BLUEGRASS INSTITUTE

‘KERA (1990-2010): WHAT HAVE WE LEARNED?’

BY RICHARD G. INNES,

EDUCATION ANALYST
KERÁ: What have we learned?

“I came on the local (school) board in 1987. What you just said to me is no different than what I heard in 1987. So why should I be hopeful?”

Kentucky Board of Education Chair Joe Brothers’ reaction to proposals to fix Kentucky’s continuing education problems during the October 2009 meeting of the Kentucky Board of Education1

Executive Summary
Not all of the policies contained in the Kentucky Education Reform Act of 1990 (KERA) have failed. However, it is evident that many Kentuckians have come to share Brothers’ frustration because a great many aspects of KERA have not worked out as promised.

In its 20 years of existence, a large number of KERA’s major efforts have proven unsuccessful. This despite the fact that educators assured citizens time and again – often using terms like “the research shows” – that their fad ideas were succeeding. In consequence, educator credibility has dropped to the level where policymakers like Brothers now feel led to make comments like the one above.

As Kentucky embarks on a serious rebuilding of the state’s education standards and the related assessment program and considers other efforts like charter schools to genuinely and thoroughly reform public schools, this paper presents a listing of major KERA initiatives that Kentuckians were assured were supported by research but which did not work in practice.

A prime goal of this report is to insure that we avoid reintroducing any of these failed ideas as Kentucky’s education standards and the related assessment program are rebuilt.

Another goal is to raise awareness of the fact that today’s educators still don’t know enough about how kids learn. Arthur Levine, former president of Teachers College at Columbia University, has noted: “There is widespread disagreement among policymakers, researchers and practitioners about what constitutes good research and how to prepare education researchers.”2

As Levine and others have realized, most education research is not conducted with sufficient rigor to “show” anything with an adequate degree of certainty.

However, we do know somewhat more than we did in 1990. We now know that many major programs failed under KERA, including the following:

• Writing portfolios for accountability – While a good instructional tool when used properly, when used as part of the assessment program, writing portfolios actually impair writing instruction. Thus, they were removed from the assessment program by Senate Bill 1 in 2009.

• De-emphasis on grammar, spelling and punctuation – The argument was made that students would learn these skills naturally as they wrote. That did not happen. Adverse
effects of the de-emphasis were intensified by arcane rules that impacted both the instruction of writing and the grading of writing portfolios. That grading process used “holistic” approaches that de-emphasized writing mechanics. All of these arcane rules were intended to prevent cheating. Thus, these teacher-limiting rules became an unintended, unwanted consequence of using writing portfolios in the state assessment program.

- **Mathematics portfolios in assessment and accountability** – Instead of doing math, students were mostly just writing about it. Math portfolios – included in the state’s assessment from 1993 to 1996 – were discarded by the legislature following numerous complaints from teachers and parents.

- **Fuzzy math** – Pushing process to the detriment of students’ learning standard math operations like division and fractions was found to adversely impact all higher level math courses from algebra up both in K-12 schools and in postsecondary institutions. The math program also tried to cover too many topics, becoming a “mile wide-inch deep” in the process.

- **Whole language reading** – The de-emphasis of phonics in the early days of KERA was misguided from the onset.

- **Ungraded Primary** (formerly Kindergarten to third grade) – The ungraded primary concept required teachers to make constant judgments of student performance and to individually determine when students were ready for fourth grade. Today, ungraded primary technically remains on the books but the real concept is generally unenforced and often totally ignored. Furthermore, new research discussed in the body of this report indicates that there is no difference in performance between Kentucky elementary schools still using ungraded primary and those schools that have reverted to a standard, grade-by-grade organization.

- **Performance Events** – These unusual assessment items required students to work in small groups to solve a problem. The students then individually wrote reports to be graded at a central site. A typical fourth-grade performance asked students to estimate the number of ladybug cartoon images presented on a piece of paper. The intent was to see if students would do something like dividing the paper into fourths, counting the bug images in that section, and then multiplying by four to get the total estimated number of images. Of course, by the fourth grade it should have been trivially easy for students to simply count all the images on the paper. In the end, issues of creating, linking and equating different performance events from year to year proved unworkable, and this “great hope” of KERA’s assessment program failed disastrously in 1996. The middle school performance event selected that year was far too difficult, and all middle schools got dramatically low scores. Eventually, the legislature threw all performance events results out of the evaluation process.

- **First testing company was an unknown with limited resources** – Despite strong support from the Kentucky Department of Education (KDE), Advanced Systems in Measurement and Evaluation proved unequal to the challenge of performance events and was terminated for cause following a scoring fiasco resulting in inaccurate scores for every elementary and middle school in the state. The erroneous scores, in turn, caused improper distribution of $25
million in reward money. Kentucky taxpayers had to provide an additional $1.5 million to pay rewards schools earned but had not received due to the scoring error.

- **Kentucky Instructional Results Information System (KIRIS), Kentucky’s first assessment program** – Disbanded for poor performance in 1998 by the legislature. Scores inflated while other indicators of performance such as the National Assessment of Educational Progress (NAEP) and the ACT college entrance test indicated Kentucky education was making little progress.

- **Multiple choice questions** – In, sort of, at the inception of KIRIS testing in 1992. Then out completely after 1994. Following considerable criticism of that 1994 decision, back in, sort of, from 1997 through 1998. Finally accepted as necessary and included in the accountability formula for the first time after KIRIS was disbanded following its 1998 administration.

- **Commonwealth Accountability Testing System (CATS), Kentucky’s second assessment program** – Launched in 1999 and disbanded for cause in 2009. Again, scores were inflated beyond anything that could be supported from other data sources, including a chronically high statewide remediation rate for entering college freshmen and continued severe inflation of proficiency rates compared to the NAEP.

- **Dropout rates for school accountability** – Used for school accountability from the beginning of KIRIS in 1992. Found highly inaccurate in an official audit by the Kentucky Auditor of Public Accounts in 2006. Focus shifted to equally unreliable graduation rate statistics based on the same inaccurate dropout figures after No Child Left Behind began. Now, slated to be replaced by a more carefully researched graduation rate formula in 2011. Internal audits by the Kentucky Department of Education, if conducted at all, never revealed problems found in the official 2006 audit.

- **Site Base (or School Based) Decision Making Councils (SBDM)** – This KERA statewide creation stripped locally elected school boards and superintendents of all real authority in schools. Instead, key decisions on matters like curriculum and final funds distribution became the purview of these teacher-controlled councils. Along the way, SBDM thoroughly destroyed the logical chain of command and accountability in education and replaced it with a chaotic system where accountability is unclear and difficult to enforce. Superintendents can still fire principals on paper, but doing so in practice is often unworkable because the SBDM can actually rehire the same person. Overall, the SBDM governance model thoroughly eliminated school accountability to local taxpayers and parents.

- **Accurate reporting of data and implications** – This problem was extensively highlighted in a study conducted for the Kentucky Office of Education Accountability in 1995. Though somewhat improved over time, KDE’s data reporting still suffers from “spin.” One notable example was a recent effort to protect the inflated CATS assessment with the KDE-operated 2008 CATS Task Force. KDE tried to hog-tie the task force so only minor changes would be recommended for CATS. That KDE effort ultimately proved a failure when the legislature voted to disband CATS in early 2009.
• **School accountability** – KIRIS and CATS identified very few schools as problematic although external data from NAEP, ACT, No Child Left Behind (NCLB) and high college remediation rates indicated problems were far more extensive. In particular, NCLB firmly exposed the very serious failure of CATS to identify and penalize achievement gaps for disadvantaged students. However, only a few schools have lost their site base council authority due to low performance. For example, in 2008, Kentucky had 34 schools in the lowest NCLB performance category of “Tier 5.” Those schools consistently failed to make adequate progress for at least six years and were supposed to implement alternative governance plans. Yet in that same year, only four schools lost their SBDM authority, all due to poor performance on CATS. Schools did not start to lose SBDM authority under federal accountability efforts such as Race to the Top and NCLB until 2010. To date, only 10 SBDMs have been impacted by federally required accountability programs, and all of those are due to Race to the Top/School Improvement Grant program related activities.

• **Educator accountability** – Despite a great deal of commentary about Kentucky wanting highly skilled teachers in every classroom, the state’s education system has no formal process to evaluate and remove tenured teachers who fail to do even a minimally acceptable job of educating students. Even in cases where teachers have committed serious crimes, it can be time-consuming and expensive to suspend or revoke the teaching certificates of individuals who clearly have no place in the classroom. For teachers who simply lack competence, no formal process exists to describe the steps to take for decertification. The Kentucky Educational Professional Standards Board informs us that no Kentucky teacher has ever had a certificate suspended or revoked for such reasons.

• **Overcorrection for funding inequity** – When KERA began, districts with low per-pupil property wealth generally also had schools with low total per-pupil funding. That process has now sharply reversed. For example, students in the high property wealth Boone County Public Schools in Northern Kentucky sometimes attend classes in portable buildings while “down state,” low property wealth schools are now housed in beautiful and modern buildings.

• **Bang for the buck efficiency has steadily declined** – Even though test scores have risen slightly over time, the costs involved have skyrocketed. As a result, the dollar cost for each test score point on the NAEP and the ACT college entrance tests have both been steadily rising since KERA began. The Kentucky Constitution requires the state to establish an efficient school system, so the rapid rise in cost versus performance may not comply with the obvious intent of this requirement.

*******************************************************************************
Forward

It must be made clear from the outset that not everything in the Kentucky Education Reform Act of 1990 (KERA) has failed.

For example, KERA’s anti-nepotism provisions are solving a uniquely Kentucky problem. Prior to KERA’s enactment, some school districts in Kentucky operated as virtual family fiefdoms. The Kentucky Legislature helped change that undesirable situation through legislation.

Also, it is now clear from the National Assessment of Educational Progress (NAEP) that in the past decade, Kentucky did make a small amount of progress. However, its rate of improvement in that national testing program has been very slow and deserves a few extra comments, as the side box shows.

Judging Kentucky’s Progress on NAEP

According to the 2009 NAEP, only 37 percent of fourth-graders and, worse, 27 percent of eighth-graders are math-proficient. With such low performance at present, it is clear that any improvements have been small.

In fact, the amount of improvement in Kentucky compared to other states is difficult to judge. Since KERA, most other states experienced large shifts in demographics while Kentucky’s have altered very little. For example, the percentage of California’s white students plummeted after 1990 from more than 50 percent to around half that figure today.

In California’s classrooms today, Hispanic students, many of whom speak English as a second language, predominate. Because they score notably lower than whites, and because there has been a dramatic increase in California’s Hispanic population, that state’s overall performance appears rather low despite the fact that the Golden State is making progress with every racial subgroup.

In fact, even despite California’s demographic disadvantages, it actually was able to statistically tie Kentucky’s rate of proficiency in the most recent eighth grade NAEP writing assessment in 2007, a fact we will explore in detail later.

If California’s demographics had been similar to Kentucky’s, the Golden State very likely would have significantly exceeded the Bluegrass State’s performance on writing in 2007.
Furthermore, the small advances in Kentucky’s educational performance have come at a very high price, as we discuss further below.

