level and from having far too many graduates lacking 12th grade skills.

Background factors

Even without the NAEP estimates that are presented in this report, the statewide data for Tennessee and the other states provides context and a backdrop of information that tells part of the story about the performance of Tennessee public schools. For example:

- By comparing the NAEP scores and the various similar testing regimes used by the states, we learn that most states "inflate" the children's performance levels wherein they place many more children in the proficient or above category than really merit that designation. In the case of Tennessee, this inflation averages 215%, 266%, and 158% above NAEP levels, respectively, in the 4th, 8th, and 12th grades. (Corresponding to inflation factors of 3.15, 3.66, and 2.58, respectively.) Over those three grade levels the average inflation factor is 3.13 (213% above NAEP levels).

- Though not discussed in the foregoing, the SchoolDataDirect website[xi] also presents information about the graduation rates from which we can infer substantial dropout rates - about 30% nationally. It shows Tennessee's dropout rate is approximately the same, estimated to be 31%.

High School Success Rate
Finally, when we define what we call the *high school success rate* to be the fraction of those entering 9th grade who graduate from 12th grade and who also attain or exceed NAEP proficiency, the statistics available from SchoolDataDirect and the NAEP allow one to estimate a nationwide *high school success rate* of approximately 15%. The same analysis for Tennessee suggests a 13.1% *high school success rate* among its public high school students who had entered the 9th grade.

A clearer picture of Tennessee public schools emerges

Removing the inflation yields more realistic proficiencies
Two basic conclusions of this report are:

- Looking at the statewide situation we have reviewed the concrete evidence showing the substantial inflation in the TCAP-reported proficiencies which, averaged over 4th, 8th and 12th grades, is about 213% above the more reliable NAEP-reported proficiency percentages. Tennessee is almost the nation's largest inflator.
- Through the use of the Piecewise ELQ mapping procedures, we have made estimates at the school and district levels showing that even the best schools often have unexpectedly low levels of proficiency, while the worst schools are performing at levels even lower than earlier presumed.

From the NAEP estimates alone, we learn that most schools in the state have a majority of their children performing below grade level. The percentages of schools in Tennessee where more than 30% of the children perform at or above grade level are:

- 46% of schools in 4th grade,
- 31% of schools in 8th grade, and
- 31% of schools in 12th grade.

Concern about Hispanic and Black Proficiencies in 12th grade
Quite evident in the tables and figures are the extraordinarily high Hispanic proficiencies reported by Tennessee's TCAP test for 12th grade students. Black student TCAP proficiencies are also much higher than would be expected. The TCAP says the Hispanic students, statewide, have a one-point proficiency edge over White students in 12th grade mathematics (60% versus 59%). In 12th grade reading, they are a point behind Whites (77% versus 78%). The reported (not estimated) statewide NAEP proficiencies in Tennessee suggest that four times as many White students are math proficient compared to Hispanics and about twice as many in reading. Given the reputation of the NAEP and its reputation as a benchmark, this suggests that there are one or more irregularities in the TCAP testing system that favors Black, Hispanic and Disadvantaged students in some way. Our first hypothesis would be to consider *special accommodations* as a potential causal factor. Appendix C provides more information about this issue.

Disproportionate inflation for disadvantaged students
While inflation factors for the overall student populations taking the various tests averaged out to approximately a 200% increase over NAEP levels, these factors are much larger for the Disadvantaged students, as is made clear in Tables 2 and 3. We see, on average, a 467% increase. When the amount of inflation varies in this way, it adds to the difficulties of determining how many children are sub-proficient.

Applicability of the piecewise continuous ELQ mapping method

There are certain limitations to this analysis, including:

- The ELQ mapping method is based on five assumptions (detailed in Appendix A) that are plausible but not rigorously proven. The Tennessee data violates the ordering assumption in the area of math testing.
- Measured and estimated errors in our previous work in other states verified our claim that our estimated proficiencies at the 4th and 8th grade levels are generally accurate to within 5% on the proficiency percentage scale. In 12th grade our earlier work rarely found errors exceeding 10%.
- However, the situation in Tennessee is different. The state-reported proficiencies display a variety of inconsistencies that in every grade level and subject violate the ordering assumption (Assumption 2 in Appendix A). The proficiency estimates are now significantly less accurate with error bars bounded by roughly 12% in 4th and 8th grades and 15% in 12th (both regard Hispanic proficiency estimates). We refer interested readers to the error discussions in Appendix C. Even these relatively large error magnitudes are still considerably smaller than the errors associated with the inflation. Thus we believe that the NAEP estimates produced by the mappings are sufficiently accurate for the purposes of this report.
- The extrapolation of statewide NAEP proficiencies to the 12th grade level (discussed in Appendix B) assumes a linear relationship that is also unproven.

Better testing is foundational to instructional and other reforms

The focus of this report is that of getting a realistic picture of the performance levels (proficiencies) of Tennessee public schools and districts. It is not our goal to recommend remedies. However, within the field of assessments, we think it helpful to outline some possibilities that education reformers might consider.

A prerequisite to reform: Aligning assessments with national standards

When a state's assessment system produces exaggerated or inflated numbers of students who are deemed proficient, it is helpful to make stakeholders aware of how many children are really meeting national standards. Recognizing that institutional inertia will tend to retard or block the public school authorities from providing this useful information, there are a number of steps could be taken to move in that direction, including:

- Keep using the inflated proficiencies in reports, but notify readers of the inflation through footnotes.
- While waiting for the authorities to act, one can estimate local NAEP proficiencies, as we do in this report.
- Eliminate the inflation by adjusting the cut scores.[♥]
- Change the content tested to better align with the NAEP and also adjust cut scores as needed.
- Develop testing in which a subset of the test is completely aligned with the NAEP while a different subset tests content that goes beyond the NAEP standards. Disaggregate the NAEP scores.

These five suggested approaches to reform are ordered in terms of the ease of implementation, with the easiest first. The ordering also reflects which proposals will be most effective, with the least effective first.

Relevance to social promotion and special accommodations

The results presented here suggest large numbers of sub-proficient children within Tennessee public schools. Using our local NAEP estimates to "drill down" to public schools in various communities shows that substantial percentages of children within the best public schools are also sub-proficient. The nexus of a student being sub-proficient versus socially promoted is one worthy of further study.

[♥] This proposal and the two following it require new public policies that could be blocked by the political process. But in the meantime, stakeholders are free to implement the first two proposals on this list.

As discussed in Appendix C, the use of incompatible testing metrics, which is implicit in *special accommodations'* policies, presents a scientific conundrum. These special accommodations are afforded children in both the NAEP and the TCAP testing systems, but the criteria for them and the details of their implementation differ. We sense that the "upshifting" phenomenon (discussed in Appendix C) is related to *special accommodations*.

In terms of our mapping methods, it is quite possible that the older Simple ELQ method would have been adequate in the absence of the effects of *special accommodations*. But with them included we think the actual mapping relationship gets sufficiently distorted to the point where it is better represented by the higher order Piecewise Continuous ELQ formulation used in this study.

A word of encouragement

The results and conclusions from our NAEP-estimated proficiency percentages for Tennessee public schools are surely discouraging. However, Tennessee is not in this "boat" alone. The public schools in many other states have similar problems. These problems, namely of large numbers of sub-proficient children, also extend into many other advanced industrial nations around the world. If Tennessee public education leaders can "engineer" workable reforms to these problems and demonstrate their successes, the state can not only improve its own schools, but also it can be an example for others to follow.

Appendix A: Proficiency Mapping Methodologies

Context of the mappings

In the Introduction we alluded to the influence of lower cut scores and easier tests and their roles in producing the higher proficiency percentages seen in most states' achievement testing systems. Since we don't control the examination environments, we are limited in the remedies we can apply. The one we have applied and discussed in this report is that of mapping inflated state-reported proficiencies to ones on the NAEP scale.

Helping our analysis is the fact that in each state there are approximately nine demographic groups for which statewide examination data is available from both the NAEP and the state assessments. From those we have chosen up to three sets or demographics of assessed students to provide "fit" points for the derivation of our mapping formulas. The ones we use are: The overall group, those designated as economically disadvantaged, and those designated as economically advantaged.

