

APPENDIX 1

RESPONSE TO ARE WE THERE YET? WHAT POLICYMAKERS CAN LEARN FROM TENNESSEE'S GROWTH MODEL

By William L. Sanders
SAS EVAAS®
SAS® Institute, Inc.

Several years ago, I was encouraged by various educational policy leaders to initiate studies on how the AYP part of NCLB could be augmented so that educators working within schools with entering populations of very low achieving students could be given credit for excellent progress with their students and avoid being branded as a failing school. From the Tennessee longitudinal database, it did not take long to find “poster child” schools. One Memphis City school at that time had 3rd graders with a mean achievement level that would map to the 26th percentile of all Tennessee 3rd graders. Yet this school had this same cohort of students leaving 5th grade at approximately the 50th percentile relative to the state distribution. Still this school was failing AYP because the 3rd, 4th and 5th grade scores had to be composited. Relative to NCLB, this was a failing school. In our view, this was not a failing school; rather it was an outstanding school and should not suffer the indignity of the failing school label.

Additionally, in our early investigations, we found schools that had passed AYP, yet had many of their students who had received the proficiency designation to have trajectories that would lead to a non-proficiency status in the future. It was our intent to develop a process that would give positive recognition to those schools that were truly ramping up their students' achievement levels, while not giving credit for those schools whose all ready proficient students were being allowed to slide. In other words, we endeavored to create a methodology to give the current schools credit for changing academic trajectories so that their students would have the opportunity to meet various academic standards in the future if effective schooling was sustained into the future. We sought methods consistent with the intent of NCLB to rectify the mislabeling of very effective schools as failing. We were encouraged to develop a process to be

an augmentation of the AYP process, not a replacement for the USDE approved AYP process.

At the time we initiated these studies, we had many years of experience working with longitudinally merged student test data and knew of many difficult problems that would have to be addressed in order to achieve a process that would have fairness and reliability. We considered and rejected several approaches. One of the first to be rejected is one of the simplest. For example, one such approach would be to take a student's current test score, subtract that score from the proficiency cut score three years in the future, divide the difference by three to yield how much progress that student must make per year. If, in an intervening year, the student's score exceeds the target score, then that student would be deemed to have made appropriate growth. Then the percentage of students making growth could be calculated for each schooling entity (i.e. district, school or subgroup) and the approved AYP rules could be applied as for the AYP status procedure.

We rejected this approach because (1) the error of measurement in any one score for any one student is so relatively large that the setting of an improvement target by this method will inevitably send the wrong signals for many students, and (2) by definition vertically scaled tests provide scores that are intrinsically nonlinear over grades, resulting in uneven expectations of growth at the student level. We believed there to be a process that avoided these two major problems and would result in greater reliability for the final estimation of whether or not a school had earned the right to be considered a non-failing school even though it had not met the regular AYP requirements.

One of our first objectives was to avoid making judgment about the progress of an individual student based upon

APPENDIX 1 (continued)

one test score—like what is done with simple approaches similar to the one outlined above. To minimize the error of measurement problem associated with one test score, we elected to use the entire observational vector of all prior scores for each student. In some of our earlier work, we had found that if at least three prior scores are used, then the error of measurement problem is dampened to be no longer of concern.¹ *This number independently of us was also found by researchers at RAND, Inc.*

The goal is to give the **current** school credit for changing the trajectory of its students so that they can meet various academic attainment levels in the future. How is this to be measured? Consider an evaluation to see if a school's fifth graders are on pace to meet or exceed the proficiency designation in 8th grade. By using the data from the most recent 8th grade completing cohort, models can be developed which will allow projections for the current 5th graders as to their likely 8th grade scores assuming the same future schooling experience as the current 8th grade cohort received. Thus, the projected score enables an evaluation to see if each student is likely to exceed the 8th grade proficient standard. If so, the **current** school receives a positive credit. The percent of projected proficient students is calculated and all of the regular AYP rules are applied to see if this school has made AYP.