Given the advantage of 20 years of hindsight, it’s painfully obvious: Kentucky’s public education system went down a significant number of expensive, dead-end streets while following poorly researched theories of “Progressive Education.”

KERA also suffered from bad research, which led to the imposition of a horde of new educational policies and practices. Never before had so many major, untested changes been implemented simultaneously, and on such a large, statewide basis. KERA in many ways was a massive, statewide experiment, conducted at public expense, using at least a generation of students as its subjects.

Unfortunately, many didn’t understand how much of KERA was theory, and how little of it was based on established, rigorously verified fact. In consequence, children were exposed to dubious education fads lacking verification from rigorous research. Worse, the manner in which KERA was implemented generally eliminated student control groups that could have efficiently and rapidly provided evidence about what was and was not working. That led in many cases to an overly extended period of use of those fads.

Thus, a major goal of this paper is to provide an overview of the “Progressive Education” ideas that were tried and failed under KERA. Hopefully, including this information will help insure that policymakers do not repeat those errors in the future, in Kentucky or elsewhere.

With that understanding, let’s examine some specific issues, starting with an overview of some of the major features in KERA:

THE KENTUCKY EDUCATION REFORM ACT (KERA) – Brief Overview

KERA is one of the most sweeping, statewide, education reforms ever attempted in American history. Some of the major changes implemented in this reform impacted:

1) Academic goals and their assessment and accountability
2) School governance
3) School funding
4) Nepotism restrictions
5) Creation of a special corps of trained educators who could be used to assist or take over problem schools and districts
6) Changes to training and certification of professional educators
7) Inclusion of special needs children in regular classrooms and in the assessment program
8) An extensive program of tutorial/remedial services
9) Family Resource/Youth Service Centers to serve needs of preschool and Kindergarten to grade 12 (K-12) students and their families

10) Preschool program changes

11) Mandatory multi-age environment for K-3 students

The final impetus for change in Kentucky’s education system came via a 1989 Kentucky Supreme Court decision in a case that originally challenged only the school funding issue. Going well beyond the narrow issue of funding equity for school districts, the court declared the entire body of Kentucky school law unconstitutional. The Supreme Court directed the legislature to completely revise these laws, a process that was completed in only nine months, culminating in KERA’s passage in 1990.

Because the court-mandated time frame to complete the reform legislation was short and the task of totally rewriting all the state’s education laws was daunting, the legislature chose to rely to a large extent on advice and direction from experts outside the commonwealth. A legal mandate existed, but few role models within the commonwealth could be found to guide Kentucky through its reform process. Thus, many portions of the reform were actually “off the shelf” ideas developed in foundation-funded policy centers and education schools around the country. They were accepted in a hurried manner by legislators who were not familiar with many of the concepts and who had insufficient time to analyze the proposed changes in detail.

Kentucky’s legislators and citizens were asked to accept at face value the assertion that “research shows” all the “Progressive Education” theories enacted in KERA were sound and would work. That leads to the next discussion point.

WHERE’S THE RESEARCH?

Based on the dubious research then available, Progressive-leaning, educators (many from outside Kentucky brought in to run KERA) confidently—and repeatedly—claimed that they really understood how kids learned. All they needed was free reign, and everything was going to be wonderful in the Bluegrass State’s schools.

The expression “the research shows” quickly became a standard defense whenever parents or legislators raised questions about KERA’s more exotic education theories. That explanation continued to be offered even when those theories didn’t seem to be working out in practice.

Now, 20 years of many negative experiences with KERA confirm: the research was faulty.

In fact, Kentucky provides extensive evidence to support 2007 statements about education research from Arthur Levine, the former president of Teachers College at Columbia University:

“There is widespread disagreement among policymakers, researchers and practitioners about what constitutes good research and how to prepare education researchers.”
Levine also provides disturbing insight into the motivation of some education researchers. In his report, “Educating Researchers,” he writes about his interviews with education school personnel responsible for training the research community. Says Levine:

“It quickly became apparent that in today’s highly charged environment, those interviewed for this study had less interest in ‘truth telling’ than in defending their positions.”

The research situation was even less satisfactory when KERA was launched in 1990. Clearly, the Progressives who took over Kentucky’s education program really could not know much about what really worked for students. The research they relied on couldn’t “show” them what worked because it simply wasn’t rigorous enough to accomplish that task.

Even today, there remains no general agreement about what constitutes acceptable research in the education community. Worse, as Levine points out, not every voice in that community is honest about the situation.

Sadly, this problem really isn’t new. Education historian Diane Ravitch points out in her book, “Left Back, A Century of Failed School Reforms,” that excessive educator self-confidence is far from a 1990s phenomenon. Ravitch writes of early 1900s progressive educators:

“They believed that they understood society as no one else did; unlike those who worked in the schools, they understood psychology – the principles of learning – and sociology – the dynamics of society. Given their supposedly superior understanding of industrial, social, economic, and educational conditions, they saw themselves as being empowered to remake the schools.”

The 1990s incarnation of excessive educator arrogance led to many problems for Kentucky education, including the fact that Progressive educators remained blind to past mistakes, perhaps due, in part, to sometimes poor documentation of those earlier failed efforts.

For example, KERA supporters actively pushed students working in groups. However, Ravitch notes the concept of students working in groups was pushed by Topeka Kansas educators as early as 1918. Yet, this nearly century old concept would reappear, presented as a new idea, when KERA got going even thought this “new” reform already fizzled out…way back when.

Sadly, that historical knowledge was apparently lost or ignored by many who reincarnated the idea in Kentucky’s 1990 education makeover. Furthermore, the group-work idea would lead to an expensive mistake when the new “Performance Events” became part of the Kentucky assessment in the mid-1990s, as discussed below.

One of KERA’s most convincing lessons is that placing blind faith in educational research is most unwarranted. Some researchers in the field get it right, but many others do not – either because they lack adequate information and the requisite analysis skills, or because they willingly defend selfish interests at children’s expense. In general, much of the research isn’t good enough to show anything.
Education funding outstripped Kentucky’s test results

The Kentucky Constitution’s requirements for the state’s education system include language often overlooked by the pro-KERA faction. Section 183 of Kentucky’s Constitution says:

“The General Assembly shall, by appropriate legislation, provide for an efficient system of common schools throughout the State.” (Emphasis added)

This language certainly indicates that an extravagant and uneconomical education program is not authorized.

One metric that explores academic efficiency is to divide test scores by per-pupil spending. The three following figures do that with NAEP fourth-grade math and reading scores, and for ACT college entrance test scores. The specific spending figures used represent the total per-pupil revenue collected each year in Kentucky and have been corrected for inflation using the US Census Bureau’s CPI-U statistics.

In every case covered by Figures 1 to 3, the number of test points achieved for each real dollar of education revenue has dropped dramatically since KERA was enacted. In other words, the cost per test-score point on these important tests has risen dramatically.
For example, the data in Figure 1 shows there was a 20 percent drop in the points per dollar in NAEP fourth-grade math performance between 1992 and 2007. A similar situation occurred in NAEP reading, as shown in Figure 2.

There was a 24 percent drop in points per dollar in NAEP fourth-grade reading scores between 1992 and 2007.

The results for the ACT follow a similar trend of reduced “bang for the buck” over time.
There was a 28 percent drop in the points per dollar in the ACT between 1993 and 2008. Thus, this analysis shows Kentucky education got significantly less efficient over time at producing test score points in relation to every inflation-corrected dollar expended. In simple terms, the state’s education system got dramatically less efficient.

As a note, we used inflation corrected dollars to create Figures 1 to 3. Had we used current year dollars instead, the decay in points per dollar over time would have been much more pronounced.

What didn’t work: Specific programs

KIRIS

The Kentucky Instructional Results Information System (KIRIS) was the cornerstone component of KERA. KIRIS was intended to be the major driver of reform in Kentucky’s public schools.

In 1990, this assessment was highly experimental. Some of its features discussed below had never been used in any large-scale statewide assessment program.

• Questionable question types

The KIRIS assessment featured a sharp reduction in emphasis on traditional multiple-choice questions. Instead, it employed much more complicated test components, including open-response questions, writing and mathematics portfolios and performance events.
Large scale, statewide use of writing-intensive questions and performance assessment formats was generally unknown territory. Getting the non-traditional assessment elements up and running not only proved very challenging, the effort ultimately was unsuccessful.

- **Poorly defined learning goals and outcomes**

At its onset, KIRIS took a huge step in the wrong direction.

The essential first step in the intelligent design of an assessment system is development of a detailed set of descriptions of desired student learning goals and outcomes. Without clear descriptions of exactly what students should know and be able to do at the end of their academic preparation, teachers have no idea what to teach. Furthermore, without specific, clearly stated performance standards and goals, people charged with designing test questions also wind up conducting an exercise equivalent to shooting in the dark. KIRIS lacked an underpinning of well-designed standards, which are detailed descriptions of learning goals and outcomes.

The KERA legislation itself only outlined some very general expectations for the state’s academic goals and standards. The original act contains only six very general statements in Section 3 about what is expected from students:

“**Schools shall develop their students’ ability to:**

1. Use basic communications and mathematics skills for the purposes and situations they will encounter throughout their lives;

2. Apply core concepts and principals from mathematics, the sciences, the arts, the humanities, social studies, and practical living studies to situations they will encounter throughout their lives;

3. Become a self sufficient individual;

4. Become responsible members of a family, work group, or community including demonstrating effectiveness in community service;

5. Think and solve problems in school situations and in a variety of situations they will encounter in life, and;

6. Connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned and build on past learning experiences to acquire new information through various media sources.”

Since KERA was legislation, lawmakers actually intended to offer incomplete guidance to teachers and test developers. Recognizing that technical expertise was required, legislators wanted the newly reformed Kentucky Department of Education (KDE) and Kentucky Board of Education (KBE) (both were fully reformed in KERA) to do the extensive, technical job of fleshing out the complete list of goals for students.
The KDE/KBE team did create “Valued Outcomes,” a list of goals and outcomes for students. However, it contained just 75 woefully inadequate, single-sentence statements to cover all K-12 education.

Despite its critical and virtually nonexistent underpinnings, the KIRIS tests launched in 1992 anyway. But by 1994, it was obvious to many – including the KDE itself – that the initial set of 75 “Valued Outcomes,” wasn’t suitable. Another curricular document, the “Kentucky Curriculum Framework,” was issued by the KDE. However, this document also proved inadequate to inform both teachers and test writers about the real knowledge and skills Kentucky students were supposed to master under KERA.

More attempts to expand on the inadequate “Curriculum Framework” would follow later, including “Academic Expectations,” which incredibly further shortened the original list of learner goals from 75 to an even more incomplete list of just 57 items.

It wasn’t until 1996 that the first document that even began to approach the level of detail needed for intelligent test development, called the “Core Content for Assessment,” was released.

All of these documentary changes were made well after the initial question banks for KIRIS were created. Thus, as the first KIRIS assessments were released in 1992, the die was already cast on a totally inadequate foundation. Problems were virtually certain as teachers and test creators were left guessing about what was supposed to be “fair game” on the KIRIS assessments and in the curriculum.

- **Dubious accountability model**

The statewide accountability model for KIRIS was unlike anything previously attempted. The basic unit of accountability was neither the student nor the individual teacher. Instead, the entire school as a unit was held accountable under the KIRIS accountability program.