As we noted above, state-reported proficiency percentages are typically double those reported by the NAEP. Even when the discrepancies are smaller they are still of concern. For example, while South Carolina and Massachusetts inflated their reading proficiencies, respectively, by about 26% and 46% in 2007, they were the only two states not inflating their math numbers. All other states significantly inflated their math and reading proficiencies above what was reported by the NAEP. According to Chester Finn, Tennessee inflates its students' proficiencies more than any other state.[xii] Which state inflates the most depends on the precise definition of the average inflation; by our estimation Tennessee is 3rd from the "top."

Statistical sampling limitations preclude making NAEP exam proficiency percentages available for individual schools or districts. However, we can estimate them by employing a mapping methodology that permits us to convert the state-reported proficiencies to the NAEP scale. In this Appendix we review two reasonably accurate mapping formulas $R(S)$ that convert the state reported proficiency percentages, S, for a tested group into more realistic proficiency percentages, R, which are aligned with the NAEP results - insofar as they are known.

Our preferred method is called the Piecewise ELQ mapping, where ELQ is an acronym for "ellipse-quartic" and is suggestive of the fact that our mapping curves (or formulas) are pieced together from a combination of curve segments, some of which are ellipses and quartic curves while others are quadratic polynomials. An earlier method, the Simple ELQ mapping, while not used in our school and district level estimates, is the foundation on which the newer method has been built.

Both mapping methods apply the Ellipse-Quartic (ELQ) formula, which we describe below. An important characteristic of the relationship of the state assessment system to that of the NAEP is the degree to which the state-reported proficiencies of the statewide tested population S_0 are "inflated" above the NAEP-reported proficiency R_0 of that same population. We define "statewide inflation" as the ratio S_0/R_0, which gives one a measure of the degree to which the statewide proficiencies have been exaggerated. However, it doesn't tell us what the local inflation factors are, nor what the local proficiency percentages are at the district and school levels.

These are proficiency mappings. They are not score mappings.

The available literature provides no examples of NAEP-equivalent district and school proficiencies mapped from state achievement test proficiencies. There are, however, a number of studies in which mapping methodologies are applied to scale scores rather than the proficiencies. The work by Jim Hull at the Center for Public Education,[xiii] work at the National Center for Educational Statistics,[xiv] and research by Gary Phillips at the American Institutes for Research[xv] are good examples of research employing score mappings.

Different kinds of statistical relationships can be used to establish links between the scoring distributions of the same tested group on different examinations. Of the types of linking discussed by Phillips, we believe that the Ellipse-Quartic (ELQ) mapping methods would most closely fall in his "projection" category where the linear regression relationship found in methods of this type now takes on an integrated form - though not explicitly in our analysis.

The specifics of achievement assessment methodologies are complex and rely on sophisticated statistical techniques, including those of psychometrics. The estimation procedures given in this report are based on a number of assumptions that do not require knowledge of those kinds of details. Whether reliance on these assumptions is justified depends on an analysis of the errors - an issue discussed below and in Appendix C.

The R_0 values of each state for reading and mathematics are specifically provided by NAEP for 4th and 8th grade students.[♠] Likewise, the S_0 value for each of these is provided by the state.

For each pair of examinations (the NAEP and the state's exam) we seek to develop a mapping formula, $R(S)$, where we have, at most, only a few data pairs and some boundary conditions to which we can fit trial formulas. In the case of the Simple ELQ mapping, the primary data pair consists of the two statewide proficiency percentages (S_0, R_0) as reported by the state's own achievement exams and by the NAEP respectively. As we noted earlier, the ratio S_0/R_0 is defined as the inflation factor. The newer Piecewise Continuous ELQ mapping fits three data pairs and is a straightforward generalization of the Simple ELQ version.

We proceed by performing numerical experiments using artificial scoring distributions to establish links between a simulated easy test and a more difficult one. Arrays of such scoring distributions are parameterized by their peak scores and standard deviations to model the range of results that might occur for a given test being administered to groups of students in which the skill levels among the groups range from very low to very high.

♠ Complications arise if we wish to study high school proficiencies due to a lack of NAEP data. Those issues are addressed in Appendix C of this report. Here our focus is on the 4th and 8th grade assessments.

From those numerical experiments we can establish numerical formulas relating the proficiencies obtained on one exam to those achieved on the other. The resulting curves display characteristics with a geometric interpretation suggesting a mixture of linear, circular, and higher order Lamé curve (quartic) components. The resulting ELQ formulas have been used to fit the data points.

The derived versions of the ELQ formulas are then subjected to error analyses in the simulated examination environments. There we see quite acceptable error levels, particularly in the proficiency ranges of most interest, where they are roughly of the same magnitude as the published NAEP standard errors. We look at measures of average errors as well as their extremes to show the adequacy of the Simple ELQ formula and the superiority of the Piecewise version.

Interested readers may refer to a more technical discussion of our mapping methodologies and analysis in the following report: **ELQ-Mappings.docx**.[xvi]

Other Statewide Data Pairs

The aforementioned data pair (S_0, R_0) could pertain to the testing of all public school students within the state for the given subject and grade level, or could be that of a tested demographic group. While the NAEP does not test all students within a state, it does test representative samples of each demographic group from which the various statewide proficiencies are measured.

Under the terms of No Child Left Behind (NCLB) legislation, state authorities and the NAEP are required to provide statewide proficiencies for a number of specified demographic groups. Thus, we usually have data pairs, (*S*, *R*), for the following nine demographic groups:

1	2	3	4	5	6	7	8	9
All	Disadvantaged	Advantaged	Female	Male	White	Black	Hispanic	Asian

In some states, American Indians are sufficiently numerous to replace one of the other groups.

In our analysis, the question arises as to how these additional data pairs or demographic data items can be used to improve the mapping methods. For each data pair one could alternatively consider using it as an additional "point" to fit or as an indicator of the method's error. We decided that use of any of the ethnic/gender categories (4 – 9) as a data fit point would introduce the appearance of improper discrimination for or against the groups involved. To avoid those issues, we decided that by using the economic status items as additional fit points we could derive a higher order, possibly more accurate, mapping method than would result from fitting just one data pair. Based on these considerations, the two following mapping models have been developed:

> **Simple Ellipse-Quartic (ELQ) Model.** Here we fit only one data pair (S_0, R_0) corresponding to the statewide proficiencies of all students in the subject and grade level denoted. The last six data pairs, 4 – 9, are used to assess the mapping errors by comparing the reported NAEP proficiencies to those predicted by the mapping formula.
>
> **Piecewise Ellipse-Quartic (PW ELQ) Model.** Here we fit three data pairs (items 1-3 in the list) while again using the last six data pairs for error measurements.

Assumptions leading to the ELQ mappings

A number of plausible assumptions have been made to make the analysis more tractable. The assumptions made for these mappings are:

1. We first imagine that all students taking the state achievement exams also took the NAEP and that the NAEP results for this full sample are the same as those reported from the actual NAEP examination, which was actually administered to a smaller sample. This is correct to within the so-called sampling error that is generally on the order of 1% or 2% for the NAEP. These sampling errors are published by NAEP and provide some of the criteria by which we conduct our analysis of errors that arise from our mapping methods or from the examination environments themselves. Appendix C is devoted to that error analysis.
2. We next assume that the two exams, those of the state and the NAEP, are similar in regard to the order of the proficiency percentages among the groups tested. (This means the group with proficiencies *nth* from the top on the NAEP would also rank *nth* on the state examination.) This means that the map will be well-ordered in the sense that any subset ranking relatively lower to a higher subset on the one exam will always rank lower than the higher group on the other exam. This means that our map will be monotone increasing. As will become apparent, this assumption is often violated, yet the monotonic mappings produced still appear to be useful – at least with respect to our goal of removing most of the "inflation" from the state-reported proficiencies.
3. We introduce the variables *S* and *R*. *S* is the percentage or fraction of students at a school, district, statewide or in some other tested group who are designated as proficient on the state's exam. *R* is the percentage or fraction of that same group that would be proficient on the NAEP. Since *S* is known, our goal is to estimate *R* if we are given *S*. Thus we are seeking to build a map relating *S* to *R*. In functional terms we seek to find $R = R(S)$. Since *R* and *S* are percentages, they each run over the range 0% to 100% or (0 to 1). Because of the monotone assumption described above, $R(S)$ is also a monotone increasing function.
4. We already know several data pairs, call them (S_i, R_i), for i=1,9. They are the proficiency percentages from the statewide assessments done by the state and the NAEP, respectively, for the various demographic groups. The two other pairs result from the mathematical requirement that R=0 when S=0 and that S=100% when R=100% (Or S=1 when R=1).
5. We assume that the inflation observed is due either to lowered cut scores or the use of an easier exam or from a combination of them. (This ignores other causal factors. For example, the provision of "special accommodations" to some assessed children or other unspecified "adjustments" can lead to different and usually higher inflation levels for them.)