What this approach accomplishes is to use all of each student's prior scores—instead of just one score as in the simple case—to give a more reliable measure of the impact of the **current** school on the rate of progress of its student population. In other words, this approach uses multivariate, longitudinal data from students from the current school to provide estimates to map into a future scale without the test error problem. Additionally, this approach avoids the inherent non-linearity problem

¹This is true because the covariance structure among the prior scores is not related to test error. For the Tennessee application, if a student does not have at least three prior scores no projection is made and a student's current determination of proficient or not is included in the percent projected proficient calculation.

of vertically scaled test data in that this approach only requires the assumption of linearity between the prior scores and the future score; an assumption that is easy to verify empirically.

The author raised questions about the reliability of the projected values.

But the use of “projected scores” by Tennessee introduces an additional error factor. The Tennessee growth model substitutes a predicted or expected score for the AMO. Tennessee shows that the correlation of the predicted scores with actual scores is about .80 ($R=.80$). This means the percentage of variance accounted for is only about two-thirds (.8 squared = 64 percent); thus, one-third of the variance in actual scores is not accounted for by the predictive model. While an R of .80 is quite respectable in the research world, it may not be adequate for making real-world decisions. Many students will be counted as being on track when they are not, and vice versa.

The projected scores have much smaller levels of uncertainty than progress measures based upon one prior score. It is true that the projected values in the Tennessee application do not consider future schooling effectiveness and will reduce somewhat the relationship between the projected scores and the observed scores in the future. However, the objective is to give the current school credit, not to hold the current educators' evaluation hostage to what future educators may or may not do! Additionally, it is most important to acknowledge and give Tennessee credit in the fact that all students' projections are used in the determination of percent projected proficiency, not merely those students who did not have the proficiency designation. In other words, students who are currently designated to be proficient but whose projected values fall below the future proficiency designation will count as a negative, providing an incentive to focus on the progress rates of all students and to minimize the focus on just the “bubble kids.”

APPENDIX 1 (continued)

Response to Zeno’s paradox assertion

The author spent considerable energy and space in the paper asserting that the Tennessee projection model is unwittingly trapped in Zeno’s paradox. He asserts that students can make small amounts of progress, yet have their projected scores exceed the future proficiency level. Since the next year the projection targets are reset to another grade, this will allow schools to “get by” with suboptimal growth, resulting in students not obtaining proficiency due to the modeling itself. We dispute the author’s assertion! The author states:

The disadvantage is that Mary’s target will always be “on the way” to proficiency because, under Tennessee’s model, Mary’s goals are recalculated each year. Her target fifth-grade score is one that is estimated to put her on the path to proficiency by eighth grade. Her sixth- through eighth-grade scores are recalculated each year based on a score projected to be on the path to a goal of proficiency by high school.

As was previously stated, the goal is to evaluate the progress made in the **current** school. The totality of Mary’s data provides the estimate of Mary’s future attainment. If Mary has a ‘good’ year, then her projected achievement level goes up. The future distribution that Mary’s projected score maps into has the same proficiency standard as has been approved for regular AYP determination. Additionally, and most importantly to the Zeno paradox argument, if the cut score for 7th and 8th grade is essentially at the same place in the statewide distribution, then it does not matter to which distribution her projected scores are mapped—so *de facto* there is no remapping. This is essentially the case for Tennessee’s 7th and 8th grade Math and Reading/Language Arts proficiency cut scores. The author’s resetting argument has no relevance and Zeno’s paradox does not apply.

Other responses

He further asserts:

As long as each student makes a small amount of progress toward proficiency, the

school could hypothetically make AYP through the growth model even though not a single student had actually gained proficiency. This is not only true in any given year, but also could be true for a period of more than the three years implied.

The idea that a school can have all students making a small amount of progress toward proficiency and yet make AYP with no proficient students is unreasonable. Just because a student makes a little progress, it does not mean that his or her projection will be greater than the target value. A school with all students not proficient would have a large number who are very low in academic attainment. These students would need to have made substantial academic progress to be projected to being proficient within three or fewer years. This would require more than a small amount of progress from each student. If the very low achieving students are projected to proficiency in three years, then their growth trajectories must have changed substantially. The author’s conjecture that only small amounts of growth are necessary to meet projected proficiency is just not accurate.