Individual students definitely had no accountability for KIRIS results. In fact, as discussed in the next section, KIRIS scores for individual students were neither valid nor reliable enough for such purposes under any condition due to the way individual student test booklets were created using a process called “matrixing” (also discussed below).

Individual teachers also were not held directly accountable under KIRIS. A weak teacher in a generally strong school might not suffer any consequences under KIRIS as originally implemented, while a strong teacher in an overall weak school could fall prey to sanctions that could lead to him or her being fired.

The KIRIS program’s key annual score was a single number called the “School Accountability Index.” It was derived from a weighted average of scores from a number of different assessment elements that included results from tests of seven subject areas. Areas covered included: reading, math, science, social studies, on-demand writing, arts/humanities and practical living/vocational studies.
Also included in the School Accountability Index were scores from writing and math portfolios along with results from a new, performance-oriented testing protocol called “Performance Events.” Very briefly, these attempted to evaluate how students working in teams solved a group problem. Performance Events are discussed in greater detail below.

Finally, a small amount of weight in the School Accountability Index came from nonacademic areas such as attendance rates, retention rates (the proportion of students held back at the end of each year), middle and high school dropout rates, and for high schools only, the proportion of graduates who successfully transitioned to either college, the military or the workforce.

Although schools received a School Accountability Index update every year, an actual accountability determination for each school was only made every second year when two years of data were averaged together. Thus, the accountable School Accountability Index score was determined only once in each two-year biennium.

The end goal under KIRIS was for all schools to reach a School Accountability Index of 100 at the end of 20 years in 2012. That goal year was later extended to 2014 as the slow progress of schools became apparent.

Progress of each school was judged against a “baseline year” set of scores that was taken separately for each school in 1992. Each school had a separate and different set of target scores that needed to be reached in each biennium to avoid accountability. An example taken from the “KIRIS Accountability Cycle I Technical Manual” shows how the process worked.19

Assume a school got a baseline year accountability index of 30 in 1992. That meant the school needed to increase the score by 70 points to reach the goal of 100 in 20 years. With 10 biennial periods available to reach that final score, the school needed to average a score increase of seven points in each of the succeeding bienniums. So, in 1994, the school would have to achieve a score of 30 plus 7, or 37. In 1996, the school would need a score of 30 plus 14, or 44 to avoid sanctions, and so forth.

However, if another school got a baseline score of 50, it only needed to improve by five points during each biennium to avoid sanctions.

Thus, low performing schools had to improve at a faster rate than better performing ones.

• **Testing in too few grades**

All of the extensive testing required initially was conducted only in grades four, eight and 12.20 This had adverse impacts on instruction as teachers in other grades – especially those for the former Kindergarten to third-grade group, which became known as “Ungraded Primary” – tended to overly relax while their compatriots teaching in the three KIRIS target grades were under extreme pressure to perform.
In addition, high schools soon begin to grumble that they were being held hostage to “Senioritis” by students in graduating senior classes who applied minimal effort on the tests because they faced no consequences from KIRIS.

This convenient complaint was soon echoed by teachers in the elementary and middle schools, even though fourth-grade students usually are not terribly attuned to the consequences of any tests they take. If the teacher says it is important, elementary school kids generally go along with them.

• **A testing surprise**

By 1994, it was obvious that something unexpected – and highly undesirable in the eyes of many – was happening. Some of the state’s strongest schools, those with very high initial baseline scores, were getting into trouble with KIRIS. Schools in the state’s highest socioeconomic areas like the Anchorage Independent schools in the Louisville suburbs and the Dixie Heights High School in Northern Kentucky found themselves under pressure after the 1993 and 1994 KIRIS results were released. This happened even though these schools were generating some of the highest School Accountability Indexes in the state. The problem was that they were not making further improvements fast enough. As a consequence, expensive remedial resources started to be targeted at both while schools elsewhere with far lower scores escaped both notice and badly needed assistance.

It made no sense to many Kentuckians to spend money and effort in the state’s wealthiest, and highest scoring, schools while other schools in places like inner-city Louisville and Covington were turning in test scores far below those of the schools in trouble.

Other undesirable issues also arose. It quickly became obvious that KIRIS results for individual students were sometimes dramatically out of line.

For example, this author’s own daughter won the Kentucky PTA’s statewide writing contest as a ninth-grader. In her tenth-grade year, she took her first set of KIRIS practice assessments (the actual KIRIS accountability grades were grades four, eight and 12 at this time). Despite her clearly superior writing ability (later confirmed by many other indicators, including her college performance), my daughter only scored “Apprentice” for writing on KIRIS. That was the second-lowest of four possible scores.

Thus, one of Kentucky’s arguably top young writers was identified by KIRIS as one of the state’s weaker writing students. My daughter’s story is far from unique.

On the other side of the coin, it was revealed in the mid-1990s that some students with learning disabilities were getting top KIRIS scores of “Distinguished.”

So many individual student scores could not be so questionable without the validity of the overall KIRIS results also suffering.

• **Test-result credibility decays in the legislature**
As early as 1994, it was obvious that KIRIS had developed some serious credibility issues. As a result, the legislature placed a hold on a key part of the KIRIS accountability program.

Under the original KIRIS plan, a school that didn’t achieve its School Accountability Index biennial target was taken over by a specially trained individual called a “Distinguished Educator.” After six months on site, the Distinguished Educator would then make “binding recommendations” to the local board of education concerning the retention or firing of teachers and the principal.

However, given sharp controversies brewing about unrealistic scores awarded high-scoring students and top-performing schools getting taken over, legislators refused to allow the 1994 KIRIS results to be used for school takeovers.

So, the law was modified. Distinguished Educators would still be sent to schools that didn’t make their required progress under KIRIS, but the authority of those educators was limited to advisory status only.

Thus, as early as 1994, KIRIS’ credibility was already in trouble.

(Note: The legislature later did allow school accountability to proceed, but only for one KIRIS biennium. At the close of the 1995-1996 biennium, nine schools were designated as “In Crisis” and a “Distinguished Educator” was sent in to take over those schools. However, this accountability program also proved problematic, and by 1998, the Distinguished Educators were re-titled “Highly Skilled Educators,” in part to denote they would no longer be given takeover powers in low-performing schools. It would be a decade before a school again lost its rights to self-manage its affairs under the SBDM program.)

- **Unstable testing grades, uneven instruction delivery**

For a testing system to create credible scores in a state assessment system, the tests must be stable over time. But KIRIS had a huge problem with lack of stability in a very basic area – the limited selection of grades where the assessments were actually given.

As previously mentioned, when KIRIS started, all elementary school testing was conducted in the fourth grade. Middle school testing was all conducted at grade eight while high school seniors did all the testing for credit in their schools. This policy led to a number of problems such as creating excessive pressure on students and teachers in grades four, eight and 12, while there was no penalty for a more lax attitude about performance other grades.

Thus, the grades where testing occurred changed somewhat by the 1994-95 school year. Math portfolios were shifted to the fifth grade while all of the high school assessments except for math and writing portfolios were shifted to grade 11.

There would be many more testing grade changes during the 1996 to 1998 KIRIS biennium. Grade four testing was eventually reduced to only include reading, science and writing. Grade
five now would be tested in math, social studies, arts and humanities and practical living and vocational studies.

In middle schools, grade seven picked up tests in reading, science and both the on-demand writing prompts and portfolios. By this time, math portfolios had been dropped from mandatory KIRIS testing. In grade 12, only the writing portfolios remained.

As a consequence of uneven subject testing from grade to grade, the delivery of curriculum from one year to the next became very uneven in some schools. Teachers over-concentrated on the subjects evaluated by KIRIS in each given grade while other subjects not tested in those same grades received reduced emphasis. This problem improved slightly after No Child Left Behind (NCLB) required annual testing of both reading and mathematics in all grades from third through eighth. However, concerns about uneven emphasis continue today for other subjects not covered by the NCLB requirements.

Furthermore, changes in the grades where tests for each subject were administered led to further concerns about the validity of the entire KIRIS program. It was critical that moving subject testing from one grade to another did not introduce instability in the scores. Otherwise, the trend lines essential to the validity of the School Accountability Index calculation would be destroyed. It was a huge psychometric challenge. Unfortunately, KIRIS never did effectively alleviate concerns that these changes in the grades where different subject tests were administered had not corrupted the assessment’s trend lines.

* Open-response questions led to low validity and reliability of individual student results

The reconstituted Kentucky Department of Education mandated by KERA in 1990 had a very strong bias against multiple-choice questions. The KDE believed multiple-choice questions tended to lead to narrowed curriculum and fostered a “teach only to the test” mentality. Thus, Kentucky’s KIRIS assessments were very heavily based on open response writing questions.

But the KDE ignored an extensive and constantly growing body of knowledge about testing within the psychometric community.

For one thing, multiple-choice tests are not called “Objective Tests” without reason. The facts are, multiple-choice questions are very efficient vehicles for determining student knowledge and can even be used to assess higher order thinking skills.

While multiple-choice tests certainly had limitations in 1990 (somewhat reduced today), they also had vast compensating values. The bias against using these questions would create some heavy “baggage” for KIRIS.

In sharp contrast, open-response questions are not efficient. It takes students much more time to answer an open-response question. If testing time is to be kept reasonable, the inevitable result is open-response questions suffer from reduced coverage of subject matter knowledge on the part of individual students.
It also takes a lot of time-consuming human effort to grade open-response questions, and doing so properly requires personnel with considerable subject knowledge and additional training on how the scoring process is supposed to work. Obtaining scorers with sufficient talent is expensive.

These realities combined to create extensive problems for the KDE’s KIRIS assessments.

For one thing, if the full breadth of the curriculum was going to be assessed in a reasonable amount of testing time, a tradeoff was essential.

**The matrixing mistake**

The department’s solution to the testing time problem with open response questions was to use a test question management system used in the NAEP, which is known as “matrixing.”

Under matrixing, although the full question bank for the test can be quite large, each student only answers a small subset of the questions. Basically, each student is only tested on a portion – sometimes a rather small one – of the overall curriculum.

While each student is only tested on a small part of the curriculum, the theory of matrixing stipulates that when a large number of those partial collections of student performances are averaged together, a reasonably accurate picture of the overall performance of the entire student group can be obtained. This is held to be true even though each student only answers a small number of questions and is, therefore, incompletely tested on the full curriculum.

The inevitable lack of complete and accurate data on individual teacher and student performance seemed acceptable to the department because KERA only required KIRIS to hold entire schools accountable. There was no requirement to provide accurate scores for individual students or to use them for assessment of individual teachers in the schools.

In the end, the matrixed test format proved highly unsatisfactory.

**Little data – High costs**

Perhaps the most severe adverse side effect of KIRIS’ matrixed testing scheme was that the scores of individual students – because they only got tested on parts of the curriculum – could not be used as valid indicators of their mastery of the full curriculum. While parents wanted to know how their children were doing in school, KIRIS’s flawed design could not accurately tell them.

This deficiency was readily admitted by a KDE official in a Kentucky Post on Aug. 11, 1994:

> “Individual student rankings aren’t statistically reliable because they are not derived from a sufficient number of test questions, said Neal Kingston, the official who oversees testing for the state Department of Education.”