While not specifically an assumption, we have noticed in our simulated examination environments a symmetry property that appears to be a fair approximation in most of the cases we have studied. The graphical representation of *S* and *R* is "drawn" on the unit square with the horizontal coordinate, *S*, running from 0 to 1, and the vertical coordinate, *R*, running from 0 to 1. As our numerical simulations confirm, consistent with much examination proficiency data, there is an approximate symmetry with respect to the "cross diagonal" of this graphical domain. That is, an observer sitting in the upper left hand corner (where R=1 and S=0) would observe the $R(S)$ curve symmetric about the line connecting to the lower right hand corner (a line that drops with a 45 degree angle). From that observation point, three curves or basis functions that obviously have this symmetry are:

- A straight line on the diagonal going from the origin to the (1,1) point.
- A circle centered at the observer position.
- A unit quartic or Lamé curve centered at the observer position.[♦]

This is evident in Figure A1, shown below.

[♦] Our use of the term "quartic" is not meant to suggest the use of a quartic polynomial, but rather to indicate that terms of 4th degree are in the equation of the Lamé curve.

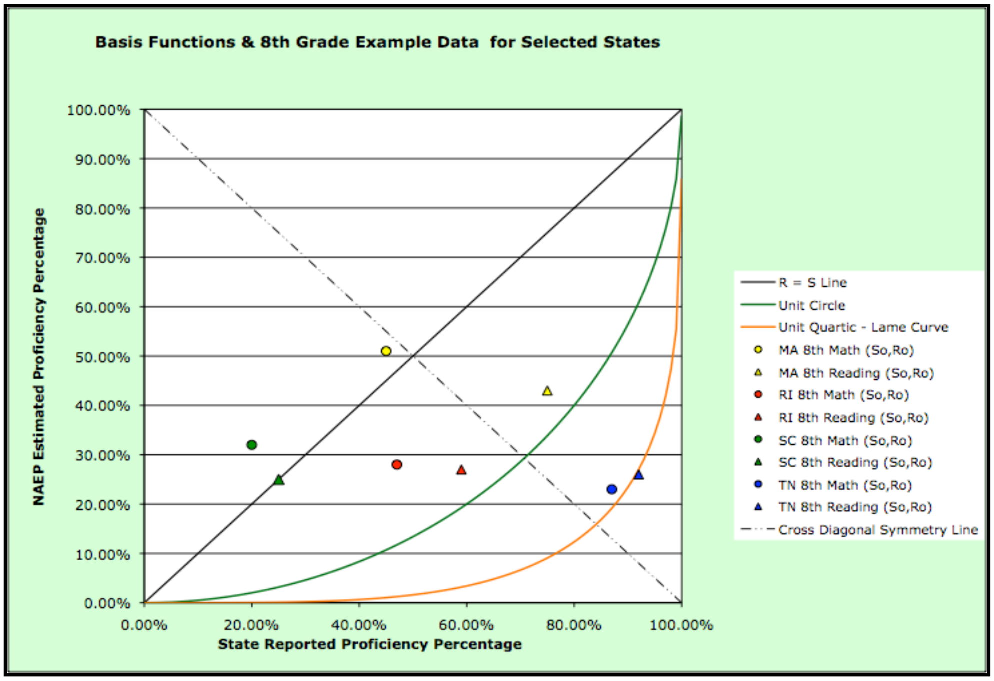

Figure A1. We display the various basis functions and some of the data pairs to be fitted. Also shown is the symmetry line (dashed) about which the basis functions are symmetric. The two data pairs for Tennessee (shown in blue) lie far below the R = S line and therefore exhibit very high levels of proficiency inflation. Also shown are the data pairs for Massachusetts (yellow) and South Carolina (green), which exhibit the lowest levels of proficiency inflation nationally.

Based on these considerations, we decided to try linear combinations of such basis functions (or curves). Strictly speaking, linear combinations of these basis functions will only maintain their symmetry if they are combined in polar coordinates centered at the $(S,R) = (0,1)$ point. The symmetry is approximately preserved, except for S near 1, by combining them in the rectangular (S,R) coordinates.

Intuitively, one might expect the polar coordinate (more symmetric) method to yield superior results, but our testing shows the combination structured in rectangular coordinates works best. Thus, the Ellipse-Quartic (ELQ) formula shown here in Eq. A1 is the one we use.

$$R(S) = \alpha S + \beta\left[1 - \sqrt{1 - S^2}\right] + \phi\left[1 - \sqrt[4]{1 - S^4}\right]$$

Eq. A1

Each of the three terms has a geometric significance: They are the equations of a straight line, a circle, and an equilateral quartic or Lamé curve. We use only the first two terms when (S_0, R_0) lies at or above the circle (plotted in green in Fig. A1) described by the functional relationship:

$$R_c(S) = \left[1 - \sqrt{1 - S^2}\right]$$

Eq. A2

and we use only the last two terms of Eq. A1 if (S_0, R_0) is below the circle.

In Tennessee's testing of 4th and 8th grade students, all of the (S_0, R_0) pairs lie far below the circle and thus in these cases we only use the last two terms of Eq. A1.

Before contrasting the two methods, we review the sub-domains of validity for Eq. A1.

- For the Simple ELQ method, the formula applies continuously over the entire domain $0 < S < 1$.
- The more advanced Piecewise ELQ method applies this formula in the leftmost domain segment of S to the left of the Disadvantaged students' data pair and in the rightmost domain segment to the right of the Advantaged students' data pair, but uses a quadratic polynomial interpolation in the remaining middle segment of the domain.

The coefficients are determined from the known (quantitative) values of the statewide proficiencies reported by the state and the NAEP, from one boundary condition, and from one additional assumption. In most circumstances only the first two terms are used while in more extreme situations only the last two terms are applied. Georgia, Illinois, and Tennessee, which are among the more egregious "inflators," provide examples of the latter.

We have measured the mapping formulas and accuracy against two different data sets:

- Model-simulated distributions of student examination results. We do this for the two major causes of inflated proficiency percentages: Lower cut scores and reductions in examination difficulty.
- We also apply the mapping formulas to real data in order to predict NAEP proficiencies for up to six different demographic groups for which the NAEP also reports proficiencies, thereby allowing further error measurements. They have been successfully applied in all but one of the 26 states we have reviewed, and from them we have generated estimates of the error levels.

We have found the Piecewise ELQ method to be almost always more accurate than the Simple ELQ method in our simulated examination environments. This accuracy advantage is greatest for groups with NAEP proficiencies below 70%, where nearly every tested group resides. The exceptions are those exceedingly rare schools or groups that have NAEP proficiencies exceeding 70% for which the two methods are comparable in their accuracy.

Then, for real data we have been able to measure the errors of our predictions in six other cases in which NAEP scores are known. NAEP not only provides proficiency percentages for statewide samples of all public school students in 4th and 8th grades, it also provides these numbers in Tennessee for seven different demographic subsets: Females, Males, Whites, Blacks, Hispanics, Disadvantaged and Advantaged students. (The standard errors for Asians are too large for their inclusion.) Since Tennessee also reports proficiency percentages for these groups, for each such S-value we can compare the actual R-value against the one predicted from the mapping formula. Often we find errors of the same magnitude as the sampling errors of the NAEP tests. But, as will become apparent in Appendix C, we also have found large errors for certain demographic groups and particularly for Black, Hispanic and Disadvantaged pupils. Are these errors from our mapping techniques or are they attributable to irregularities in the state-run assessment systems? Most evidence suggests the latter. A more detailed discussion of the errors is presented in Appendix C.