Another comment:

But relative to the Tennessee growth model, the safe harbor provision has three advantages. First, it is based on real data, rather than a projection. Second, it is based on students achieving a set, policy-driven target—proficient—rather than a moving, amorphous, and norm-referenced target (i.e., a projected score), which has many more variables.

This whole passage is misleading. As was previously mentioned, percent proficiency calculations, as used in safe harbor, are estimates based on test scores with errors. The projections are estimates based upon much more data and contain more information than is in the one test score for the safe harbor estimates. The statement, “*rather than a moving, amorphous, and norm-referenced target,*” is totally wrong. There is not a norm-referenced target: the projections are to established proficiency cut scores for future grades.

APPENDIX 1 (continued)

The author further states:

The Tennessee growth model also means that schools in Tennessee will be able to make AYP in 2014 without all students being proficient.

This statement could be applied to all growth models, safe harbor, and the present AYP status model as well if test errors are taken into account. To single out the Tennessee model with such a declarative statement without careful consideration of all of the uncertainties around the estimates from other models is inappropriate. As was stated earlier, many of the opinions expressed in this paper ignore the innate test errors in one set of AYP approaches yet attempt to magnify uncertainties in the Tennessee projection model. In fact, this is an area in which the projection model has advantage over the others.

The following is a statement that is clearly and provably wrong:

Schools will be more likely to make AYP under the projection model than the other two models.

In Tennessee, many more schools make AYP with the status approach than either with safe harbor or projections, and many more schools make AYP through safe harbor than do through the growth model.

Not in the current paper, but in a recent blog by the author the author stated:

Some of the methodology and most of the data are proprietary, meaning they are privately owned, i.e., no public access. This all makes it very difficult for even Ph.D. and J.D.-level policy experts to get a handle on what is going on (which I found as a peer-reviewer last year), let alone teachers, parents, and the general public.

Exact descriptions of the methodology deployed in the Tennessee projection calculations have been published in the open literature since 2005.² Additionally, the proposals to utilize this methodology have been reviewed and

approved by four different peer review teams assigned by the USDE. Also, at the request of Congress, the early approved proposals were reviewed by a team from the GAO. In that review, the software for the Tennessee calculations was reviewed and evaluated to give an independent evaluation as to computational accuracy.

Agreements with the author

We agree with some of the author's comments.

The use of growth models represents an opportunity to improve upon the state accountability systems currently in use under NCLB. NCLB's focus on a single criterion, proficiency, and its lack of focus on continuous academic progress short of proficiency, fails to recognize schools that may be making significant gains in student achievement and may encourage so-called "educational triage."

The model does offer some advantages. By setting goals short of, but on a statistically projected path to, proficiency, the model may provide an incentive to focus efforts—at least in the short-term—on a wider range of students, including both those close to and those farther away from the proficiency benchmark. It also may more fairly credit, again in the short-term, those schools and districts that are making significant progress that would not be reflected in the percentage of students who have met or exceeded the proficiency benchmark.

We also agree with the author that Congress and the Secretary should review and learn from the growth models which have been approved. After working with longitudinal student achievement data generally for nearly three decades and working with various models to be used as augmentations for AYP specifically, I have

²Wright, Sanders, and Rivers (2005, "Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards", in R. W. Lissitz (ed.) *Longitudinal and Value Added Modeling of Student Performance*, Maple Grove, MN, JAM Press).

APPENDIX 1 (continued)

formed some opinions that I hope are worthy of serious consideration:

- Simplicity of calculation, under the banner of transparency, is a poor trade-off for reliability of information. Some of the more simplistic growth models sweep under the rug some serious non-trivial scaling, reliability and bias issues. The approved models for Tennessee, Ohio and Pennsylvania represent a major step in eliminating some of these problems.
- Reauthorization of NCLB should put more focus on the academic progress rates of all students, not merely the lowest achieving students. Our research has shown for years that some of the students with the greatest inequitable academic opportunities are the early average and above average students in schools with high concentrations of poor and minority students.

