

**Testimony to Georgia House's
Federal Government's Role in Education Study Committee**

Regarding:

Common Core State Standards and Related Testing Issues

Richard G. Innes

Staff Education Analyst

Bluegrass Institute for Public Policy Solutions (www.bipps.org), Kentucky

Mr. Chairman, members of the committee, my name is Richard Innes, and I am the Staff Education Analyst at the Bluegrass Institute for Public Policy Solutions. The Bluegrass Institute is a Kentucky public policy think tank similar to your Georgia Public Policy Foundation and the Georgia Center for Opportunity. In fact, both of those Georgia organizations are our State Policy Network partners in your state.

Who is Richard Innes and Why is He Addressing Us?

To begin, I probably need to share a little information about who I am.

I do have experience in instruction and some of it is very out-of-the-box, though it is not from the K to 12 arena.

In fact, after joining the United States Air Force and learning to fly high performance jets at Moody AFB in Valdosta, I spent nearly half of my military career as an instructor pilot in both training and heavy transport aircraft.

By the way, much as is done with our public school teachers, Air Force instructors go through a carefully developed, formal training program in how to teach as well as what to teach.

Also pertinent to the committee's interests, while serving as an instructor pilot at Columbus AFB in Mississippi in 1971, I became one of the very first pilots trained to program the very first generation of what I'll call automated teaching machines that were ever used operationally in Air Force pilot training. This "Learning Center" equipment marked a huge advance in instructional technology, a forerunner of what we refer to as digital learning today. In preparation for this new assignment, I received additional training in a number of areas such as the development of educational standards.

So, I understand something about how real education standards must be put together to create a really outstanding educational system.

There are other reasons why I was approached to talk to you which relate to Kentucky's very aggressive attempts at education reform over the past 24 years. For example:

- Kentucky has a very extensive history with so-called higher order thinking skills types of assessments, dating way back to 1992. This experience relates to your new assessment plans in Georgia. In fact, a briefing package from your Georgia Department of Education regarding your new Georgia Milestones Assessment System specifically points to Kentucky testing products as examples of what to expect on your new tests.
- Kentucky was the very first state to adopt the Common Core, doing so in February, 2010, about four months before the final version of the Common Core was even released. My state has the most experience of any state to date with these standards.
- Kentucky also was the first state to launch Common Core aligned testing systems in reading and mathematics, doing so in the spring of 2012, a year ahead of New York, which was the next state to adopt Common Core aligned testing. Kentucky has now conducted three years of Common Core aligned testing in reading and mathematics, although the year three data still is not available – a topic I will further address in a moment.

Thus, Kentucky already has experience with Common Core and even more experience with advanced testing programs. Georgia would be wise to consider the lessons to be learned from that experience. If you research Kentucky, you will be better prepared for what is likely to come. You also might avoid some pitfalls that cost Kentucky lots of time, money and aggravation.

Now, let's talk about education standards in general, the Common Core State Standards in particular, and how those standards will drive your new Georgia Milestones Assessment System. However, due to the title of your committee, let's first talk briefly about some the federal involvement issues that have recently surfaced in my state.

Federal Involvement in State Education Business

Action from US Department of Education, often referred to by educators as USED, has lately become a very hot topic in Kentucky.

As in other states, Kentucky has been pushed rather hard over the past few years by federal programs like the Race to the Top sweepstakes, where a few winners took virtually all the money at a time when all states were facing severe financial challenges. Kentucky was also heavily pushed by the more recent use of waivers from the No Child Left Behind Act. Those waivers are being used as a means to push federal desires into the state education programs across the country.

Apparently, USED pressure finally reached the boiling point with Kentucky's Commissioner of Education Terry Holliday recently when he posted several blogs (<http://kyedcommissioner.blogspot.com/>) about the increasingly objectionable nature of these intrusions.

Dr. Holliday's first criticism appeared in his August 15, 2014 blog, "**The good news and bad news on NCLB waivers.**" Dr. Holliday charges in this blog:

“While the initial waiver process was something we supported in Kentucky, it has become problematic. When the state chiefs talked with Sec. Duncan about what would happen at the end of the initial waiver period, we recommended a “streamlined and expedited” process for one-year extensions.”