In practice, we have found that substantial numbers of states, including Tennessee, employ testing regimes inconsistent with some of the assumptions used to develop these ELQ methods. Nevertheless, we believe the resulting mappings used to generate local NAEP proficiency estimates are reasonably accurate for our purposes. The errors in our analysis are generally much smaller than the proficiency differences that arise from the state-administered examination environment's inflation. Thus the estimates of NAEP performance, though not precise, give one a more realistic and less distorted measure of actual student performance than the state-reported numbers would indicate.

In this section we have attempted to provide an overview of the derivation of the two ELQ mapping formulas. As we indicated above, a presentation of the details can be found in a more "elaborate" report available from the author.[xvii]

Appendix B: Generating High School Mappings

The high school maps require additional approximations

Making estimates at the high school level is more difficult than what is involved in doing the 4th and 8th grade analyses. A number of problems and impediments confront any attempt to convert state reported proficiencies into NAEP scale estimates. They include:

- The NAEP tests do not report state-by-state proficiencies; only national proficiencies are available.
- NAEP tests at the 12th grade while most states test at the 10th or 11th levels. Tennessee tests at 12th grade.

In the next subsections we show how these problems have been overcome. However, the additional assumptions and approximations used will have the effect of introducing additional errors into the analysis. As we think the results will show, using the ELQ method to estimate high school proficiencies still provides useful information about individual schools and districts that is not obtainable any other way of which we are aware.

Approximating 12th grade NAEP proficiencies

We now consider how one might obtain a reasonable approximation of what Tennessee 12th grade public school students would have achieved statewide on the NAEP. Different approaches have been considered:

1. Extrapolate the 4th and 8th grade Tennessee NAEP results linearly to provide a 12th grade number.
2. Apply the scaling factor measured nationally that relates 12th grade NAEP results to those at the 8th grade and then use this proportionality factor in each subject area to generate a Tennessee 12th grade proficiency percentage from its NAEP-measured 8th grade proficiency.
3. Using a least squares method, analyze the national relationships of the 4th, 8th, and 12th grade NAEP proficiencies for the several different demographic groups to determine a formal linear relationship between the 12th grade proficiencies and those at the 4th and 8th grades. Then apply this same linear relationship to the Tennessee 4th and 8th grade NAEP scores to provide an approximation of the 12th grade scores. Refine this approximation by using the nationally determined errors for each demographic group to provide a correction increment to be added to the linear approximation.

To get an idea of how these methods compare, we look at their predictions for national proficiencies as shown in the following tables. These numbers pertain to 2005, the last year for which national NAEP proficiencies were reported.

What The Various Methods Predict Nationwide Versus The Actual Measurements of Mathematics Proficiencies									Sum of
Method	Disadvantaged	White	Black	Hispanic	Advantaged	Female	Male	All	Squares Error
Simple Extrapolation	8.0%	28.0%	4.0%	8.0%	29.0%	22.0%	24.0%	22.0%	0.15%
Scaling	10.2%	29.1%	6.3%	10.2%	30.6%	21.2%	23.6%	22.0%	0.35%
Least Squares	7.8%	25.5%	4.3%	7.8%	26.1%	21.1%	22.7%	20.8%	0.04%
Actual USA	7.0%	28.0%	5.0%	7.0%	26.0%	20.0%	24.0%	22.0%	
Maximum Error	3.2%	1.1%	1.3%	3.2%	4.6%	2.0%	1.3%	1.2%	
What The Various Methods Predict Nationwide Versus The Actual Measurements of Reading Proficiencies									Sum of
Method	Disadvantaged	White	Black	Hispanic	Advantaged	Female	Male	All	Squares Error
Simple Extrapolation	23.0%	43.0%	18.0%	21.0%	42.0%	43.0%	29.0%	36.0%	0.40%
Scaling	17.6%	43.4%	12.9%	16.4%	44.6%	39.9%	28.1%	34.0%	0.39%
Least Squares	20.1%	40.2%	15.2%	18.2%	39.4%	39.9%	26.3%	33.1%	0.01%
Actual USA	19.0%	41.0%	15.0%	18.0%	39.0%	40.0%	27.0%	34.0%	
Maximum Error	4.0%	2.4%	3.0%	1.6%	0.4%	0.1%	0.7%	3.0%	

Table B1. Applying the three methods to the nationwide NAEP scores of 4th and 8th grade students produces these estimates for 12th grade mathematics and reading proficiencies. The actual 12th grade scores are known so this provides a means to compare the three optional methods. The least squares method is clearly superior with its smaller errors.

As the table shows, judged by the sum of the squared errors, the least squares method is superior to the others in its ability to predict national scores. We assume this advantage carries over to statewide NAEP scores. When we apply the three methods to Tennessee's statewide data for 2007 we find a range of results as shown in the next table.

What The Various Methods Predict Statewide for Mathematics Proficiencies, Showing Their Ranges of Values								
Method	Disadvantaged	White	Black	Hispanic	Advantaged	Female	Male	All
Simple Extrapolation	8.0%	25.0%	6.0%	12.0%	25.0%	15.0%	22.0%	18.0%
Scaling	9.4%	23.6%	5.5%	10.2%	25.1%	15.7%	20.4%	18.1%
Least Squares	7.3%	26.5%	7.9%	12.1%	23.2%	13.3%	22.8%	18.5%
Maximum Value	9.4%	26.5%	7.9%	12.1%	25.1%	15.7%	22.8%	18.5%
Minimum Value	7.3%	23.6%	5.5%	10.2%	23.2%	13.3%	20.4%	18.0%
Error Range	2.1%	2.9%	2.4%	1.9%	2.0%	2.4%	2.3%	0.5%
What The Various Methods Predict Statewide for Reading Proficiencies, Showing Their Ranges of Values								
Method	Disadvantaged	White	Black	Hispanic	Advantaged	Female	Male	All
Simple Extrapolation	22.0%	38.0%	16.0%	24.0%	39.0%	39.0%	25.0%	33.0%
Scaling	16.4%	37.5%	9.4%	21.1%	41.0%	35.2%	24.6%	30.5%
Least Squares	18.0%	36.0%	12.9%	21.1%	36.0%	36.0%	23.1%	31.0%
Maximum Value	22.0%	38.0%	16.0%	24.0%	41.0%	39.0%	25.0%	33.0%
Minimum Value	16.4%	36.0%	9.4%	21.1%	36.0%	35.2%	23.1%	30.5%
Error Range	5.6%	2.0%	6.6%	2.9%	5.0%	3.8%	1.9%	2.5%

Table B2. Applying the three methods to the statewide Tennessee NAEP scores of 4th and 8th grade students produces these estimates for 12th grade mathematics and reading proficiencies. Here the actual 12th grade scores don't exist so it is not possible to measure errors directly. However the "Error Range" shown provides a rough measure of the disparity of the results, which we take to be a measure of a range of errors.

In considering the statewide predictions of these three methods, the least reliable is that for the Black students where the range of predicted proficiencies is close to 7%. This appears consistent with our claim that the Piecewise ELQ errors will be less than 10% at the high school level. While consistent, it is not sufficient to verify it.

Appendix C: Accounting for Errors

To understand the measured errors with respect to the proficiencies reported for the demographic groups, we consider the known errors stemming from the NAEP sampling techniques and those theoretically calculated for the two mapping methods. As we compare these known errors against the measured ones, we sometimes find discrepancies or larger errors than expected. For such cases, additional causal factors, such as *special accommodations*, may affect the reported proficiencies.