25
Other reports would also make the lack of individual student score validity and reliability very clear, as discussed later.

In addition to not providing accurate data about individual students, KIRIS could not provide useful information about individual teachers, either. Because matrixed test results only become reasonably accurate when a large number of student scores are added together, the validity of scores collected at the individual classroom level was highly suspect.

While some teachers may not have been necessarily disappointed about a testing system incapable of holding them individually accountable for their performance, the lack of detailed data served schools, administrators and the public very poorly. Not only could KIRIS not be confidently used to identify under-performing teachers, it couldn’t identify exemplary instructors, either.

The decisions to place excessive reliance on inefficient written-answer testing formats did result in Kentucky spending a lot of money on KIRIS. The assessment quickly became one of the nation’s most expensive testing systems (see side box) – one providing very little useful data in return for its very high costs.

**How Much Did KIRIS Cost?**

Estimating the total cost to administer the KIRIS assessments is a challenging task. For one thing, the contract costs for the prime testing contract, which were considerable, only represented a small amount of the overall costs of this time-consuming assessment. There were significant added costs for such things as district and school-level test preparation and testing support tasks.

Perhaps the most exhaustive attempt to accomplish this difficult cost analysis is contained in a report from Lawrence O. Picus and Alisha Tralli titled *Alternative Assessment Programs: What Are the True Costs?* This is published as CRESST Technical Report 441 and is online at: [http://www.cse.ucla.edu/products/Reports/TECH441new.pdf](http://www.cse.ucla.edu/products/Reports/TECH441new.pdf).

Table 13 in that report shows the total opportunity cost per student tested with KIRIS was an astronomical $1,791.96!
• **Open-response questions slowed turnaround times for results**

An additional advantage of multiple-choice tests is that unlike open-response assessments, they can be rapidly scored by machine. Even with experienced and highly skilled readers, open-response questions take much longer to grade.

Using large numbers of open-response questions slowed return of KIRIS results to schools. The next school year was always well under way by the time KIRIS scores were finally released, resulting in schools being unable to use the results to make intelligent adjustments to their curriculum before the school year began.

Thus, another very important benefit expected from quality testing systems was lost in the KIRIS system.

• **Mathematics portfolios in assessment and accountability – Failed assessment element during the KIRIS era**

The KIRIS mathematics portfolios were the first major Progressive Education fad idea to fail in the Kentucky state assessments.

The theory behind these portfolios – collections of students’ papers on the subject of math – was that they provided a way to determine if students were really learning about math, specifically how to apply it in new situations.

The problem is that writing about math isn’t doing math. Also, writing about math took time away from instruction and practice in important concepts and skills such as computational accuracy and developing proficiency with fractions.

From the outset, Kentucky’s rank and file math teachers were upset about math portfolios. They argued that math portfolios were significantly interfering with math instruction.

Parents also quickly became alarmed as it became evident their children were not learning important math elements, especially computational skills and fractions. Meanwhile, students wasted many hours on what was too often alleged to be just a writing exercise with very little math value.

ACT college entrance mathematics scores provided an early clue that math portfolios, which were introduced in 1993, were not beneficial. As Figure 4 shows, Kentucky’s public school graduates’ ACT math scores declined in the four years that math portfolios were used. The scores sharply rebounded in 1997 after the portfolios were discontinued in the state assessment, although some of this rise is due to the ACT permitting calculators on the exam for the first time during that year. However, notice that the state’s ACT math score continued to rise the following year – beyond what could be explained by the introduction of calculator use on the ACT.

*Figure 4*
Reacting to widespread complaints, the Kentucky Legislature voted in 1996 to remove mathematics portfolios from the state’s accountability program.

The KDE attempted to do more research on the dubious mathematics portfolios in succeeding years with the hope of reintroducing them to KIRIS. But all discussion of them had long since ended by the time CATS, the state’s second assessment program, was disbanded in 2009.

- **Performance Events: Failed assessment element during the KIRIS era**

Performance events as used in Kentucky were an interesting concept.

One fourth-grade example of a performance event involved an 8 ½ by 11 inch piece of paper covered with a large number of life sized cartoon images of the lady bug insect. Students were to work in small groups, generally composed of four students, to develop a method of estimating the number of images present on the paper. Students were supplied with things like rulers, protractors and compasses to conduct their work.

The intent was that students might do something like dividing the paper into four or more equal-sized areas so they could count a smaller number of images and then extrapolate to the likely total number on the paper.

(Note: By the fourth grade, students should have been able to rapidly count up all the bug images on the page with ease, but that obvious potential corruption of the event does not seem to have occurred to the people who created the problem. This quick example demonstrates that creating really worthwhile performance events is a very demanding and time-consuming task.)
While one professed goal of Performance Events was to evaluate how students worked in teams, the final judgment of performance was based on written reports separately created by each student.

To preclude biases and possible cheating, Kentucky’s performance events were administered by trained facilitators (not teachers) sent by the testing company to each school to conduct the assessment element under controlled conditions.

After the student team’s work was completed, the individual reports were collected from each student for scoring at a central site by scorers who had not witnessed the group work. Thus, in essence, the performance events became largely just another writing exercise, which further confounded the real meaning of the results. Were the final scores a fair indication of how students worked in groups on problems, or just writing exercises for individuals?

There were other problems with the performance events program in Kentucky, including the fact that the performance events were dramatic and easy to remember. Thus, they had to be changed each year to prevent “teaching to the test.”

However, the constant need for new “events” created another host of problems. It is an enormous challenge to annually create different performance events that have comparable difficulty and can be scored on some sort of common scale. If the scoring can’t be kept consistent, then trend lines for the performance events become meaningless.

This was all new territory for statewide public school assessments in 1990, which meant the KDE was largely working in the dark.27

By January 1997, it was obvious that something had gone very wrong with performance events. The 1996 scores deviated wildly from earlier years’ performance events and even from the current year’s results from other academic testing areas. All attempts to correct the raw scores for these problems were unsuccessful, and the KDE reluctantly decided not to use performance events results for 1996.28

As a consequence, the KIRIS performance events and management of the contract for these radical and expensive assessment elements came into serious question.

At the request of the chair of the Kentucky Legislature’s Interim Joint Committee on Education, the Kentucky Finance and Administration Cabinet reviewed the KDE’s administration of the performance events contract. In a Feb. 3, 1997 letter, the commissioner of the Finance Cabinet identified a number of problems.29

For example, despite regulatory requirements to do so, the KDE failed to provide any notice of contract amendments to the Finance Cabinet, although at least 17 change-of-scope orders had been negotiated between the department and Advanced Systems in Measurement and Evaluation (ASME). Perhaps most importantly, the Finance Cabinet letter says, “To our knowledge, no audit has been performed in the six year history of the ASME contract.” The letter also
recommends that an independent financial and management audit be performed over the entire contract period. That suggestion would be initiated within the next year.

As a consequence of the independent audit that followed and the failure to operate a successful performance events program, the Kentucky Legislature permanently deleted performance events from the state’s accountability program amidst considerable expressions of disgust about the failure of the “great hope.”

The performance events debacle had further consequences. As 1997 wore on, the extreme dissatisfaction with the performance events developments resulted in legislators expressing growing concern about the overall performance of the state’s testing contractor, Advanced Systems in Measurement and Evaluation.

- **Final failure of KIRIS**

Aside from the total failure of math portfolios and performance events, the credibility of results from remaining KIRIS elements came under increasing pressure from the moment the assessment program began.

  - **Questions Questioned**

Concerns about the quality of questions on KIRIS arose as soon as the public began receiving sample problems from the test. Parents became upset, for example, when it was revealed that the “fuzzy math” grading didn’t require students to compute correct answers to get top scores.

For example, the question in Figure 5a was included in the 1992 KIRIS grade 4 released items in “World-Class Standards…for World-Class Kids,” a publication intended to inform teachers and parents about the new KIRIS assessments.30
Figure 5a

You and your friend, Katy, are in charge of designing a border to go across the top of the chalkboard. You each agree to make a design that would work well on the chalkboard. You make strips of graph paper like the one in your answer folder to use to draw the designs. On Saturday, Katy calls you and tells you how she constructed her design so you can draw her design.

a. First, read Katy’s instructions below.
   1. Put your pencil at START and, drawing line segments as you go, go to the right 3 units.
   2. Go down 2 units.
   3. Go left 1 unit.
   4. Go up 1 unit.
   5. Go left 1 unit.
   6. Go down 2 units.
   7. Go right 4 units.
   8. Go up 3 units.
   9. Now you have drawn the basic design. Repeat the design 2 more times.

Now, draw Katy’s design on the GRID for open-response question 3 in your answer folder.

Figure 5b shows the answer that the KDE says would earn a top score of “Distinguished.” But, this top-scored answer isn’t correct.

Figure 5b

As shown by added dotted lines in the figure, the student stopped drawing the line in the answer too soon. The student should have included the additional portion shown by the dotted line.

Math wasn’t the only area of concern. This next question was used on KIRIS for an unknown number of years up through 1997-98 before being released to the public, at which time the question was deemed scientific nonsense.

The picture, presented in the original in color, is a very good artist’s rendition of the walking stick insect common to Kentucky. While the question writer is obviously looking for answer “B,” “camouflage,” the facts are that none of the world’s “walking sticks” insects – most definitely including those that Kentucky’s fourth graders would be likely to be know – catch any prey. They are strictly plant eaters.

Figure 6

31
Many more dubious KIRIS questions were discovered throughout the life of the test.

In fact, it was a 1994 discovery of a bad 12th grade science question that launched the author’s first investigations into the performance of Kentucky’s testing program. This particular question asked students to select the proper circuit to hook up a battery, switch, light bulb and voltmeter so that all worked properly. None of the four multiple-choice answers offered on the test were correct.

Professional testing experts also had concerns about the quality of the KIRIS questions. When science and math questions were reviewed by a panel of testing experts for the Kentucky Office of Education Accountability (OEA) in 1995, their report said,

“Our inspection also reveals, especially for the science and social studies items, a striking de-emphasis of subject matter knowledge. It was almost as though the student did not have to attend science or social studies classes to be able to answer the questions.”

This is a remarkable statement for educators to make. Essentially, the OEA research team says the questions on the test were unrelated to the academic subject matter that students should be studying in science and social studies. That makes whatever the tests were measuring irrelevant to core academic knowledge.

Kentucky’s excessive reliance on open-response questions came with serious liabilities. It requires a lot of expertise to create and grade these questions. Such talent is expensive. It wasn’t available at a price the state felt it could afford.

- Reports growing more critical over time

Several key reports on KERA’s performance were issued in 1994 and 1995. The first one, which was issued by ACT, Inc., was titled: “A Study of Core Course-taking Patterns for Kentucky

In a strikingly direct summary comment, the ACT’s report concludes,

- “ACT strongly advises that no judgements (sic) regarding individual student decisions can or should be made, at this time, on the basis of the present Kentucky performance test results.”