Dr. Holliday’s blog continues:

“There is significant evidence from many states that the waiver extension process has not been streamlined,” and he also remarked, ***“Nor has the waiver extension process been expedited, as we were promised. We submitted our extension request May 1 and it was mid-August before we got word on its status.”*** Dr. Holliday shockingly points out, ***“Our initial waiver took less time to approve.”***

Dr. Holliday got considerably more aggravated in his second blog about this issue, which appeared on August 22, 2014. He got right to the point with the blog’s title, **“USED action contrary to state, federal law.”**

Among other complaints, Dr. Holliday angrily pointed to USED forcing Kentucky to use an out-of-date science test this coming spring even though the state’s classrooms have already switched over to teaching the new Next Generation Science Standards. These new standards reportedly will require rather complex assessments and there isn’t enough time to create those new science tests in an orderly manner for use this coming spring.

Continuing this saga, Dr. Holliday recently turned his aggravation into action with a trip to discuss problems with US Secretary of Education Arne Duncan on September 12, 2014. Whether or not any truly fruitful outcomes will result probably will not become evident for some time, however.

Furthermore, salt has already been rubbed in Holliday’s must-use-old-science-test wounds. Kentuckians recently learned that just north of our border in Indiana, following that state’s rejection of the Common Core and adoption of new, state-developed standards, USED demanded that the Hoosiers had to rapidly develop new tests aligned to the new curriculum and must use those new tests this coming spring. So, while Kentucky is being forced to use old tests not aligned to the state’s new science curriculum, the exact opposite testing demand is being forced on our northern neighbor. Imagine that.

Certainly, given this backdrop, the formation of your committee is most appropriate, and I sincerely hope your state will soon add to those like Kentucky and Louisiana where complaints are starting to turn to actions.

Now, let’s shift focus and talk about some key background data your committee needs to understand to better perform its primary investigations. I am going to discuss several things:

- What a good, standards based education program looks like,
- How the types of testing we are told Common Core requires already have an extensive track record in Kentucky, and what your committee should know and has probably not been told about that history.

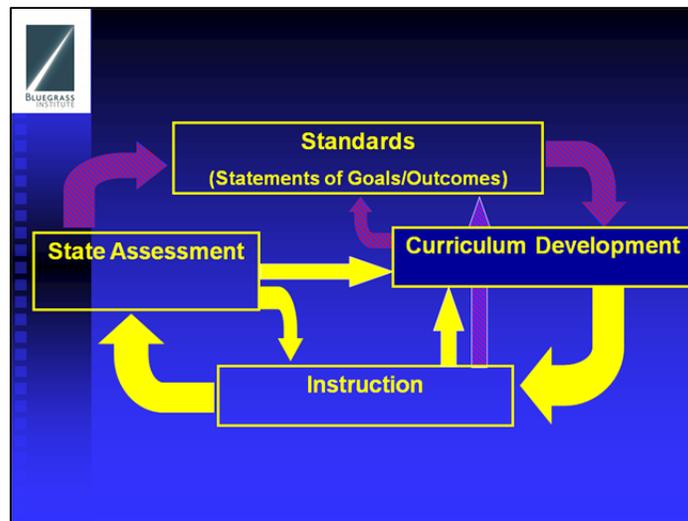
What Does a Good Standards-based Education Program Look Like?

Years ago the US Air Force taught me about how a high quality, standards-based education system really needs to operate. Pilot training, after all, is very much about standards, and the penalty for a low standards program can be steep indeed.

In a high performance education system, which I graph out for you in Figure 1, the education standards element, which includes the goals and outcomes, are at the absolute center of the program. The standards are not an add-on, not something peripheral, not only vaguely related to everything else. The standards actually are the axel around which everything else will revolve.

Figure 1

High Performance, Standards-Based Education System Relationship Chart



I learned that real education standards, to be effective, need to be detailed and specific so everyone in the education program really is on the same page. The standards also need to define in measurable terms exactly what level of performance is good enough, a feature lacking in the Common Core and Next Generation Science Standards. Without measurability information that tells us what level of performance is good enough, everyone else in the education process – and most particularly the test writers – are left to founder on their own about what proficiency really means. Such confusion can drag down the entire process.