Too many of these students are meeting the proficiency standards, yet their academic attainment is sliding.

- Serious consideration should be given to setting future academic standards to various attainment levels. For instance, for Tennessee we provide projections to proficiency levels (regular and advanced), to minimal high school graduation requirements, to levels necessary for a student to avoid being vulnerable to taking a college remedial course, and to levels required to be competitive in various college majors. Some or all of these could be included in an AYP reauthorization with some careful thought. States which presently have these capabilities should be encouraged to move forward. Moving to these concepts will tend to avoid the conflict over what cut score the word 'proficiency' should be attached.

APPENDIX 2

RESPONSE TO THE COMMENTS OF DR. WILLIAM SANDERS RE: ARE WE THERE YET? WHAT POLICYMAKERS CAN LEARN FROM TENNESSEE'S GROWTH MODEL

By Charles Barone

First, I appreciate Dr. William Sanders taking the time to converse about the “Are We There Yet?” paper.

The Tennessee growth model, like those of the other 14 states in which growth models are in use, is a pilot program being conducted through time-limited waivers of federal statutory requirements. The purpose is to try something out, learn from it, and use the results to inform future policy efforts. This was the very reason I wrote “AWTY?” and why Education Sector published it.

I actually think that the paper addresses all the points raised in Sanders’ response, and here, for the sake of brevity, I will focus only on the key points. In most, though not all cases, it is, in my opinion, a matter of emphasis rather than real difference.

The Fallacy of “Failing Schools.” There are a couple of points raised in the opening paragraph that I will address later, but there is an overarching point that I think is implicit in this debate about NCLB in general and AYP in particular that I want to bring into the open.

In talking about a school that was doing well “normatively” i.e., relative to other schools (in terms of percentile ranks) at some grade levels, Sanders states:

Relative to NCLB, this was a failing school. In our view, this was not a failing school; rather it was an outstanding school and should not suffer the indignity of the failing school label.

Nothing in NCLB labels a school as “failing.” Why this is a common misperception (and why Sanders has bought into it) is a topic for another discussion, but it’s indisputable that many perceive the law as ascribing this label to schools “in need of improvement.” It seems

to me that the school Sanders cites was likely neither failing nor “outstanding” but somewhere within the wide gulf between those two poles. The whole point of growth models, I thought, was to calibrate the differences between extremes, not to throw schools into one of two “either-or” (or dichotomous) categories.

The real purpose of federal law is to identify areas where students are falling short—by grade and subject area—and to direct resources to them early and as intensively as is necessary and appropriate. Doing otherwise is a state and local choice and, I would add, a misguided one.

Those involved in creating the NCLB law felt that, **prior to enactment of the law in 2002**, schools were able to hide behind average across groups and, it appears in Tennessee, across grade levels i.e., **were used in a way that obscured areas in need of improvement rather than illuminated them.** Average elementary school scores can hide deficiencies in third grade that would be better to address early. Average scores of all students can hide gaps between black and Latino students and their non-minority peers. Composites across subjects can hide subject area-specific shortcomings.

By bringing those problems to light, and funneling resources to those areas as early and as surgically (or radically) as needed and as is possible, it is hoped that students will get a better education and that potential long-term problems will be addressed sooner rather than latter.

Hence, in the paper we make this point:

The Tennessee growth model will also reduce the number of schools identified by NCLB

APPENDIX 2 (continued)

as falling short academically. This could be a positive change if it allows the state to focus more intensely on the lowest-performing schools. However, it will also mean that some schools may not be able to avail themselves of resources that could help address student learning problems early enough to prevent future academic failure.

It sounds like what we had in Tennessee was a labeling problem—calling all schools that did not make AYP “failing” rather than an AYP problem per se. I think most educators seeing a third grade with scores in the 26th percentile statewide (with one of the lowest set of standards in the nation) would want to address that problem promptly in the antecedent years (i.e., by improving what happens in pre-K, kindergarten, first, and second grade) rather than waiting two years to see what happens in fifth grade. Other states have gradations of not making AYP and target their interventions accordingly (such as at one grade level or in one subject) including interventions at grades prior to the grades in which testing begins. The law offers wide leeway to do so.