The 1994 ACT study was followed by three more critically important studies in 1995. Two of these reports were fairly similar. One was from the Evaluation Center at Western Michigan University and was created under contract for the Kentucky Institute for Education Research. This institute was formed by the Kentucky Legislature in 1990 to generate independent research on the progress of KERA. The report in question is titled “An Independent Evaluation of the Kentucky Instructional Results Information System (KIRIS).” 35

The second of this similar pair of reports, and perhaps the most important of all, was created for the OEA by a six-person panel of testing experts. Its title is, “Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991 – 1994.” (June 20, 1995) 36 The OEA also was established by 1990 legislation to provide “in-house” expertise within the Kentucky Legislative Research Commission, which is an advisory organization operated by the legislature.

There was considerable overlap in the findings of the reports from Western Michigan and the OEA Panel’s report. Some of these findings include:

- “The Panel concluded that the KIRIS portfolio assessments are currently inappropriate for use in the KIRIS accountability system.” In this case, the Panel found portfolio scores were seriously inflated (OEA Panel, Page 6). There were similar comments from the Western Michigan group (Western Michigan, Pages 32 and 46).

- Evidence that the KIRIS assessments, regardless of the subject, were mostly exercises in writing. The Western Michigan report said, “Students who have content knowledge of the discipline but lack adequate writing skills are precluded from doing well on the assessment.” (Western Michigan, Page 33).

- The school level scores are not “sufficiently reliable to inform policy, administer rewards, guide school improvement efforts, and administer sanctions,” and further pointing out that, “There is not sufficient evidence available in the technical reports for us to assess whether the accountability index is or is not a reliable measure of a school’s effectiveness in meeting KERA’s standards” (Western Michigan, Pages 36 and 37). This finding was echoed by the OEA panel (OEA Panel, Page 5), which also concluded that year to year comparison of scores in KIRIS is of “questionable validity” (OEA Panel, Page 6).
“There is a considerable amount of measurement error associated with individual student level scores” (Western Michigan, Page 37). Along the same lines, the OEA Panel said, “The Panel considers the performance standards used to classify students as Novice, Apprentice, Proficient, and Advanced (sic) to be untrustworthy.” (Note, the Panel report confused the NAEP score of “Advanced” with the KIRIS score of “Distinguished”) (OEA Panel, Page 6).

“The assessment reports and reporting schedule are not designed to provide timely, detailed assessment feedback that school staffs can use to diagnose students’ learning deficiencies and focus and improve instruction accordingly” (Western Michigan, Page 52).

Findings that a narrowing of the curriculum had already occurred. (Western Michigan, Page 52). The OEA Panel said it as, “There has been a shift toward process at the expense of content in the curricula,” which the Panel urged the state to reconsider (Page 9).

Strong adverse comments about KDE’s closed-minded approach to multiple-choice questions that ignored the fact that these questions could be written to assess more than simple recall to answers. (Western Michigan, Pages 56 and 57). The OEA Panel echoed, “The elimination of multiple-choice items by the Department of Education from all of the important analyses unnecessarily restricts content coverage, lowers the reliability of the school accountability index, reduces the stability of the equating or linking of assessments from one year to the next, reduces the stability of performance standards, and creates less reliable and valid scores for schools and individual score reporting” (OEA Panel, Page 7).

Evidence from the NAEP, though limited, fails to show the large gains in KIRIS from 1992 through 1994. The OEA Panel flatly stated, “KIRIS markedly overstates actual gains in student achievement” (OEA Panel, Page 5-6).

Technical documentation was deficient (OEA Panel, Page 7).

Press releases were ignoring alternative explanations for results and missing important caveats (OEA Panel, Page 9).

Testing time was long enough that much more extensive student reporting should be possible (OEA Panel, Page 7-12).

A most remarkable and unique part of the OEA Panel Report is the technical legal analysis at the end of the document, indicating that KIRIS tripped a number of legal issues.

The OEA Panel concluded their recommendations section with this strongly critical comment, which refers to a testing standards document frequently cited by the testing industry and identified here as the “Test Standards”:

“In summary, the Panel feels that the Kentucky Department of Education set unrealistic expectations for itself and its contractor to deliver an accountability and assessment system that would meet all of the essential Test Standards within a period of several years. The result
is a system which has many technical shortcomings and, therefore, is not meeting the accountability and assessment needs of the Commonwealth.”

- Interfering with the public’s right to know?

It should be noted that the Western Michigan study’s release was delayed for six months. There was speculation at the time that this was due to an attempt by the KDE to unduly influence the findings from this supposedly independent effort. Those speculations were increased by comments in a Feb. 17, 1995 Courier-Journal news article:

“After months of delays and weeks of behind-the-scenes wrangling, an outside evaluation of Kentucky’s controversial new tests of student achievement will be released today, intensifying the ‘test wars’ that may go unresolved until the legislature meets next year.”

- A different, but critical, report

The third critical report from 1995 was created at the University of Kentucky. This was titled “An Evaluation of The Kentucky Education Reform Assessment System (KIRIS) and Academic Performance at the University of Kentucky (UK).” (March 1, 1995)

Unlike the Western Michigan and OEA studies, the UK report was a reaction to pressure placed by KDE on the university to start using KIRIS test scores as an admissions criterion for students. However, the university refused to cave to this demand.

Instead, the university commissioned its own study to see how well KIRIS scores predicted college performance for the freshman class of 1993. The report investigated results for 1,516 of the university’s students who had KIRIS scores and had graduated from Kentucky high schools in 1993. It concludes:

- Using KIRIS scores does not increase the ability of the University to predict how well a student will perform in specific introductory courses in English, mathematics, social science and the physical sciences.” (Page 4)

The report also showed there wasn’t a very strong relationship between the scores students got on KIRIS and their grades in a number of first year college courses, including English 101, Writing I, Math 109 (college algebra), Chemistry 105, General College Chemistry I, and Psychology 100.

In the end, UK politely declined to use KIRIS scores for admissions purposes.

It must be noted that the department’s proposal to UK flew in the face of the earlier public comments that KDE Associate Commissioner Neal Kingston made only a few months earlier when he stated, “Individual student rankings aren’t statistically reliable because they are not derived from a sufficient number of test questions.”
Thus, at the same time Kingston was publicly, and correctly, stating the KIRIS test could not be used for individual student purposes, the department was secretly attempting to make the test appear valid for such uses. And, it was maneuvering to trap UK into being an unsuspecting and unwilling agent in creating this invalid impression.

- **More interference with the public’s right to know**

Like the Western Michigan report, the UK study’s public release was also delayed for many months. Though dated March 1, 2009, its existence did not become public until five months later when it finally surfaced in an article in the Lexington Herald-Leader on Aug. 2, 1995.40

There was speculation at the time that the KDE was unhappy with the UK study’s findings and didn’t want the public to know about them.

Thus, the public’s access to important reports about KERA was at best delayed, and possibly actively interfered with, twice within the single year of 1995.

- **Exit Advanced Systems**

The last straw in the Advanced Systems testing contract with Kentucky came in 1997 when a KDE statistician made a most untimely discovery: Somehow, Advanced Systems improperly calculated the 1996 KIRIS results for every elementary and middle school in the state. The discovery came long after the $25 million KIRIS reward “pot” had been distributed based on earlier, erroneous scores.

Research also revealed that many more schools among those whose progress had been under-reported as a result of the error, were entitled to a share of the reward money. Since the reward money had already been distributed, taxpayers had to make up the nearly $1.5 million difference.

Kentucky Commissioner of Education Wilmer Cody promptly fired Advanced Systems for cause in the late spring of 1997.41 The firing was probably a pre-emptive strike because the legislature, quite riled already about the performance events debacle, certainly would have taken action had KDE continued to defend its contractor.

Thus, Kentucky’s first KIRIS testing contractor ultimately proved incapable of handling the highly intricate and futuristic assessment program, fulfilling a number of prophecies that this small company was a poor choice to handle such a complex and unprecedented project.

KIRIS itself wouldn’t last much longer than the contract with the company that created it.

- **KIRIS terminated**

By the time the 1998 regular legislative session convened in Frankfort, the KIRIS assessment was in serious trouble.
As an outcome of the disintegration of the performance events and the shock of the expensive scoring errors, the legislature commissioned an audit in late 1997 to determine if Advanced Systems could be held liable for damages and if the KIRIS program was academically defensible.

The lead audit, conducted by the firm of Coopers & Lybrand, L.L.P., concerned contract performance overall and the liability issue. It proved a disappointment. The KDE had done such a poor job of administering the Advanced Systems contract that it was impossible to determine exactly what the deliverables were. Change orders had been made in a very informal manner. Written records were often sparse. Advanced Systems in Measurement and Evaluation would escape all questions of liability as a consequence.

The companion academic audit of the KIRIS assessment was less equivocal than the contract performance audit. This separate audit report was titled “Kentucky Instructional Results Information System: A Technical Review.” (January 1998)42

This report’s lead author is Professor James S. Catterall, from UCLA, so this report is often simply referred to as the “Catterall Report.”

A great deal of the Catterall Report dealt with issues that had been raised by the OEA Panel Report of 1995, specifically looking for any progress that had been made.

A key issue was to determine if the scoring of the performance events and the other KIRIS items met, “A threshold of quality that would indicate, when compared to the reasonable standards of quality generally accepted by the psychometric evaluation trade community for similar services…that there was or was not substantial compliance with the contracts.”43

In addition, the team was to evaluate, “The validity, reliability, and appropriateness of specific KIRIS components and practices as measures of student achievement and performance as well as measures of school success in producing student achievement and performance.” 44

While some findings in the report were positive, there were also many cautionary findings that echoed the 1995 reports previously discussed. Some of the Catterall Report’s adverse comments included:

- Overall, not enough validity research had taken place since 1995 (Page 35).
- The lack of adequate research on performance events was only recognized after the fact (Page 42).
- Teachers were definitely teaching to the test and untested material and important content were being deemphasized (Pages 69, 73).
- Few teachers felt KIRIS was focusing teachers on more real-life applications, which was a KERA goal (Page 69).
- Teachers reported a de-emphasis on spelling, punctuation and grammar (Page 70).
- Expectations for low-achieving students had not increased, although expectations for high achievers had (Page 70).
There were mixed views about whether teachers had increased use of problem solving activities in the classroom (Page 71).

While students thought some KIRIS/KERA period changes were productive, they were still worried that they were not being adequately prepared for college. (Page 71).

Writing portfolios were taking too much time (Page 73).

There was too much KIRIS test preparation (Page 73).

The Catterall Report needed to have a lot of positives if KIRIS supporters were to prevail in saving the assessment from the rapidly growing concerns about its many perceived shortcomings. It didn’t provide that, and the legislature voted in the 1998 legislative session to disband the KIRIS program and replace it with a new assessment to be named the Commonwealth Accountability Testing System, or CATS.

However, KIRIS’ Progressive Education camp supporters were not totally overruled. As a consequence, the state’s new CATS assessment would inherit some of the same, unsatisfactory characteristics that eventually doomed KIRIS’ credibility. Those congenital defects would ultimately prove fatal to the CATS assessment, as well.

CATS, Kentucky’s second assessment program: Terminated in 2009

Kentucky’s second attempt at a reform assessment program had the catchy acronym of CATS, which stood for the Commonwealth Accountability Testing System. The CATS was created by the passage of House Bill 53 (HB 53) in 1998.

Like KIRIS, the CATS was intended to be a prime mover for reform activity. Unfortunately, that bill did not go far enough to correct the many problems in KIRIS.