Measured errors

We check the accuracy of each ELQ mapping formula by showing the actual (*S*, *R*) pairs for the demographic groups that are not used as data fit points, and which are also tested by NAEP in each state. Here, *S* and *R* denote the TCAP and NAEP measured proficiencies, respectively. They are denoted by the X marks shown on the graphs presented in this report's main body - in Figures 1a- 3b. The vertical distance between the X positions and the blue or orange mapping curves, for the Piecewise and Simple ELQ formulas respectively, are the measured errors with respect to the prediction of the mapping formula for each of the several demographic groups. The NAEP organization also publishes the "standard errors" associated with each of these data pairs, which are due to the effects of sampling methodologies.

Expected errors

In the mappings, there are the *expected errors* that result from the approximations in the mapping methods themselves and from the sampling errors in the NAEP proficiencies that are input into the model.

ELQ method error bounds

In the derivation of the two versions of the ELQ mapping method ix, artificial numerical scoring distributions were used to generate a theoretical numerical mapping relationship *R(S)*. The Simple and Piecewise ELQ formulas are algebraic approximations to those numerical relationships that are, respectively, constrained to fit either one numerical data pair or three of them. In that theoretical modeling, we were able to calculate the errors encountered by comparing the algebraic formula result with the numerical one. The error or displacement is a signed quantity defined as the algebraic formula value for *R* less the numerically generated value for *R*. When the numerical value exceeds the algebraic one, called here the topside situation, the displacement is negative, and vice versa for the bottomside situation. Table C1 shows the error bounds measured in this environment for the two methods.

Error Bounds in Artificial Simulation Environment				
	For R < .3		For R > .3	
	Topside	Bottomside	Topside	Bottomside
Simple ELQ	-3.11%	1.00%	-4.60%	1.60%
PW ELQ	-2.20%	0.40%	-4.50%	0.60%

Table C1. We show the error bounds that always enclose the numerically generated errors in our artificial simulation environments. Shown for two ranges of R values.

NAEP standard errors

The NAEP tests students on a statewide basis and does this for several different demographic groups. Given that the NAEP uses sampling methods to determine average statewide proficiency percentages, there is a sampling error, or standard error, from that sampling which the NAEP organization reports with its proficiency numbers. The standard error is the standard deviation of the error distribution (assumed to be a normal distribution) that would be obtained if the testing were repeated over a large number of random samplings. As such it would encompass approximately 68% of the random errors. If, instead, one considers an error interval twice that of the standard error, one would have two standard deviations in the error interval, which then contains about 95% of the random errors. This double-wide interval is usually called the "confidence interval," wherein data will be found in 95% of the instances if the errors are purely random and do not have other so-called systematic error components. Thus, our analysis assumes that the contribution of the NAEP standard error to our expected error bounds will be this double-width interval.

Topside and bottomside error bounds

Our error analysis is based on constructing an interval in the variable *R* that brackets the mapping curve. A so-called error bar will extend above the mapping curve by adding the absolute magnitude of the topside error in Table C1 to twice the standard error for the demographic group/grade level/subject being studied.

Likewise, we construct the displacement below the mapping curve that matches the bottomside error bound. That negative displacement is found by subtracting the bottomside error in Table C1 from minus twice the standard error for that case.

With this interval constructed, we can now examine the measured error to determine if the error is within the bounds or exceeds them.

Let's first consider the errors found when the Simple ELQ mapping curve is employed. Figure C1 depicts the error analysis in graphical form as it applies to Tennessee 8th grade mathematics proficiencies.

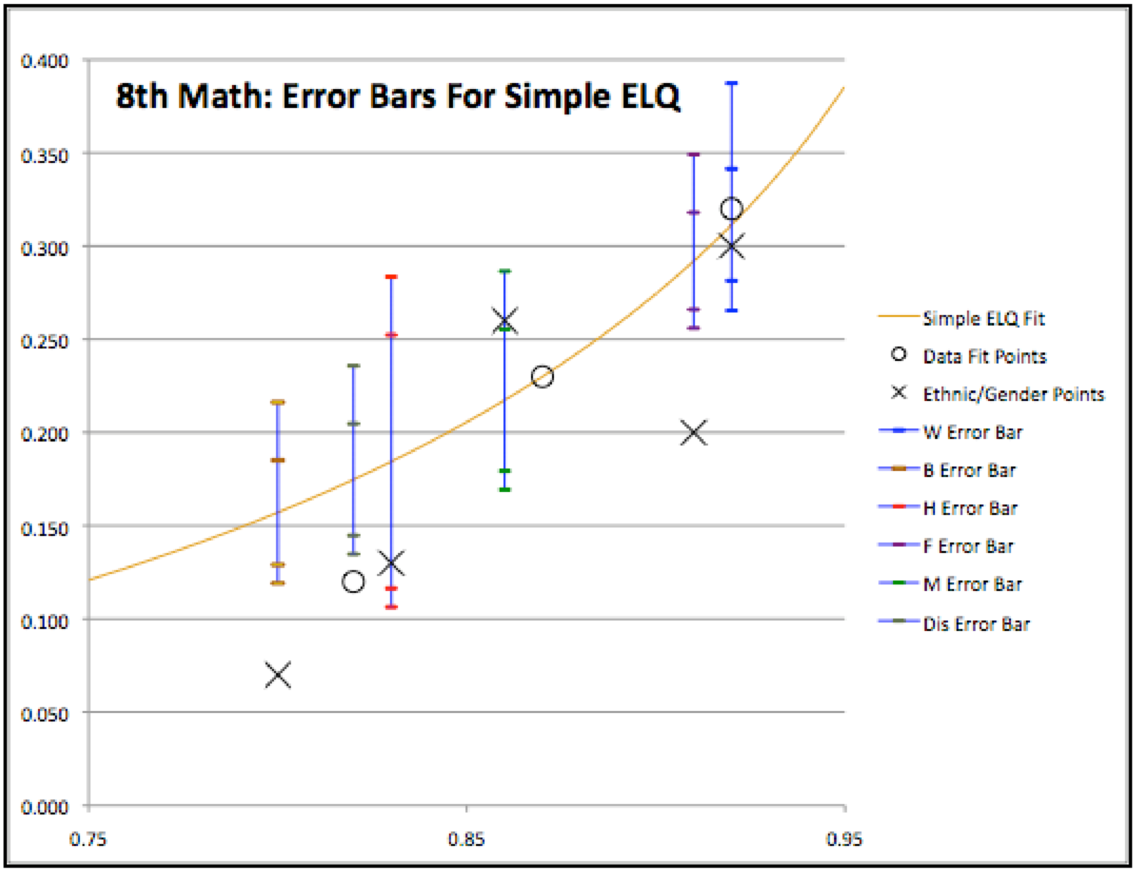

Figure C1 shows the mapping curve for 8th grade mathematics for the Simple ELQ formula, the data pairs for various demographic groups for which statewide proficiencies are reported by both the Tennessee authorities and by the Nation's Report Card, and the error bars that represent the expected errors (which ideally should bracket the x or o markers).

In the figure, left to right, the X data points represent Black, Hispanic, Male, Female, and White students, while the O markers show the data pairs for Disadvantaged, All, and Advantaged students. The error bars show an inner interval defined by twice the NAEP standard error, as well as their extension from the errors calculated in the theory of the mapping method. The latter typically have a higher topside extent than on the bottomside as is also evident in the figure. When the data points indicated by the X or O markers lie outside of the error bars, it is highly likely that the error has additional causal factors not considered in the analysis- such as the effects from special accommodations in the testing environment.

Figure C1 shows three such data pairs lying below the indicated error ranges, including the data points for Blacks, Disadvantaged, and Female students. In our analysis of other states, similar situations were often found. Moreover, among the states it has been rare to find the data pairs for the lower-performing groups that lie above the mapping curves. In fact, it was this observation that helped motivate doing a three point fit mapping formula, in which the Disadvantaged and Advantaged data pairs are also fit by the Piecewise Continuous ELQ mapping curve. Figure C2 shows the representation of errors for the Piecewise ELQ.