Once the standards are created, it's time to write the curriculum.

Please note, I didn't say tests; I said curriculum writing is next after standards creation. There is a reason for this. In the process of creating the curriculum, it is inevitable that questions about the interpretation of the standards will arise. When that happens, curriculum writers need to be able to talk to experts on the standards creation team to iron out the confusion. In some cases, it may turn out the standards need more work. It is best if those standards adjustments occur as much as possible before the education system hits the classroom. Those adjustments cannot happen if there is no feedback.

After the curriculum is in hand, it is time to try it out in the classroom. In this step, teachers interacting with real students may run into difficulties. So, teachers need to be able to talk with both the curriculum writers and the standards writers because those classroom implementation problems could be caused by either standards or curriculum issues. Once again, a good standards-based education system requires feedback loops.

With some classroom experience now in hand, we are finally ready to look at test creation. Again, the people writing the tests need to communicate with all the other education elements shown in Figure 1. Test writers may surface additional problems related to the standards, or the curriculum or the teaching in this system. Thus, still more feedback pathways are required as testing personnel create the new tests.

Even more feedback is needed when the test scores come out. Did the tests accurately capture student performance? Why did students perform this way? Do the results highlight problems with standards, curriculum or teaching? All the players need to be communicating with one another.

By the way, I asked Dr. Sandra Stotsky, former associate commissioner of education in Massachusetts, about the process outlined in Figure 1. She said that this was indeed pretty much the way her state operated back when that state was producing top math and reading performances on the National Assessment of Educational Progress. So, the process outlined in Figure 1 isn't just theory. It works.

Let me reemphasize a key point: feedback is absolutely crucial in the real, high performance education world. That includes availability of continuous and active feedback loops with the standards creation team.

This need for feedback is where I see a major flaw in the Common Core State Standards that will underlie your Milestones Assessments. After the Common Core was handed down in June of 2010, the creation teams folded their tents and departed the scene. There is no central support team in place to service these non-living standards. That is why all the feedback arrows to the standards element in Figure 1 are shadowed. They simply don't exist with Common Core. That's a huge – perhaps fatal – error that Georgia must carefully consider.

Perhaps in part due to this lack of feedback support, some states like Indiana and Oklahoma have departed the Common Core camp. Other states like South and North Carolina are launching reviews, as well. Already, thanks to the lack of feedback to the standards creation process, these supposed national standards are becoming less common.

Very recently, Kentucky also launched its own, state-level review and comment process for the Common Core. If Kentucky can work around the copyright on the Common Core State Standards and those pesky restrictions from the federal government, any changes Kentucky makes will make the Core still less “common.”

Reviews of Common Core like the one Kentucky just launched actually should be going on at the national level. The very fact that there isn't even a mechanism for such feedback and review to occur is

evidence of a massive failure of responsibility on the part of those who created Common Core. You cannot just create a standards document and walk away. But, that is exactly what has happened with Common Core State Standards.

This leaves states like Kentucky and Georgia out on their own. Sooner or later, just as is happening right now in Kentucky, Georgia will have to come to grips with problems in Common Core that won't really become apparent until you attempt to create the curriculum, the classroom experience, and the tests that are intimately related to those standards. Absent someone taking up the responsibility again in Washington, the Georgia Core will sooner or later become less common with other states. You will be better off to recognize that sooner rather than later. Rigid adherence to non-living education standards is a certain recipe for disaster.

History of Kentucky Assessments and Issues

Kentucky started to use large numbers of open-response and extended-response types of questions in its state accountability system way back in 1992 with the launching of the Kentucky Instructional Results Information System, or KIRIS. The Bluegrass State is closing in on nearly a quarter of a century of experience with these types of tests, and Georgia will ignore some of our lessons learned at great peril for your students and great expense for your taxpayers.