The third grade case cited by Sanders is particularly in need of attention, as stated in the “AWTY?” paper:

Slowing down the pace at which students are expected to learn academic skills in elementary school may create long-term problems for students and create larger and more difficult burdens for public education in junior high, high school, and beyond. A wide body of research suggests, for example, that children who do not acquire language skills in the early grades have an increasingly difficult time catching up to their peers as they progress. This parallels neuropsychological research that shows critical periods for brain development in language and other areas of cognitive functioning.

Statistical Error. Sanders states that Tennessee rejected looking at non-statistically-derived scores (i.e., hard targets, rather than estimates) in part:

Because (1) the error of measurement in any one score for any one student is so relatively large that the setting of an improvement target by this method will inevitably send the wrong signals for many students.

Here, as at other points in the paper, Sanders seems to assert that the projected score model gets rid of measurement error. It doesn't. **Measurement error is inherent in any test score** (as in any grading system). Sanders' method uses the same tests as every other AYP model in use in Tennessee and the other 49 states.

What the projected score model does is introduce an additional source of error, “prediction” error (the difference between a projected score that a multiple regression analysis *estimates* will put a student on the path to proficiency and the actual score that *would* do so).

This is pointed out in the paper, but unaddressed in Sanders' comments:

...the use of “projected scores” by Tennessee introduces an additional error factor. The Tennessee growth model substitutes a predicted or expected score for the AMO. Tennessee shows that the correlation of the predicted scores with actual scores is about .80 ($R=.80$). This means the percentage of variance accounted for is only about two-thirds (.8 squared = 64 percent); thus, one-third of the variance in actual scores is not accounted for by the predictive model. While an R of .80 is quite respectable in the research world, it may not be adequate for making real-world decisions. Many students will be counted as being on track when they are not, and vice versa.

Sanders goes on to state that:

To minimize the error of measurement problem associated with one test score, we elected to use the entire observational vector of all prior scores for each student. In some of our

APPENDIX 2 (continued)

earlier work, we had found that if at least three prior scores are used, then the error of measurement problem is dampened to be no longer of concern.

But what he does not mention is that current law allows this option (using “rolling three year” averages of scores) whether or not a projected model is used.

Zeno’s Paradox Issue. The “AWTY?” paper concludes that many students under the Tennessee model will take longer than three years to reach proficiency even if they meet their minimum “projected” score three years in a row.

Sanders states, through reasoning I could not quite follow:

The author’s resetting argument has no relevance and Zeno’s paradox does not apply.

I stand by the conclusions of the paper. I challenge Sanders, or anyone else for that matter, to show me an instance where:

- 1) there is a long-term goal (e.g., X distance in Y years)
- 2) there is an interim goal that is some fraction of X progress for some fraction of Y years and;
- 3) the interim goals are re-calculated each year for a fraction of the remaining distance to Y;

in which it doesn’t take longer than Y years to get there.

Sanders could of course clear all of this up by taking, say, 100 cases where we can see the projected scores for each student, each year, and where the student exactly makes each interim goal, to show us what happens in Tennessee in this instance over three successive years. As the paper shows, however, since the data and exact methods are proprietary, none of us can do this on our own, or we would have simulated such an instance in the paper.

On this point, Sanders states:

Exact descriptions of the methodology deployed in the Tennessee projection calculations have been published in the open literature since 2005. Additionally, the proposals to utilize this methodology have been reviewed and approved by four different peer review teams assigned by the USDE.

It is true that the multiple regression formula that Sanders uses can be found in Tennessee’s application for the federal waiver, as well as in most elementary statistics books. Tennessee’s materials also include descriptions of some of the methods and adjustments that are specific to the Tennessee growth model.

But the details—including standard deviations and standard errors of measurement of students within in a school, and the histories of individual students over multiple years—are not available. Thus no one, at least no one that I have talked to, can do a real replication.