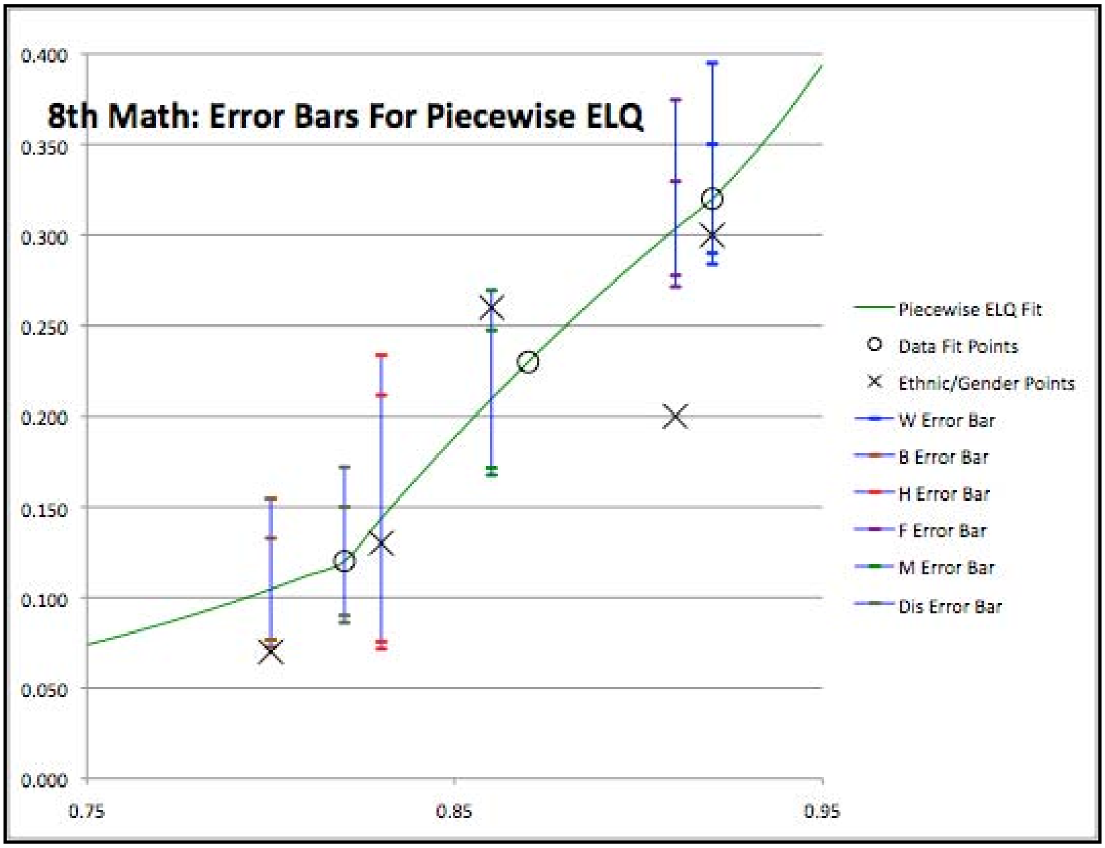

Figure C2 shows the analogous error configuration for the Piecewise ELQ mapping as applied to 8th grade math students. While the data point for Black students lies outside the error range, it is very close to its lower bound. As you would expect, the Disadvantaged O marker now is fit perfectly. The one glaring large displacement is for female students, which was also seen in Fig. C1. There is clearly a problem when, as we see here, the state test is showing girls better than boys in math, while the NAEP shows the opposite.

These two figures are just two among twelve such representations that cover the three grade levels, the two subjects, and the two mapping versions that were applied to the Tennessee TCAP data. Space doesn't permit display of the others. Rather we present tables of the same information in the next section to cover all of the relevant situations.

Error profiles in tabular format

The next several tables present in numerical form the error bounds like those depicted in the preceding two figures. The first group of tables describes the results for the Simple ELQ - sometimes called the Old ELQ - mapping method as applied to the Tennessee TCAP data and that of the NAEP.

Error analysis results for the Simple ELQ mapping method

We begin by presenting the error intervals for the upper and lower bounds. In Figure C1, the topside or upper error bounds corresponded to the vertical distance from the crossing of the mapping curve to the top of each error bar which, incidentally, is a vector quantity (pointing up) of positive sign. The lower error bounds, likewise, correspond to the vertical distance from the crossing of the mapping curve to the bottom of each error bar, which here is a vector quantity (pointing down) of negative sign.

Table C2 shows these numbers for the six situations corresponding to the three grade levels and two subjects.

	Topside Error Bound For The Old ELQ Mapping Formula					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	5.51%	5.51%	5.91%	5.31%	5.91%	6.80%
Disadvantage	5.51%	5.71%	6.11%	6.31%	6.11%	6.31%
Advantaged	8.40%	8.20%	8.40%	7.80%	6.91%	7.80%
White	7.80%	7.40%	7.60%	7.20%	6.11%	7.20%
Black	5.51%	5.71%	5.91%	6.31%	5.91%	6.31%
Hispanic	11.91%	13.11%	9.91%	14.91%	9.91%	16.40%
Female	7.40%	7.60%	5.71%	8.20%	5.71%	8.20%
Male	6.71%	6.31%	6.91%	5.91%	6.91%	5.91%
	Bottomside Error Bound For The Old ELQ Mapping Formula					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-3.40%	-3.40%	-3.80%	-3.20%	-3.80%	-3.80%
Disadvantage	-3.40%	-3.60%	-4.00%	-4.20%	-4.00%	-4.20%
Advantaged	-5.40%	-5.20%	-5.40%	-4.80%	-4.80%	-4.80%
White	-4.80%	-4.40%	-4.60%	-4.20%	-4.00%	-4.20%
Black	-3.40%	-3.60%	-3.80%	-4.20%	-3.80%	-4.20%
Hispanic	-9.80%	-11.00%	-7.80%	-12.80%	-7.80%	-13.40%
Female	-4.40%	-4.60%	-3.60%	-5.20%	-3.60%	-5.20%
Male	-4.60%	-4.20%	-4.80%	-3.80%	-4.80%	-3.80%

Table C2 shows the topside and bottomside error bar intervals within which random errors from the NAEP samplings and errors inherent in the mapping technique could likely explain any data pair that "lands" within the error bar. The relatively larger bars for Hispanic students are primarily attributed to their larger NAEP standard errors. The larger numbers for Advantaged students are mostly due to somewhat larger errors from the mapping technique that arise when it is applied to higher proficiency levels.

	Bad Error Excess Topside For The Old ELQ Mapping					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-5.51%	-5.51%	-5.91%	-5.31%	-5.91%	-6.80%
Disadvantaged	-10.89%	-12.84%	-11.58%	-11.80%	-14.78%	-11.24%
Advantaged	-7.49%	-4.62%	-7.54%	-9.46%	-3.59%	-9.39%
	Cells Outside Error Bars Are Shown In Yellow					
White	-8.42%	-8.82%	-8.74%	-8.53%	-2.39%	-10.00%
Black	-15.82%	-16.05%	-14.62%	-16.51%	-11.18%	-11.89%
Hispanic	-24.39%	-11.45%	-15.35%	-12.84%	-21.28%	-32.93%
Female	-13.85%	-12.04%	-14.91%	-11.53%	-12.25%	-5.27%
Male	-3.19%	-5.74%	-2.65%	-7.33%	-2.02%	-9.99%
	Bad Error Excess Bottomside For The Old ELQ Mapping					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-3.40%	-3.40%	-3.80%	-3.20%	-3.80%	-3.80%
Disadvantaged	1.98%	3.53%	1.47%	1.29%	4.67%	0.73%
Advantaged	-6.31%	-8.78%	-6.26%	-3.14%	-8.12%	-3.21%
	Cells Outside Error Bars Are Shown In Yellow					
White	-4.18%	-2.98%	-3.46%	-2.87%	-7.72%	-1.40%
Black	6.91%	6.74%	4.91%	6.00%	1.47%	1.38%
Hispanic	2.68%	-12.66%	-2.36%	-14.87%	3.57%	3.13%
Female	2.05%	-0.16%	5.60%	-1.87%	2.94%	-8.13%
Male	-8.12%	-4.77%	-9.06%	-2.38%	-9.69%	0.28%

Table C3 presents the displacement of the data points with respect to the upper and lower bounds of the error bars. Only points outside of these boundaries will show a positive displacement, which are shown in yellow shaded cells. None of the measured errors had a data point above the upper bound as indicated in the top half of the table where there are no yellow shaded cells. The lower half, which shows the displacement below the lower bounds, is positive for a number of cases. They are shown in yellow.