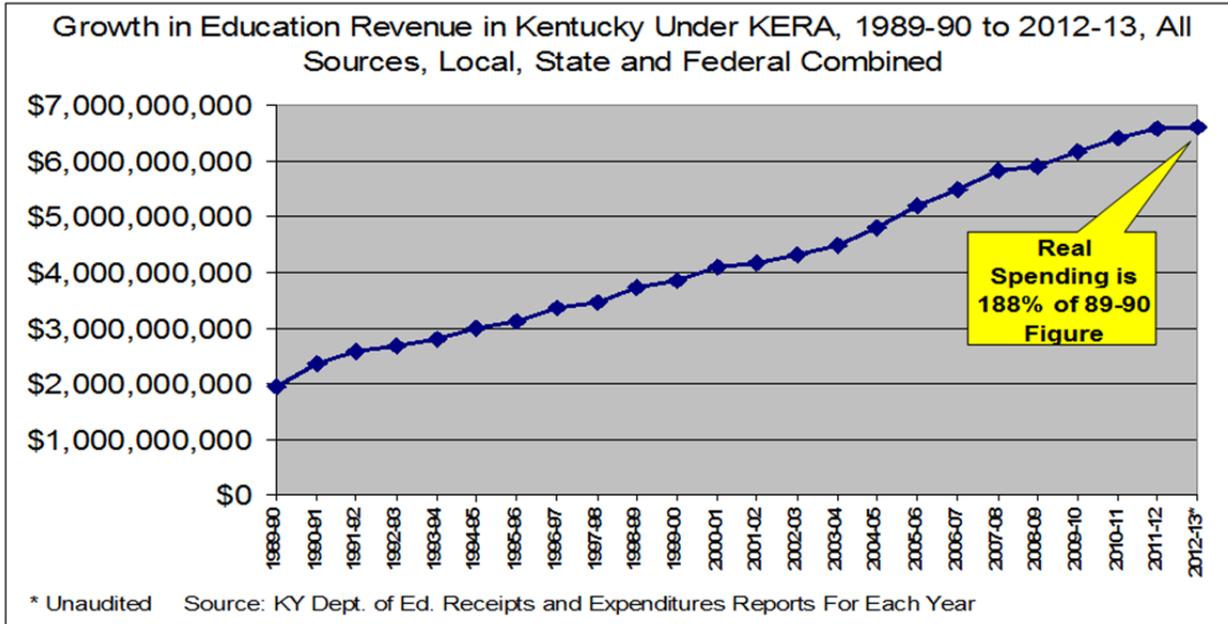
The Costs

With a standards-driven system, you are not just talking about increased costs for tests. You are adopting a new system and, a system like the one Kentucky adopted in 1990 isn't cheap. This new system requires significant investments in:

- Professional Development,
- Textbooks,
- Curriculum,
- Digital Technology, as well as
- Assessments.

In Kentucky the resulting increases expenditures over the years have been dramatic, as the graphic in Figure 2 shows. The line on this graph traces Kentucky's total education revenue figures in unadjusted dollars from the year prior to adoption of the state's radical education reform in 1990 to the most recent year of data for the 2012-13 school term. I also ran a calculation of the increase in spending after adjusting for inflation, as shown by the flag in Figure 2.

Figure 2



The cost increase as Kentucky chased highly Progressive education reforms, even if we correct for inflation, nearly doubled over the past 23 years.

You might find it interesting to compare Kentucky’s increases in education spending to those Georgia experienced in the same time interval. Those are shown in Table 1. From the year prior to Kentucky adopting its education reform to the most recent data available from the US Census Bureau, while Georgia’s per pupil spending rose 241 percent, Kentucky’s was soaring by 335 percent.

Table 1

State and Local Government Expenditures for Education for Georgia and Kentucky, 1988-89 and FY 2012 Compared

	1988-89 Total Expenditures	Enrollment 1989	1989 Per Pupil Expenditures (Calculated)	FY 2012 Total Expenditures	Enrollment 2012	2012 Per Pupil Expenditures (Calculated)	2012 PP Expenditures as Percent of 1989 PP Expenditures
Georgia	\$4,698,246,000	1,107,994	\$4,240	\$17,387,814,000	1,699,156	\$10,233	241%
Kentucky	\$2,076,138,000	637,627	\$3,256	\$7,442,661,000	681,827	\$10,916	335%

Data Sources: Public Education Finances 1989, US Census Bureau, Tables 11, 22
Public Education Finances 2012, US Census Bureau, Tables 1, 19.