To be sure, HB 53 confirmed the deletion of math portfolios and performance events from the assessment, but the bill also set the stage for continued over-reliance on open-response questions and writing portfolios in the new state assessment program.

Because CATS maintained a focus on open-response questions, test-book matrixing would be continued as well, which led to the inevitable consequence that CATS scores for individual students would continue to lack high levels of reliability. As an additional consequence, this prevented CATS from meeting another requirement in the 1998 CATS legislation that called for the state to move to a longitudinal assessment system – one that tracked student performance over time.

Thus, while other states were adopting state-of-the-art value-added assessments, like the Tennessee Value Added Assessment System, Kentucky remained trapped in its 1990 era test mentality.

• Truth on individual student scores comes out
The CATS basically continued without modification the individual, matrixed KIRIS academic test elements in reading, math, science, social studies, arts and humanities, and practical living and vocational skills. These academic tests were collectively renamed the Kentucky Core Content Tests (KCCT). As a consequence of keeping the same academic test formats with matrixed, writing-intensive questions, the CATS also inherited the same problems that plagued the old KIRIS test’s individual student score validity and reliability.

However, greatly amplifying a situation that started with KIRIS tests, defects in the usefulness of individual student academic scores from CATS became increasingly contentious with both parents and legislators.

The problems of the lack of student score validity were directly exposed in a 2005 letter from a panel of testing experts hired by Kentucky. This panel, titled the National Technical Advisory Panel on Assessment and Accountability, was created by the 1998 CATS legislation to provide legislators with an additional technical advisory source outside the education department.55

The Technical Advisory Panel’s letter says,

- “The test scores alone have important limitations as indicators of knowledge of the full Core Content.” In other words, the scores didn’t really provide accurate information about the students’ mastery of the academic content Kentucky wanted them to learn.

- “There is no assurance students are in fact evaluated against the exact same set of skills.”

  Basically, matrixing could lead to different students being evaluated on different skills.

- “There is enough error in student level scores to advise caution in interpreting and using these scores – and reason not to base high-stakes student consequences on these scores.”

- “We cannot recommend placing KCCT scores on transcripts.”

While the KDE would continue to make changes to CATS in an attempt to overcome these concerns, at no time in the history of CATS were its scores determined to be valid and reliable for individual students.

The statute creating CATS required scores to be placed on individual student transcripts if they were ever found acceptable – an action never taken because the scores were never found to be valid.

- **Score inflation continues with CATS**

The scoring inflation that started with KIRIS continued after CATS began.

The initial CATS benchmark testing was completed during 1999 and 2000. Then, the KDE significantly revised the scoring scheme inherited by CATS from KIRIS.

In the process of resetting the scores, which included rescoring the 1999 and 2000 assessments against new CATS standards, tremendous and very obvious inflation was introduced. The worst situation occurred in middle schools (highlighted in Figure 7), where sudden score jumps
boosted all those schools much closer to the proclaimed final goal for 2014 of an index score of 100.

Once CATS rescoring took place, the average middle school accountability index in Kentucky suddenly jumped by about 20 points (on the 140-point scale) – a huge, unearned bonus considering CATS maintained the same end goal of a score of 100 in 2014.

The inflation becomes even more pronounced when you consider that during the entire time KIRIS operated, middle schools were only able to raise their accountability index by about five points.

Furthermore, providing more evidence that the new CATS scoring was watered down, the middle school scores started trending upward, again. During the period of KIRIS from 1992 to 1998, the statewide average middle school Accountability Index actually started to flatten after 1994-95.
In the end, the trend in the CATS School Accountability Indexes inflated beyond anything that could be supported by other evidence such as NAEP scores.

For example, Figure 8 shows the proficiency rate on NAEP fourth-grade reading in Kentucky improved very little in the two decades of the state’s reform assessments (Note: NAEP reading is not given every year).

In very sharp contrast, the KIRIS assessment, which first reported much lower proficiency rates than the NAEP, later inaccurately showed reading proficiency rates that were somewhat better than the national assessment.

The situation changed more dramatically when CATS began. Fourth grade reading proficiency rates on the CATS (technically on the Kentucky Core Content Test part of CATS) suddenly skyrocketed in the first year of CATS in 1999, after the rescoring took place. The rates then continued to escalate much faster than the very sluggish rise shown by the NAEP.

Kentucky’s NAEP performance shows that the dramatic jump in CATS rates simply were not credible.
There was other evidence that the CATS results were inflated.

As shown by Figure 9, remediation rates for recent Kentucky high school graduates entering the state’s postsecondary institutions were very high and improved only very slowly during the CATS era. The blue bars show that during the six year period covered in Figure 9, the drop in the percentage of students who needed remediation in at least one college subject area was only reduced, on average, by 1 percent per year. If this rate isn’t improved, it will take another 38 years to reduce remediation in our college system to zero.
More examples of CATS testing inflation are available at FreedomKentucky.org, the Bluegrass Institute’s wiki site.

To summarize, it became apparent over time that CATS painted a different – and far rosier – picture than other evidence about the performance of Kentucky’s public education system. As the results from CATS continued to implausibly rise, Kentuckians again began losing faith in their state’s public school assessment program, just as they had with KIRIS.

• **Writing portfolios for accountability: Failed assessment element during the CATS era**

While writing portfolios are generally considered today to be good classroom instructional tools, there are doubts about their value as part of a state assessment program. For example, Vermont, which actually launched its statewide writing portfolio assessment just before Kentucky did, has now abandoned portfolios for state assessments.

Kentucky’s writing portfolios were one of the major Progressive Education ideas included in the KIRIS and later CATS school assessment programs. These portfolios took students an entire year to create and were supposed to be collections of the final, best examples of students’ writing products following an extensive process of instruction and editing.

Under the accountability concepts adopted in KERA, which assumed the state assessment would be a major driving force for reform, it was inevitable that writing portfolios would be included in the KIRIS accountability program. The driving theory was, “What gets measured, gets taught.” The theory was that by including portfolios in the state’s assessments, teachers would be forced
to emphasize writing instruction. On the other hand, if writing were excluded from the assessment, the theory held that teachers would not teach it.

As with many other education ideas, the theory of using writing portfolios in the state’s assessment didn’t mesh well with the realities of trying to run a valid and reliable program free of cheating.

Over time, the writing portfolio program became hampered by a number of stifling rules created to prevent cheating and undue teacher influence on students’ submissions.

To avoid the appearance of cheating, many time-tested methods of writing instruction were specifically not allowed. Instead, students were left to guess about elements such as punctuation. A stilted set of rules only allowed teachers to circle a questionable punctuation mark and make a cryptic entry in the margin that something was wrong. Specific, explicit and detailed instruction on errors – and how to fix them – was not allowed.  

It was also widely reported that teachers and students spent inordinate amounts of time on the portfolio pieces. Students were required to rewrite articles beyond the point of diminishing returns. Questions abounded about whether the extensive rewriting resulted in teachers’ rather than students’ work winding up in the final graded pieces, and students reportedly developed strong negative attitudes about writing due to the endless revisions they were forced to make.

Instructional time for other important subjects was also limited in the grades where portfolios counted in the KIRIS or CATS assessments.

Over time, portfolios became very controversial in Kentucky. One of the most severe challenges to Kentucky’s writing program came with the release of the 2007 NAEP Writing results. Data in Table 1 will help explain some dramatic credibility issues this assessment created for writing portfolios.

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>White Percent Scoring At/Above Proficient</th>
<th>White Percentage of Test Sample</th>
<th>Black Percent Scoring At/Above Proficient</th>
<th>Black Percentage of Test Sample</th>
<th>Hispanic Percent Scoring At/Above Proficient</th>
<th>Hispanic Percentage of Test Sample</th>
<th>Asian/Pacific Island Percent Scoring At/Above Proficient</th>
<th>Asian/Pacific Island Percentage of Test Sample</th>
<th>American Indian Percent Scoring At/Above Proficient</th>
<th>American Indian Percentage of Test Sample</th>
<th>Unclassified Percent Scoring At/Above Proficient</th>
<th>Unclassified Percentage of Test Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>30</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>12</td>
<td>48</td>
<td>46</td>
<td>12</td>
<td>17</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>27</td>
<td>14</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

‡ Reporting standards not met.
Notice that in this assessment, California had significant percentages of minority students in all racial groups except American Indians and Unclassified. In California, whites formed a definite minority with just 31 percent of the tested sample, while in Kentucky whites comprised 86 percent of the tested sample. Still, whites in California badly outscored Kentucky whites, with proficiency rates of 38 percent versus 27 percent. Except for the Asian/Pacific Islanders, all other racial groups in California scored very low on the assessment.

Hispanics, which comprised nearly half of all the students tested (48 percent) in California in 2007, scored about the same as Kentucky’s relatively small population of blacks did. That is especially remarkable because in California, 21 percent of the entire tested sample was classified as English Language Learners, compared to only 1 percent in Kentucky. Hispanics in California actually matched the writing performance of Kentucky’s blacks (within statistical sampling error) even though nearly half of California’s Hispanics were still learning English and the assessment is only given in English.

Kentucky’s racial mix was far less diverse in the 2007 NAEP writing assessment. Except for whites and blacks, Kentucky had so few representatives from the other racial groups that their scores were unreported due to excessive statistical sampling errors that exceeded NAEP’s reporting standards.

Still, as can be seen in the table, despite their many racial and language disadvantages, all of California’s minority students basically matched, or seriously outperformed, Kentucky’s lone minority group across the board (proficiency rate differences of 1 point in this assessment are not statistically significantly different).

Overall, averaged across all students, the writing proficiency rates in the two states were statistically equal.

Thus, the 2007 NAEP writing results created a stunning indictment of the writing program in the Bluegrass State, an indictment made even more dramatic by the fact that California’s poverty rate, as measured by the percentage of students in the federal free and reduced cost lunch program, was equal to Kentucky’s.

Despite more than a decade and a half of intense concentration on writing, Kentucky’s use of portfolios in the state assessment program created a situation that tied writing teachers’ hands. The national assessment brought that message home in very stark detail when the Bluegrass State was barely able to match the performance of California and its huge disadvantages.

- **De-emphasis on grammar, spelling and punctuation**

Aside from the portfolio issues, there were other, related problems with writing instruction in Kentucky.

In the early years of KERA, a number of strange Progressive Theories about writing instruction were strongly pushed in Kentucky. At the same time the state tried to emphasize writing by
including portfolios in the assessment, spelling, grammar and punctuation were de-emphasized. Progressive Educators argued that students would learn the important technical features of good writing naturally by simply writing more.

The de-emphasis on writing mechanics was aided and abetted by portfolio scoring methods. Portfolio scoring used a “holistic” approach where mechanics barely counted and where expression of ideas was supposed to be key.

Of course, the mechanics were often so bad that it wasn’t possible to really understand what students were trying to say. But, as sample grading of written response problems would later show, graders used a lot of “creativity” in awarding students high marks for writing what – in some cases – was seriously factually wrong material (See side box).

(Side Box) Scoring in Kentucky’s assessments was often dubious.

For example, the sample 3-point student response shown in Figure 10 was included in a 1995-96 KIRIS Grade 11 Science “released item” package from the KDE. The original question asked students to discuss four or more variables that might make it easier or harder to suck liquid up a long straw followed by four examples from real life where liquids are moved against the force of gravity.