These tables present the error information for all of the considered cases, unlike the previous Figures that only showed the situation for 8th grade mathematics. The third column of these tables, pertaining to 8th grade mathematics, shows the numbers behind the plots in Fig. C1.

Another way to describe Tables C2 and C3 is to say that the first shows the range of allowable errors, while the second shows whether the errors of the actual data pairs fell within the allowable ranges. Most of our discussion focuses on Table C3, where we can see which cases had errors out of bounds.

Probably the most obvious characteristic of the errors encountered in these mappings is that there are no cases for which the errors are topside - that is to say, the data points always lie below the upper error bound. This is surely the case for Figure C1, which shows the 8th grade math mappings, data, and errors.

We also find no cases in which students in the categories of Whites, Advantaged students, or All, have errors to the bottomside that exceed the lower error bounds.

Also common to all cases is the fact that the demographic groups of Disadvantaged and Blacks always have excessive bottomside errors, indicated by the "yellow" rows in Table C3.

The situations for Hispanic, Female, and Male students are mixed.

In the case of Hispanic students, two of the three math data pairs fall below the lower bound and the one that doesn't is close to the lower bound. Instead of thinking of these errors as vertical downward shifts of the NAEP proficiencies, it is equivalent to consider them as horizontal shifts in which the TCAP reported numbers, *S*, are shifted significantly to the right. We call these "upshifts." So in this table we see evidence that Hispanic math proficiencies have upshifts in all three tested grade levels.

Female students also show out of bounds errors for mathematics proficiencies at all three tested grade levels, but not for reading. Fig. C1, for example, shows a large upshift for female math proficiencies not seen for male students. The female math proficiencies are always larger than those of the males as measured by the TCAP and always smaller than those of males as reported by the NAEP. This reordering of math proficiencies violates one of the assumptions discussed in the previous Appendix, with the effect that the mapping formula will display larger errors than generally found when its derivation's assumptions are met.

Finally, the 12th grade reading proficiencies of Hispanic and Male students have data pairs laying below the error bar lower bound. The excess for Males is fairly small and may not be significant, but the one for Hispanic students, though not big, is sufficiently large to warrant discussion.

The causal factors will be discussed briefly farther along.

Error analysis results for the Piecewise Continuous ELQ mapping

The analogous tables for the Piecewise Continuous ELQ mapping method are shown next:

	Topside Error Bound For the PW ELQ Mapping Formula					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	4.60%	4.60%	5.00%	4.40%	5.00%	6.70%
Disadvantage	4.60%	4.80%	5.20%	5.40%	5.20%	5.40%
Advantaged	8.30%	8.10%	8.30%	7.70%	6.00%	7.70%
White	7.70%	7.30%	7.50%	7.10%	5.20%	7.10%
Black	4.60%	4.80%	5.00%	5.40%	5.00%	5.40%
Hispanic	11.00%	12.20%	9.00%	14.00%	9.00%	16.30%
Female	7.30%	7.50%	7.10%	8.10%	4.80%	8.10%
Male	5.80%	5.40%	6.00%	5.00%	6.00%	5.00%
	Bottomside Error Bound For The PW ELQ Mapping Formula					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-2.80%	-2.80%	-3.20%	-2.60%	-3.80%	-3.80%
Disadvantage	-2.80%	-3.00%	-3.40%	-3.60%	-4.00%	-4.20%
Advantaged	-4.40%	-4.20%	-4.40%	-3.80%	-4.80%	-4.80%
White	-3.80%	-3.40%	-3.60%	-3.20%	-4.00%	-4.20%
Black	-2.80%	-3.00%	-3.20%	-3.60%	-3.80%	-4.20%
Hispanic	-9.20%	-10.40%	-7.20%	-12.20%	-7.80%	-13.40%
Female	-3.40%	-3.60%	-3.20%	-4.20%	-3.60%	-5.20%
Male	-4.00%	-3.60%	-4.20%	-3.20%	-4.80%	-3.80%

Table C4 is very much the same as Table C2 except it measures its errors with respect to the Piecewise Continuous ELQ mapping curves.

	Bad Error Excess Topside For The PW ELQ Mapping					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-4.60%	-4.60%	-5.00%	-4.40%	-5.00%	-6.70%
Disadvantaged	-4.60%	-4.80%	-5.20%	-5.40%	-5.20%	-5.40%
Advantaged	-8.30%	-8.10%	-8.30%	-7.70%	-6.00%	-7.70%
	Cells Outside Error Bars Are Shown In Yellow					
White	-9.58%	-12.30%	-9.50%	-8.13%	-4.90%	-8.31%
Black	-9.84%	-8.44%	-8.46%	-10.20%	-2.82%	-6.71%
Hispanic	-22.68%	-3.84%	-10.36%	-6.71%	-23.82%	-31.23%
Female	-14.82%	-15.18%	-17.46%	-11.13%	-14.66%	-5.09%
Male	-1.48%	-2.32%	-0.96%	-4.38%	0.80%	-7.94%
	Bad Error Excess Bottomside For the PW ELQ Mapping					
Level/Subject	4th Math	4th Reading	8th Math	8th Reading	12th Math	12th Reading
All	-2.80%	-2.80%	-3.20%	-2.60%	-3.80%	-3.80%
Disadvantaged	-2.80%	-3.00%	-3.40%	-3.60%	-4.00%	-4.20%
Advantaged	-4.40%	-4.20%	-4.40%	-3.80%	-4.80%	-4.80%
	Cells Outside Error Bars Are Shown In Yellow					
White	-1.92%	1.60%	-1.60%	-2.17%	-4.30%	-2.99%
Black	2.44%	0.64%	0.26%	1.20%	-5.98%	-2.89%
Hispanic	2.48%	-18.76%	-5.84%	-19.49%	7.02%	1.53%
Female	4.12%	4.08%	7.16%	-1.17%	6.26%	-8.21%
Male	-8.32%	-6.68%	-9.24%	-3.82%	-11.60%	-0.86%

Table C5, like its analog in Table C3, displays the excess by which a data point might lie outside the error bounds. This happens when the excess is a positive number as holds true for the cells shaded in yellow. Unlike before, there is one excess error, for 12th grade male math, exceeding the topside error bounds.

As in the previous section, which was devoted to the errors found in the Simple ELQ mappings, here we are trying to understand the kinds of errors encountered when the more accurate Piecewise Continuous ELQ mappings are used.

As before in the Simple ELQ cases, we have no instances of topside errors being out of bounds.

There are substantially fewer data pairs now out of bounds on the bottomside than before: 11 such pairs now compared to 19 for the Simple ELQ mappings. There are none for 12th grade reading, regardless of the demographic group.

As might be expected when fitting the three data pairs for the demographic groups Disadvantaged, All, and Advantaged, none of them will display out-of-bounds errors as they are forced to lie on the mapping curve. Males also have no proficiency data pairs out of bounds.

White students' proficiency in 4th grade reading was almost 2% below the lower error bar and needs further investigation. 4th and 8th grade Black students had out-of-bounds proficiencies in both subjects, while Hispanics' proficiencies were out of bounds in 4th and 12th grade mathematics.

The largest excesses pertained to female students and particularly for mathematics proficiencies.

The upshifting phenomena for the Simple ELQ mapping

As we previously mentioned, excessive downward displacement of a data pair from the mapping curve can equally well be considered as a rightward displacement from the mapping curve. This displacement is a measure of the mapping errors as measured in units of S rather than the conventional measurement in units of R.

In the theory that led to the Simple ELQ mapping, only two mechanisms were used to differentiate the model for the NAEP versus the model for the state test: varying cut scores and varying peak scores. The latter amounted to making tests easier or more difficult than the reference test. Almost independent of the relative proportions of these two ingredients, the derived mapping formulas took the Simple ELQ form.