Georgia, I guarantee this isn’t going to be cheap.

What Were the Impacts of Higher-Order, Authentic, Performance-Based Assessments in Kentucky?

Regardless of actual quality, assessments that include a lot of student writing often are called “Higher-Order” and “Performance-Based.” Supposedly, relying on such written answer formats makes these tests more “Authentic.” The new Milestones Assessments will feature increased reliance on such a format. Thus, because Kentucky has used such testing formats for many years, it is important for your committee to review the results of Kentucky’s experience.

How did such testing perform in Kentucky?

I’m not a great believer in reinventing well-constructed wheels, so to answer that question I will quote from a new report titled “*The Revenge of K-12*” by Richard Phelps and James Milgram (on line here: <http://pioneerinstitute.org/?wpdmdl=424&>). Writing about such advanced testing ideas as the Kentucky Instructional Results Information System (KIRIS) test from the 1990s, the authors say:

“Those testing programs failed because of unreliable scores; volatile test score trends; secrecy of items and forms; an absence of individual scores in some cases; individuals being judged on group work in some cases; large expenditures of time; inconsistent (and some improper) test preparation procedures from school to school; inconsistent grading on open-ended response test items; long delays between administration and release of scores; little feedback for students; and no substantial evidence after several years that education had improved. As one should expect, instruction had changed as test proponents desired, but without empirical gains or perceived improvement in student achievement. Parents, politicians, and measurement professionals alike overwhelmingly rejected these dysfunctional tests.”

That remarkable paragraph largely says it all, but let me expand on a few of the comments that pertain to Georgia’s somewhat less radical Milestones testing model.

Scoring Open-Response and Extended-Response Written Questions Is Always Problematic

For starters, if the answers are hand-written, there is an immediate problem with the quality of the handwriting getting in the way with what is actually written. This is particularly problematic with younger children, but some older students produce difficult to decipher scrawl, too, especially under the pressure of a testing environment.

If you shift to using computers to collect your written answers, then keyboarding skills replace handwriting problems as a confounding variable in determining what the student actually is trying to say.

If you also decide to let computers score the answers, you are sailing into largely uncharted waters. After watching the grammar, spelling and punctuation correction features of Microsoft Word in action, and after considering that the sheer volume of users of that powerful word processor program provide

Microsoft with a lot of cash to make improvements in the software, I think reservation about the fairness of proposed computer scoring of student work is in order, at least for now.

Another issue is the educational background of affordable human scorers, a concern that grows at increasing student grade levels. Kentucky's overall written answer scoring costs have always been high although at one point it was reported individual scorers were only being paid about \$10 per hour. Results from scorers willing to work for that pay rate might not be terribly good, especially when teachers being judged by the results earn a lot more, but this was the best Kentucky could finance.

I don't think you want a truly gifted child's responses to be graded by scorers earning that wage level. Even in middle school, I have seen an example of a student's science test answer that was well above the heads of the readers who scored the child. The student's rather low grade seemed mostly based on somewhat awkward sentence structure rather than the knowledge of science. In other words, it is quite possible for open-response answers in subjects other than writing to turn into mostly an assessment of creative rather than factual writing skill. That generally does not serve students or their teachers very well.

Since we are talking about grading of written answers, let's not forget another potential pitfall: scorer bias. When it comes to writing, one man's wine can be another's poison. I am sure that fact of life has been played out in the halls of your legislature on more than one occasion.

In the case of testing, take a somewhat controversial reading passage, a child from a conservative home, and a highly liberally oriented scorer, and a mix for scoring inaccuracy mischief is at hand.

Thanks to these and other issues, score validity and reliability from written answer questions is always going to be lower than what you are used to with Georgia's past multiple-choice tests. Use too many written questions, and the judgments you form about schools and students could become both inaccurate and highly erratic from year to year.

Scoring Written Answers Takes Time, Slows Score Turn-Around

It takes a lot longer to administer a written answer test, and that means you cannot ask as many questions. In Kentucky, they tried to work around that by doing something called "Matrixing," where each student only answered a portion of the questions in the total question databank. I strongly advise you against allowing matrixing. It guarantees non-comparable scores from student to student because each student takes a different test. Matrixing will take the "common" right out of any assessment, even one developed around the Common Core State Standards.