The student got credit here for incorrectly stating that a “large” (i.e. wide) straw would make it more difficult to suck liquid up the straw.

Anyone who has tried to suck liquid up a relatively wide straw such as found in fast food restaurants and compared that to the difficulty of sucking liquid up a thin straw such as often found in bars knows this student’s answer is wrong.

But the grader and question writers who put this annotated sample student answer together clearly did not know much, if anything, about science.
Figure 10
Sample Student Response to a KIRIS Question

KIRIS ASSESSMENT ANNOTATED RESPONSE
GRADE 11 SCIENCE

Sample 3-Point Response of Student Work

Students describes three real-life examples, i.e. the human circulatory system, water in home plumbing, and oil in a car motor. The fourth example is incorrect.

Student explains how each variable might affect the outcome.

This experiment may be affected by many different outcomes. Depending on the thickness of the liquid, the width of the straw, and the temperature, the experiment would be off. Another factor in doing this would be the temperature outside. If the temperature was high, the liquid would be likely to move faster than if it were cool because the warm molecules would be moving around faster and this would cause easier flow. Many situations in life also cause liquids to move against the pull of gravity other than just a straw. For example, your circulatory system in your body faces this problem everyday when transporting your blood. Water flowing from the ground to our houses everyday also faces this. Oil moving through your car motor also has to face this because the oil pan is at the bottom and the oil has to flow up into the motor when you start your car. Also, some rivers face this problem when they flow.

Student demonstrates an application of scientific ways of thinking and working and uses those methods to solve real-life problems.

Student demonstrates an ability to design investigations by clarifying the variables.

Student demonstrates a knowledge that objects change their motion only when a net force is applied. Laws of motion are used to predict and/or calculate the effects of forces on the motion of objects.
Ignoring writing mechanics was another example of KERA ideas driven by bad research and, as ACT English results show, resulted in little improvement in the subject. Figure 11 shows that ACT college entrance test scores for English for Kentucky’s high school graduates generally declined in the early days of the writing portfolios and didn’t even start to recover from the loss until after 2002 when pressure and feedback from colleges began to have impact.

Figure 11

(Note: Kentucky’s ACT trend line was lost in 2009 after Kentucky required ACT testing of all students; hence the 2009 data is not included above).

- **NAEP raises more questions about scoring of Kentucky writing**

The 2007 NAEP Writing Report Card was issued a year before Senate Bill 1 passed during the 2009 legislative session, which ended writing portfolios as part of the state’s assessment program. A comparison of Kentucky’s middle school writing proficiency rate from the NAEP with the writing portfolio proficiency rates clearly reveals inflation in portfolio scores. The disparity worsened following the very obvious further inflation in Kentucky writing portfolio scoring that occurred in 2007.
The credibility of CATS writing scoring was sorely tested by the nearly flat trend in NAEP proficiency rates.

- **Writing portfolio audits also showed scoring was questionable**

The many portfolio audits conducted by KDE over the years always showed that scores initially assigned by the teachers, who normally graded portfolios for students in their own school, were consistently inflated.

For example, in the 2007-08 audit -- the final one conducted -- one in four students initially awarded a top score of “Distinguished” got downgraded by two score levels to “Apprentice.” Among students originally awarded a “Distinguished” score, more than half deserved the score of “Apprentice” rather than the top CATS score.  

By 2009, it was clear to the Kentucky Legislature that using writing portfolios for accountability had failed. Adverse effects were intensified by arcane instructional rules for writing to prevent cheating that were an unfortunate side effect of using portfolios in the state assessment program. And, regardless of continuous attempts to control cheating, scores were hopelessly inflated, anyway.
• Termination of the CATS assessments

In recognition of the many obvious shortcomings in CATS that had appeared by 2009, the Kentucky Legislature voted in Senate Bill 1 during the 2009 General Assembly to terminate the assessment. The legislature also removed the writing portfolios from the assessment program in this legislation. This far-reaching bill requires the KDE to do things in a much better manner as the state rebuilds its assessment and accountability program.

As a first step, before a new assessment program is even fashioned, the KDE must revisit the state’s entire set of academic standards, updating them with the latest national and international theories and practices in education. These new, and potentially different and more comprehensive, standards will then guide the development of a new, hopefully much improved, testing program.

Summarizing on both KIRIS and CATS

• School and educator accountability mostly in name only

When all the dust settled from the KIRIS and CATS experiences, it was clear that neither program effectively identified many schools as problematic. This despite external evidence from NAEP, ACT and high college remediation rates showing Kentucky had extensive educational problems.

A handful of schools did get taken over at one point in the KIRIS process, and a few more got identified by CATS for state assistance. However, the numbers of schools involved were always small. Furthermore, not one educator ever lost a license to teach in Kentucky due to poor teaching skills. Despite a great deal of commentary about Kentucky wanting highly skilled educators in every classroom, the state’s education system today has no formal process to evaluate and remove tenured teachers who fail to do at least a minimally acceptable job of educating students.

In fact, it is even time-consuming and expensive to suspend or revoke the teaching certificates of individuals who have been convicted of serious felonies and clearly have no place in the classroom. However, no statute or regulation exists for removing incompetent teachers or those unwilling to perform. Plus, the Kentucky Educational Professional Standards Board informs us that no teacher in the commonwealth has ever had a certificate suspended or revoked for that reason.

In a similar vein, while some principals were removed by their superintendents, in a few cases it was shown that SBDM councils had the authority to hire the same person right back again as principal. Furthermore, even if removed, most principals had tenure rights and had to be retained in another position in the local school district.

Some of the real failings of KERA’s accountability program were vividly highlighted when the federal No Child Left Behind Act of 2001 came along.
For one thing, the failure of CATS to expose and penalize learning gaps was suddenly and dramatically revealed. NCLB resulted in many schools with significant performance gaps being highlighted for the first time. In 2009 NCLB reporting, 461 schools failed to meet goals, many due to performance gaps in reading and/or mathematics. The schools involved numbered hundreds more than the entire KIRIS and CATS program together ever exposed. For example, in the last year of CATS accountability, only 26 schools were identified in the “Assistance” level that entailed sanctions.

After 20 years of KERA, only a handful of schools have ever lost their SBDM authority. Since CATS testing began in 1999, only four schools (Western Middle School, Iroquois Middle School and the Southern Leadership Academy middle school in Jefferson County and the Holmes Junior/Senior High School in Covington) ever lost their self-governing authority due to poor performance. And, those schools didn’t lose their authority to self-govern until 2008. These schools were joined by 10 more schools in 2010 that were identified not by a Kentucky-developed accountability program, but rather under rules required by the federal School Improvement Grant program.

Meanwhile, performance gaps continue to haunt many Kentucky minorities including children of color, children in poverty, and children who learn differently.

The good news, however, is that the Kentucky Legislature has recognized many of these issues, and Senate Bill 1 passed by the 2009 General Assembly mandates a complete refashioning of the state’s education standards and the assessment program. If well done, this could be a strong step towards a much better Kentucky school system.

Supplemental Issues –

Other controversial KERA initiatives

Fuzzy math

One year before KERA’s passage, the National Council for Teachers of Mathematics (NCTM) released a document with a model math curriculum, “Curriculum and Evaluation Standards for School Mathematics” (1989). Heavily influenced by Progressive Educators, this document would spawn a large number of math programs that soon earned derisive nicknames like “Fuzzy Math” as parents and better trained math educators in both the public school system and colleges were exposed to the resulting curriculums. Certainly, the 1989 National Council for Teachers of Mathematics document was long on assertions but woefully short on scientific research to substantiate its main features.

After the KDE was reformed by KERA, department staff members adhered very closely to the 1989 NCTM standards. Fuzzy Math was heavily pushed in Kentucky in the 1990s. Vague concepts of “math understanding” were heavily stressed to the detriment of instruction in
classical math fundamentals such as multiplication tables and fractions. The KIRIS assessment questions also were heavily impacted by philosophies in the 1989 NCTM standards.

This de-emphasis of the “fundamentals” fit well with the Progressive idea of writing about math rather than insuring students could do math calculations leading to precise, reliable and correct answers. The vagueness of the math programs also tended to hide the serious deficiencies in the approach.

In KERA’s early years, the KDE actually defended such quaint notions as the idea that as long as the student grasped the mathematical concept, getting correct answers was not terribly important. There was a lot of negative discussion in Frankfort about the need for practice in math. Also, the pejorative term “drill and kill” was coined and frequently muttered by Progressive math proponents whenever more rational viewpoints surfaced concerning the need for practice to develop math skills.

After a decade of Progressive math instruction, severe problems with students’ performances were becoming strikingly evident. Students who had fuzzy math classes in elementary school were increasingly unable to master algebra and later ran into enormous problems in college. College math professors increasingly bemoaned the dreadfully poor preparation of incoming freshmen.

Eventually, college math professors around the nation led a charge on the NCTM and its 1989 curriculum document, which resulted in the document being revised around the turn of the new decade to become the “Principles and Standards for School Mathematics.”

That document was amended with the council’s “Focal Points” paper in 2006, which now very clearly indicates that instruction in such “basics” as multiplication tables and fractions is indeed a fundamental requirement in any respectable math curriculum. While some of the council’s spokespeople contended that even the 1989 document required the basics, most of the math programs that were built on the 1989 standards indicate that was not the common interpretation of the document. Also, state education departments’ interpretations of that document led to fuzzy math curriculums across the country, including here in Kentucky, while NCTM failed during most of the 1990s to address the deficiencies.

Unlike some other education fads discussed in this paper, fuzzy math didn’t firmly die in Kentucky due to a specific legislative action. However, the legislature has kept considerable pressure on math educators to get the commonwealth’s math program into better shape. One positive result of this has been the formation of a mathematics research center at Northern Kentucky University that opened in 2006. That center is to research and disseminate better ideas on math instruction, and it certainly isn’t pushing the idea that concepts trump process-mastery in math.

Still, there was a price to pay for these mistakes. The next figure shows that math performance for Kentucky’s dominant student group – white students – on the recent 2009 NAEP Eighth Grade Mathematics exam seriously lags the rest of the nation.
Whites comprised 85 percent of Kentucky eighth-graders in 2009.\textsuperscript{63} Nevertheless, only West Virginia had a statistically significantly lower eighth grade NAEP math score for white students in 2009 than Kentucky. A handful of other Southern states’ scores were not statistically different from Kentucky. However, white students in the vast majority of the nation got statistically significant better scores on this assessment than Kentucky.

\textbf{Figure 13} \textsuperscript{64}

\textit{White Student Grade 8 NAEP Math Results in 2009
States that Outscored, Tied and Performed More Poorly Than Kentucky
(States in Green Outscored Kentucky)}

Unfortunately, one thing hampering the effort to move Kentucky to a more robust math curriculum is the fact that a considerable number of its elementary school teachers were not strong math students when they were in school and college. It is possible for elementary school teacher candidates in Kentucky to complete college while taking only one mathematics course – called “Elementary Mathematics” – which is taught below the level of college algebra. Teachers with such low levels of mathematics ability can have great difficulty in reading research that can better inform their classroom practice.
Such shortcomings in teacher preparation will be challenging to overcome. Furthermore, many of these teachers pass along their unease with math to their students, creating a perpetuating effect.