As we have applied the Simple ELQ mapping formula to the schools and demographic groups of various states, we have seen similar characteristics of the error patterns, particularly for groups having significant numbers of disadvantaged students. We rarely see any excess errors to the topside; they are almost always to the bottomside.

One way such patterns could arise would be from additional state-imposed policies or practices that affect different demographics differently. To put this in terms of the data, let's look at the errors as measured in terms of the upshift displacements. Table C6 shows the displacements for Tennessee students with respect to the Simple ELQ mapping curves.

	Upshift Displacements to Right of Simple ELQ Mapping Curves					
	For ethnic and gender groups, those that are +4% and above are shown in pink.					
Level/Subject	**4th Math**	**4th Reading**	**8th Math**	**8th Reading**	**12th Math**	**12th Reading**
All	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Disadvantaged	6.00%	9.00%	7.00%	5.00%	17.00%	6.00%
White	1.00%	1.00%	1.00%	1.00%	-5.00%	2.00%
Black	15.00%	17.00%	15.00%	12.00%	11.00%	9.00%
Hispanic	12.00%	-2.00%	6.00%	-2.00%	17.00%	17.00%
Female	4.00%	3.00%	7.00%	2.00%	10.00%	-3.00%
Male	-2.00%	0.00%	-3.00%	1.00%	-7.00%	5.00%

Table C6 measures horizontally in the S coordinate to obtain the displacement of the data pair from the Simple ELQ mapping curve. Positive numbers reflect a rightward displacement, or "upshift." Black students' proficiencies display large upshifts (indicated by pink cells) while Hispanic and Female students often show large upshifts. Males, perhaps oddly, show a significant upshift in 12th grade reading proficiencies. The 12th grade Hispanic upshifts are sufficiently large to bring their state-measured proficiencies to the same level as White's.

One is tempted to investigate additional causal factors that would lead to these sometimes large upshifts in the *S* values. There are a number of possibilities that might be explored:

- The state testing regime provides *special accommodations* more liberal than those of the NAEP.
- The state content standards are narrower or qualitatively much different than those of the NAEP.
- The state testing administration improperly manipulates testing scores to produce better results.

The scope of our work in this report is limited and prevents us from exploring much in these areas except for some observations about so-called *special accommodations,* which are discussed next.

Upshifting brings 12th grade Hispanic proficiencies to equal those of White's

For Hispanic students in Tennessee, the magnitude of the upshifting phenomenon is sufficiently large at the 12th grade level that it brings the statewide TCAP proficiencies in both math and reading to the same level as those of White students. In those same subjects, the NAEP tests, to the contrary, show a wide gulf between Hispanic and White students in the proficiency percentages they report. One might argue that the precise level of proficiencies is not very important because of all the approximations and errors in the data. However, when the errors (here due to upshifting) are so large as to change the qualitative picture of an ethnic group's performance, it degrades the utility of the results. We think the use of *special accommodations* for many of these students may be part of the explanation, while we do not rule out other causes when the effect is so large.

Six kinds of errors

We have shown that the expected errors that derive from the ELQ derivations and from NAEP sampling errors sometimes do not explain or contain the actual proficiency numbers reported by the NAEP and TCAP testing systems. We have seen that data pairs falling outside of the expected ranges nearly always fall below the mapping curves of either ELQ method and correspond, alternately, to an upshift or rightward displacement of the TCAP proficiencies from the mapping curves. Among possible causes, the use of *special accommodations* is almost certainly part of the story. The causes of the upshifts should be the subject of subsequent studies. They would analyze with more rigor and detail the relationships of the various errors and how they relate to the two testing environments.

In such an analysis there are several kinds of errors to be considered. A partial list:

1. The NAEP Standard Error (these are published by NAEP)
2. The errors intrinsic to the Piecewise ELQ mapping method
3. The errors (systematic errors) caused by the use of special accommodations for students taking the tests.
4. The errors caused by other data manipulations that are consistent with the legal and regulatory systems.
5. The errors caused by incompetent or illegal test administration procedures and data manipulations.
6. The inflation itself as a type of error (one of exaggeration).

Whatever the magnitude of these various errors (1 - 5) their combination is nearly always considerably smaller than the inflation error (6). This means our mappings will still remove most of the inflation. It means that our estimates of NAEP proficiencies for various schools and districts will not be precise, but will be significantly more realistic than the exaggerated ones reported by the TCAP assessments.

The bottom line: Our NAEP estimates of student proficiency in Tennessee are more realistic than the reported TCAP numbers. Until the TCAP system can be reformed, education stakeholders should use these NAEP estimates, as they should provide more transparency to Tennessee parents, educators, and others interested in improving the state's public school systems.

Endnotes

[i] About the Author: David V. Anderson, Ph.D.
1971: University of California, Davis. Ph.D. in Physics
1968: University of Nevada, Reno. M.S. in Physics
1963: Johns Hopkins University. B.A.
Dr. Anderson began his career as a classroom teacher in Oakland, CA. Following completion of the Ph.D., he served for several years as a research physicist, first at the Naval Research Laboratory in Washington, D.C. and then at the Lawrence Livermore National Laboratory in Livermore, CA. In recent years, his work has focused on educational research, development, and policy analysis.

ii Please see the glossary at http://nationsreportcard.gov/glossary.asp.

iii For the 25 states and the District of Columbia, the reported assessment proficiencies have been obtained from or checked against the database at http://www.schooldatadirect.org/. Although many states have reported more recent results than the 2007 testing year, all of our data items are taken from the 2007 testing year to maintain temporal consistency with the NAEP data. This state-generated data is usually also available from the respective state departments of education. However, in comparing the states, we found it easier and more defensible to take the state assessment systems' data from the one source provided by SchoolDataDirect.

iv All of the NAEP-reported proficiencies and standard errors have been taken from the *NAEP Data Explorer* located on the Internet at http://nces.ed.gov/nationsreportcard/naepdata/. The 2007 testing year data is the most recently available at the time of this report's writing.

v *SchoolDataDirect.* All of our proficiency data reported by the Tennessee Department of Education was obtained indirectly from the data repository of School Data Direct. Please visit www.schooldatadirect.org for details.

vi Chester Finn, *E Pluribus Unum? Two longtime school reformers debate the merits of the national curriculum*, Education Next, Spring 2009, p. 55, available at http://educationnext.org/e-pluribus-unum-2/.

vii While some of the data for Tennessee public schools was available from the Tennessee Department of Education's website at: http://edu.reportcard.state.tn.us/pls/apex/f?p=200:30:1369775186832476::NO::P30_SELECTION:AAAD there were certain inconsistencies in the numbers that made these data sets unusable.

viii The NAEP data was obtained directly from the NCES website's NAEP Data Explorer at http://nces.ed.gov/nationsreportcard/nde/.

ix Many of the details of the derivation and analysis of the two ELQ methods are in the downloadable reports titled **ELQ-Derivation.xlsx** and **ELQ-Mappings.docx.** These are available at http://www.education-consumers.org/ELQ-Derivation.xlsx and http://www.education-consumers.org/ELQ-Mappings.docx.

x Ibid.

xi *SchoolDataDirect.* Op. cit.

xii Finn, Op. cit., p. 55.

xiii Jim Hull, *Mapping state cut scores against NAEP: The proficiency debate*, Center for Public Education website at http://www.centerforpubliceducation.org.

xiv (NCES2007-482) Found in *Mapping 2005 State Proficiency Standards Onto the NAEP Scales*, Research and Development Report, June 2007, U.S. Department of Education, Institute of Education Sciences. Report is downloadable from http://nces.ed.gov/nationsreportcard/pubs/studies/2007482.asp.

xv Gary W. Phillips, *Expressing International Educational Achievement in Terms of U.S. Performance Standards: Linking NAEP Achievement Levels to TIMSS,* American Institutes for Research®, April 24, 2007. This report downloads from http://www.air.org/news/documents/naep-timss.pdf.

xvi We provide many additional details in the report, **ELQ-Mappings.docx**, and its accompanying spreadsheet, **ELQ-Derivation.xlsx** These are available at http://www.education-consumers.org/ELQ-Derivation.xlsx and http://www.education-consumers.org/ELQ-Mappings.docx.

xvii Ibid.