In addition, I sat on a growth model peer review panel in 2008 in which other states submitted models based on that of Tennessee. Not a single person at the Department with whom I inquired understood that in 2013, the goal for Tennessee was still proficiency in 2016, not 2014, and I think any casual observer of former Secretary Spellings’ comments over the last few years can attest to that.

Size of Adequate Yearly Progress. Sanders disputes the paper’s contention about the interaction between the growth model’s incremental progress (extending the proficiency target over a number of years rather than proficiency each year) and Tennessee’s low standards. But he merely skirts over the latter point.

First, I don’t understand the artificial focus on those grades in which testing is done. If proficiency within three years is a doable goal, why not start with proficiency in fourth grade as a goal beginning in first grade (or kindergarten or pre-K) where research shows schools (and programs like Head Start or high-quality child care) can have an incredible impact? The state and its localities have all these resources at their disposal to

APPENDIX 2 (continued)

impact the years within and outside the 3-8 testing box imposed by NCLB. Why not do so? Is more federally imposed standardized testing, in earlier grades, what is required to bring this about? (I, for one, hope not.)

Second, no matter what grade you begin in, the Tennessee standard for proficiency is low compared to the NAEP standard—lower than virtually any other state.

Again, let me re-state a section of the paper which Sanders does not address:

In fourth-grade reading, for example, the NAEP benchmark for “basic” in reading is a score of 243; for “proficient” it is 281. The NAEP-equivalent score of the Tennessee standard for fourth-grade proficiency in reading is 222, which is about as far below the NAEP standard of basic as the NAEP standard for basic is below the NAEP standard of proficient. The NAEP benchmark for basic in fourth-grade math is 214; for proficient, it is 249. The NAEP equivalent of Tennessee’s standard for proficient in math is 200.

So if reaching proficiency in Tennessee is a low goal relative to other states (which the state of Tennessee acknowledges and is trying to change) relative to NAEP, then fractional progress toward that goal is, by definition, even lower.

How could it possibly be otherwise?

Linearity. Sanders asserts, regarding the Tennessee model, that:

This approach avoids the inherent non-linearity problem of vertically scaled test data in that this approach only requires the assumption of linearity between the prior scores and the future score; an assumption that is easy to verify empirically.

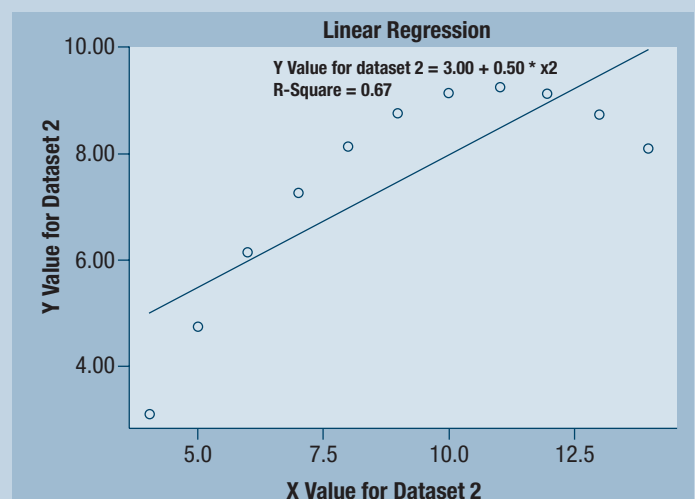
I chose not to go into this in the paper (for obvious reasons), but since the issue is being opened here, I think it should be addressed.

Linearity is a double-edged sword (stay with me until at least the chart on the next page). With vertical scaling, different tests can be equated across grades by re-scaling scores to make them comparable. We can’t go into all the relative advantages and disadvantages of vertical scaling here. (Sanders is right that there are disadvantages.)

But I must point out that Sanders’ assertion of the linearity of non-vertically scaled scores in Tennessee—which he says are easy to verify empirically—may not always be true. (Note that Sanders does not supply empirical verification but only asserts it can be verified.) In turn, applying a linear regression, as Tennessee does, to estimate future scores may distort the relationship between real growth in student scores and those scores projected through the statistical model.