It also takes time to score written answer questions, and that can really slow the reporting of results to schools and the public. In general, Kentucky's schools have not received accurate, proofed scores until well after the start of the succeeding school term. In fact, the public scores release from Kentucky's spring 2014 testing cycle, which were promised for September 2014, are now delayed further to early

October, at the very best. That delay creates problems because it is very difficult for schools and teachers to alter curriculum on the fly during an active school term while teachers are heavily engaged in actual classroom activities.

I note Georgia already is planning for some delays. A Georgia Department of Education Power Point briefing on “Assessment Update: Georgia’s Changing Assessment Landscape,” says score turnaround is going to be delayed in the first year of the Milestones Assessment program. I’d be ready for more delays after that.

Even Multiple-Choice in the Higher-Order Testing Realm Can Be Problematic

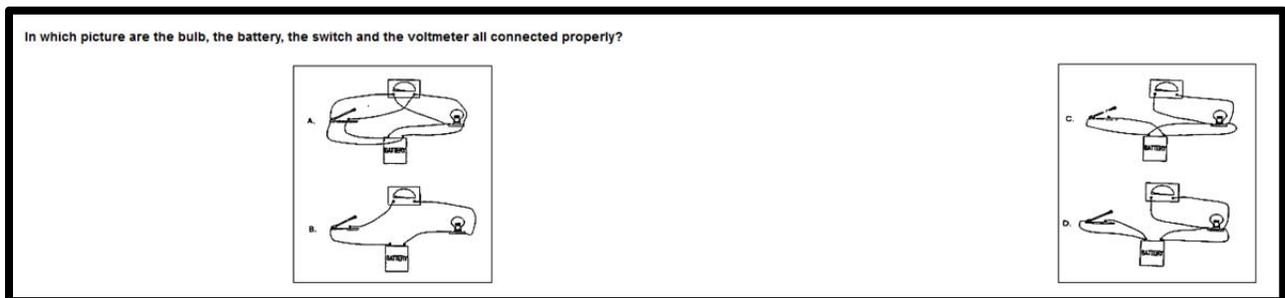
Figure 3 shows a question from the first KIRIS 12th Grade Science Test from 1992. I discovered this question in a booklet of released questions in 1994.

This multiple-choice question is what first showed me there were some serious problems with the quality of Kentucky’s new, supposed higher-order thinking skills tests.

Very simply, this question does not include a correct answer.

When I asked the Kentucky Department of Education who wrote this question and what their credentials were, I was advised there was no record of that information. I learned in my Air Force days that this is not the way quality tests are created. Good test creation requires careful record keeping and accountability for test creators. Kentucky’s KIRIS tests clearly didn’t have that.

Figure 3
Question from 1992 KIRIS 12th Grade Science Test in Kentucky



By the way, thanks to the heads up I got from this wiring diagram question, I’ve been looking at education in my state ever since.

Sustainability and Valid Trend Lines

Closely aligned to the previous problems with tests using written answer problems is the issue of maintaining valid trend lines over time.

It is pretty challenging to change out multiple-choice questions in a way that continues to sample the same educational material at the same level of difficulty. However, doing this is vital if valid trend lines are to be maintained.

This challenge gets a lot worse when written answer questions are thrown in the mix. Creating new written-answer questions that cover the same material to the same level of difficulty is really demanding work.

As a caution, if Georgia later decides to dabble in even more advanced assessment formats such as what the Smarter Balanced Assessment Consortium calls “Performance Items” and Kentucky’s KIRIS labeled “Performance Events,” the challenge to maintain accurate trend lines becomes impossible. Kentucky’s Performance Events failed in only four years in large measure due to the trend line issue. Georgia isn’t currently contemplating performance items, but you will soon experience a push to use them for your next science assessments. That push has already started in South Carolina and Kentucky, so be ready.

Score Inflation

Another problem with using subjectively graded questions on a state assessment is maintaining control over grade inflation. Such inflation has been a major issue in Kentucky as Figure 4 shows.