Let’s say that over time, non-vertically scaled scores for some students are not linear but are parabolic (curvilinear) with accelerated growth in early years and a leveling off, and then a decrease in later years (a phenomenon not unknown in education research). Then let’s say we try to map a linear regression onto this model (with an R squared of .67, similar to the Tennessee model with an R squared of .64).

The chart below illustrates this scenario.



Source: From SPSS Textbook Examples, Applied Regression Analysis, by John Fox, Chapter 3: Examining Data. UCLA: Academic Technology Services.

APPENDIX 2 (continued)

Here, the projected scores in the early years would be lower than the actual scores that would be seen over time. In this scenario, the linear model would set AYP goals below that which we should expect for students between ages 6 and 11. Conversely, the model would overestimate what we should expect for students over age 11.

This is just one of the many (virtually infinite) scenarios possible depending on student characteristics, age, and patterns of variance of scores for students in a particular school. The point is a linear regression only approximates, and in some cases can distort, educational reality.

Transparency. In closing, I would like to address the issue of transparency. In his remarks, Sanders says:

Simplicity of calculation, under the banner of transparency, is a poor trade-off for reliability of information. Some of the more simplistic growth models sweep under the rug some serious non-trivial scaling, reliability and bias issues. The approved models for Tennessee, Ohio and Pennsylvania represent a major step in eliminating some of these problems.

This paper only speaks to Tennessee, and so we will leave the issue of other states aside.

But, as the paper shows, and as demonstrated here, the Tennessee growth model is not necessarily more reliable, accurate, or valid than those of other states using other growth models or the statutory “status” or “safe harbor” models. All represent tradeoffs.

While eliminating some problems, the Tennessee model creates others. For now, each state can reach its own conclusions about the relative strengths and weaknesses, and it was the hope that the “AWTY?” paper, and this discussion, will help better inform those decisions.

I do not, however, think transparency is an issue to be taken lightly. Real accountability only takes place when

all participants in the education system—including parents, advocates, and teachers—can make informed choices.

I talked to a reporter from Texas this week (which is implementing an adapted form of the Sanders model, with at least a couple of key improvements per points raised here) who recalled her school days of independent reading assignments through the “SRA” method.

For those of you who do not remember, SRA was a box of large (roughly 8 x 11) cards, with readings and structured questions. The box progressed in difficulty from front to back (easiest to most difficult) with color-codings for varying levels of difficulty.

What the color-coding did was make understandable where you were—for yourself and the teacher—in progressing through a set of skills. The reporter pointed out that with the traditional method you would know, for example, that if you were at say red (the lowest) rather than violet (the highest) you knew you were farther back than you wanted to be by a certain time. Depending on the assigned color of where you were at (say red or orange) you also knew where you were *relative to the end goal*.

She then pointed out that with the Tennessee growth model method, we never know what the target color (or level of difficulty)—i.e., the interim “projected” score for a student by the end of the school year—is. It could be any color of the rainbow from red (below basic) to violet (proficient), and all we would know is that it was *somewhere* in between.

I think that all players in the educational policy and practice arena—educators, consumers, parents, advocates, and taxpayers—want, just as this reporter does, something a little more like the color-coded SRA system.² That is, they would like quite a bit more clarity than “trust us, your child is on a projected path to proficiency” within Y years (which, as we see here, is really Y + unknown # of years) according to the following formula:

$$\begin{aligned} \text{Projected Score} &= MY + b_1(X_1 - M_1) + b_2(X_2 - M_2) + \dots \\ &= MY + \sum x_i b^i \end{aligned}$$

APPENDIX 2 (continued)

And, as much as I love statistics, I would assert that given that these individuals—educators, consumers, parents, advocates, and taxpayers—are the primary sponsors *and* intended beneficiaries of the educational system, we owe it to them to strive, as much as humanly possible, to meet their needs and expectations toward the goal of a better education for each and every child.

Endnotes

- ¹ SRA is now the “Open Court” reading system. Its citing here by the author does not represent or imply any appraisal or endorsement.
- ² Where MY, M1, etc. are estimated mean scores for the response variable (Y) and the predictor variables (Xs).