Figure 4

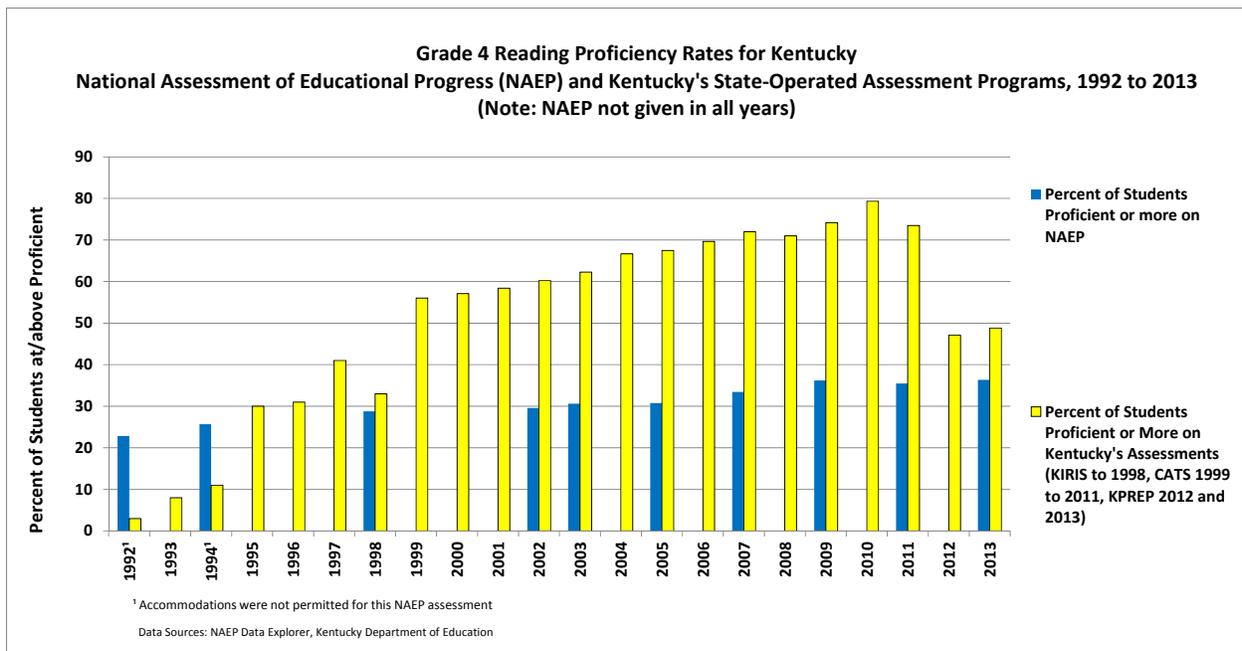


Figure 4 compares the proficiency rates for fourth grade reading over time from the National Assessment of Educational Progress, usually called the NAEP, and Kentucky’s own state assessments.

The NAEP results are shown by the dark bars and Kentucky's own testing results are shown by the yellow bars.

Incredibly, when Kentucky's KIRIS testing began in 1992, it was scored much more rigorously than the already very rigorous NAEP. That was a handy result if you were trying to make a case to justify very radical education ideas, but it clearly was questionable from the standpoint of accurate assessment of the real reading ability of Kentucky's fourth graders.

Then, over time, the KIRIS results exploded while the NAEP fourth grade reading proficiency rate for Kentucky scarcely budged. According to the NAEP, the reading proficiency in Kentucky only rose from 23 to 29 percent between 1992 and 1998. KIRIS, on the other hand, claimed proficiency rates had jumped from just 3 percent in 1992 to 33 percent by 1998.

That inflation was one of the major reasons why KIRIS lost credibility by 1998 and was thrown out after that year.

In 1999, KIRIS was replaced by Kentucky's Commonwealth Accountability Testing System, or CATS. As you can see from the 1999 through 2011 data in Figure 4, the Kentucky Core Content Tests portion of the new CATS bred a lot more inflation.

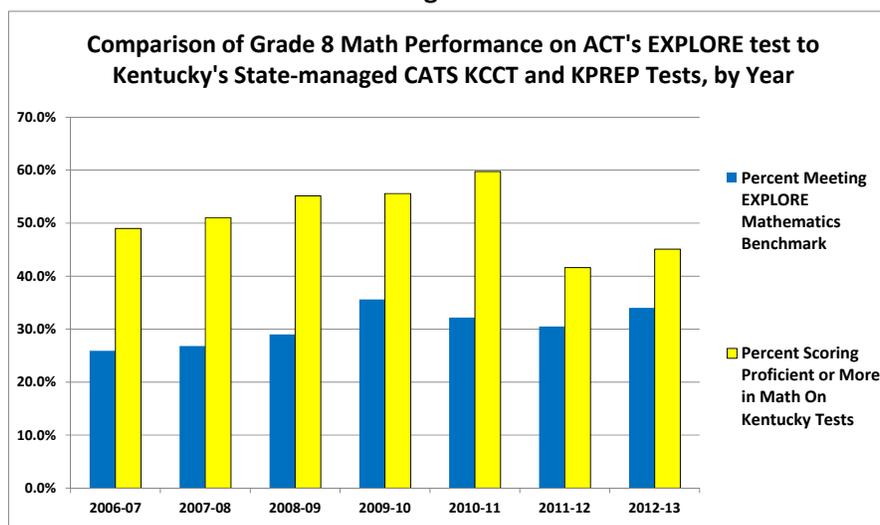
In the initial CATS fourth grade reading results from 1999, Kentuckians were told that the reading proficiency rate had skyrocketed to 56 percent, a figure nearly 30 percentage points higher than the NAEP reported just one year earlier. When the 2002 NAEP reading results were released, the excesses of CATS were confirmed. In that year, NAEP said there was a 30 percent rate of reading proficiency in Kentucky. CATS claimed it was over 60 percent.

The NAEP to CATS KCCT proficiency rate discrepancy for Kentucky's fourth grade readers continued to grow until the last year of the KCCT program in 2011. In that year, the difference between NAEP and CATS reading proficiency rates for Kentucky's fourth graders was over 38 percentage points.

To be clear, largely thanks to the No Child Left Behind Act, many other states also were reporting inflated proficiency rates in the latter part of the time covered in Figure 4. However, the incredible inflation in Kentucky's scores over time is objectionable and provides more evidence that at the very least use of tests that rely on subjectively scored written questions can open the door to serious problems with score inflation due to the great difficulty of maintaining valid trend lines over time posed by these sorts of assessments.

By the way, another, high quality test product from the ACT, Inc. has been used in Kentucky since the 2006-07 school year. That test is the ACT EXPLORE test. You can think of it as an ACT College Entrance Test that is grade-appropriate for these younger students. EXPLORE is linked and equated to the actual ACT college entrance test. Figure 5 compares student proficiency rates from Kentucky's home-grown tests in eighth grade math to the truly college and career readiness rates from EXPLORE.

Figure 5



Once again, the score inflation in Kentucky's supposed higher-order thinking skills tests is apparent. Notice there even is some inflation in Kentucky's new, Common Core aligned KPREP scores for 2011-12 and 2012-13.

Closing recommendations

Using Common Core as the basis for your new education system in Georgia can be problematic. You will probably find that changes to the standards are needed, but as you make Georgia-unique adjustments to Common Core, your education program will obviously become less "common." Due to the total lack of feedback from Common Core's technical experts team, however, I would advise being willing to make those changes. You subject your students to a non-living, certain-to-disappoint system, otherwise.

Try not to get the cart before the horse as you build your system. You probably are already doing that, unfortunately, because rapid test construction apparently is being forced from Washington. May I ask you to consider joining states like Kentucky and Louisiana that are pushing back on that undue pressure from DC? Your kids deserve better.

The world of higher-order thinking skills tests is well-known in Kentucky, and it has been something less than a blessing. I would suggest providing a relatively low weighting to the more subjective written response question types. Otherwise, you run a considerable risk of repeating Kentucky's troubled testing history which has seen the rise and demise since 1992 of not one, but two, higher order thinking skills testing systems.

Bluegrass Institute for Public Policy Solutions
P.O. Box 11706
Lexington, KY 40577