Kentucky also continues to use dubious mathematics curricular programs in many schools. Remarkably, even as of 2009, only two school districts in the state are reportedly using the highly ranked Singapore Math program. This is particularly troubling because a North American edition of Singapore’s top-performing-in-the-world curriculum has been available for years. Kentucky claims it wants world-class standards but refuses to follow through on those aspirations.

In any event, fuzzy math is on the way out, albeit very slowly, in Kentucky.

**Whole language reading**

Another Progressive Educator fad idea – with notable parallels to the poorly researched fuzzy math concepts – involves early KERA era theories about teaching reading.

“Whole Language Reading,” pushed notions like kids could learn to read “naturally” by immersion as they read, similar to the way babies learn to speak by being surrounded by older speakers.

Typical of Progressive Education theories, the Whole Language concept was utterly hostile to the idea that practice makes perfect. Whole Language theorists also argued that teaching phonics was detrimental to children learning to read because it required practice and rote memory. In the early days of KERA there were concerns about the KDE’s de-emphasis of phonics instruction.

Progressives ignored the fact that English is a largely phonetic language, claiming it was unnecessary to teach text-decoding skills because children would pick them up in some vaguely described (and unscientifically researched) manner.

But as the results show, their approach – based on faulty research – led to bad conclusions and even worse results. Even many experienced teachers who had successfully taught with phonics approaches questioned the Progressive ideas about reading instruction.

Erosion of credibility for the Whole Language Reading philosophy accelerated in 2000 when the report of the National Reading Panel came out. This congressionally chartered panel was directed to examine reading instruction in considerable detail. The Panel based its findings on research that met at least minimal requirements for scientific rigor.

The National Reading Panel was not composed merely of educators, but also included representatives from the psychological and medical research areas, including individuals who understood the difference between scientific and other types of less definitive research in human-related studies.
Not surprisingly, the Panel found that early phonological instruction was crucial if many students were to become effective readers. The Panel determined that proper instruction in phonics gave students the word-decoding skills they needed to read quickly and efficiently. That, in turn, set the stage for more advanced stages of reading instruction where comprehension of those words also became important. However, if the first step – developing good phonics based text decoding skills – was bypassed, then long-term reading difficulties resulted for many children.

It was a chicken-versus-egg challenge. While Progressives were correct about the eventual need to develop student comprehension, they wanted to bypass and ignore an early, and crucial, step in the process – the “drill to skill” (not “drill and kill”) exercises of learning phonics.

The Progressives’ push to downplay phonics in KERA’s early years probably led to stagnation in Kentucky’s reading performance highlighted by the NAEP. Between 1992 and 1994, the state’s fourth grade NAEP reading score actually dropped trivially by a point.

There was no reading improvement until after the KDE issued new guidelines indicating it was again acceptable to teach phonics.

Also, ACT results show that Kentucky’s reading performance didn’t begin to improve until well after the KDE stopped pushing Whole Language Reading in the latter 1990s.
Ungraded Primary (formerly grades from Kindergarten to 3rd grade)

One of KERA’s most optimistic Progressive ideas is the concept of the ungraded primary school. This term does not refer to the absence of letter grades for students, but rather to the concept that children in the early school years will thrive if allowed to learn at their own pace. The Ungraded Primary theory holds that students should only advance to the fourth grade as their individual academic growth warrants.

A key Progressive argument to support the Ungraded Primary concept involves student self-esteem. The idea is that Ungraded Primary would eliminate student-to-student comparisons based on age that tend to make slower learners feel inadequate. Students would be grouped, and regrouped, according to their developmental levels, not age. Thus, it was hoped, students would not develop feelings of inadequacy.

Ungraded primary was a wonderfully wistful idea, but it unavoidably collided with the real facts of life. For one thing, the Ungraded Primary concept requires teachers to make incredibly incisive, on-going judgments of student performance. Even for highly experienced teachers, this is a very demanding, and ultimately quite subjective, process. For less experienced teachers, it’s simply impractical.

Plus, parents never really accepted the Ungraded Primary concept. They still expected their children to be promoted on time. Heated reactions often occurred when students were retained in Ungraded Primary.
One spectacular example of the tension Ungraded Primary could generate surfaced at Booker T. Washington Academy, a Lexington elementary school, in August 2007. The school’s principal departed under a big cloud of controversy fueled partly by allegations that she purposely held students back so they would not be tested. However, the Ungraded Primary concept actually required the principal to take such actions if students were not ready for the fourth grade. Neither the parents nor school district’s administration, however, supported this Ungraded Primary action. The principal lost her job for doing what she was supposed to do. Instead, she was accused of cheating on the CATS assessments for holding unready students back in class (a charge later denied by the Kentucky Educational Professional Standards Board).

- Performance of Kentucky’s primary and non-primary schools

Initially, KERA required all elementary schools to shift to a full, multi-age environment shortly after the act’s passage. However, continuing criticism from parents and educators led to a significant weakening of that provision as the first decade of KERA closed.

As a result, as the second decade of KERA was drawing to a close, there was a considerable range of implementation of the multi-age program in Kentucky. Some schools still faithfully followed the Primary concept, but about half had fully returned to a traditional grade orientation.

In 2008, the KDE surveyed all the elementary school principals in Kentucky concerning the degree of implementation of the multi-age environment in their schools. The survey results were summarized in an Excel spreadsheet, which this author obtained from the KDE. The existence of this data offered this author an opportunity to compare the degree to which Kentucky’s elementary schools currently implement Ungraded Primary to those schools' combined average math and reading proficiency rates on the KCCT.

- How Ungraded Primary implementation was categorized

The various levels of compliance with Ungraded Primary used in the Kentucky Department of Education’s 2008 survey and the number of schools in each category include:

"Single Grade" - Students are not multi-age mixed – 425 Schools (57.6 percent)

"Dual Graded (e.g., P3 and P4)" - Some multi-age mixing occurs, but not more than two age groups in a classroom – 102 schools (13.8 percent)

"Non-Graded Primary P1-P4" - Complete multi-age grouping of all students who would normally be in Kindergarten through third grade – 188 Schools (25.5 percent)

"Non-Graded P2-P4 and graded P-1" - Students entering school (formerly Kindergarten) held separate from multi-age environment of what was formerly considered grades one to three – 23 Schools (3.1 percent)
This author merged the Excel spreadsheet containing the Primary implementation information with another Excel document containing each school’s 2008 proficiency rates from the KCCT. The math and reading proficiency rates for each school were averaged together, and then the spreadsheet was analyzed to see how the degree of Ungraded Primary implementation corresponded to each school’s performance on the KCCT reading and mathematics assessments.

The results of the analysis are summarized in this graph.

**Figure 15**
Combined Math and Reading Proficiency Rate in Kentucky Elementary Schools Versus the Degree of Implementation of Ungraded Primary Concepts
2007-08 School Term

The first bar, "Single Grade Program" (shaded yellow) shows the combined average proficiency rate for reading and math was 71.64 percent in those schools which did not use any multi-age groupings in the 2007-2008 school term.

The next bar, "Average for all Multi-Age Models," (shaded pink), shows that overall across all the various multi-age groupings, the averaged math and reading proficiency rate was somewhat lower at 70.7 percent.

Following bars show the combined average math and reading proficiency rates for several different classifications of multi-age environments. Only one of those environments, “Dual Graded (e.g. P3 and P4) Only,” where students are multi-age mixed only with two ages maximum, shows trivially better performance than the totally non-multi-age “Single Grade Program.” That difference is only 0.57 percentage point, which probably isn't significantly different.
Based on the 2008 data, none of the various the multi-age environments currently used in Kentucky produces a notable advantage in academic performance over the traditional, graded structure. In fact, most of the implementations result in slightly lower performance. Therefore, the current rather lax enforcement of the requirement to implement multi-age environments in Kentucky is probably wise.

To summarize, Ungraded Primary is only loosely observed today in Kentucky today. A majority of the state’s elementary schools have returned to a classical grade-by-grade orientation. Finally, based on 2008 performance for Kentucky’s elementary schools, there is no notable advantage to this type of school organization.

(Note: An expanded discussion of this Ungraded Primary study is available in a freedomkentucky.org Wiki item, including a link to the detailed spreadsheet where the data used to assemble Figure 15 can also be found.)

**Dropout rates for school accountability**

Few outside of Progressive Education circles argue against the idea that a part of the accountability determination for high schools should be based on successfully graduating students. KIRIS and CATS both included a factor in the School Accountability Index for high school dropout rate performance. When NCLB came along, it explicitly required a graduation rate metric for accountability.

However, the details about how to actually measure this performance indicator proved problematic from the start of KIRIS.

Unfortunately, the dropout rate calculation chosen for use by the KDE was virtually impossible to audit and enforce. The dropout rates that resulted were almost immediately challenged as unrealistic. As the department continued to publish its dubious dropout rate data, considerable unease developed within the legislature that the figures were not trustworthy.

Legislative concern heightened once the federal NCLB law was enacted. ‘No Child,’ which required all high schools to report the closely related high school graduation rate statistic, focused more attention on the need for accurate data.

These concerns led to a hearing in October 2003 by the Kentucky Legislature’s Program Review and Investigations Committee concerning the accuracy of the state’s official dropout rates. Following testimony from the author of this paper and OEA, the committee voted to request the Kentucky Auditor of Public Accounts to officially audit the state’s dropout rates.

Completed in October 2006, the auditor’s report found significant errors in the state’s dropout data. Those errors created considerable overstatement of Kentucky’s dropout and graduation rate performance. Many more students were dropping out than the official data indicated.

Kentucky’s current dropout and graduation rate reporting program is now scheduled to be replaced by a much better graduation rate reporting system that is aligned with a standard federal
formula derived from a process developed and accepted by the National Governor’s Association in 2005. The current adoption schedule was announced in the KDE’s 2010 Nonacademic Data Report. Under this plan, Kentucky will first transition to a much more accurate interim graduation rate calculation in 2011. The KDE then starts to report a fully accurate graduation rate in 2014.

However, ample evidence, including the previously mentioned official audit, is available to show Kentucky’s dropout rate reporting has been inaccurate ever since KERA’s enactment.

Today, extensive federal research shows the most credible formula for computing high school graduation rates, which are now preferred to unreliable dropout rates, is the called the ‘Averaged Freshman Graduation Rate’ (AFGR). Figure 16 shows federally calculated AFGRs for the nation and Kentucky since KERA was enacted.

![Figure 16](image)

It is clear that Kentucky’s actual graduation rate performance has fluctuated significantly since KERA began. The best performance the state ever posted occurred way back in the 1993-94 school year, before much impact from KERA was felt by high school students who spent the vast majority of their school years in pre-KERA classrooms. As KERA began to have more impact on graduating students, the graduation rate fell very dramatically until 2001-02. The rate then began a recovery, but never attained the peak from 1993-94. Sadly, in the most recent two years of available data, federal statisticians report that Kentucky’s high school graduation rate is actually declining again, and in the most recently reported year slid back down below the national average.
Site Base (or School Base) Decision Making Councils (SBDM)

Unlike the situation in every other state, KERA’s School Based Decision Making Councils (SBDM) became the key governing group in Kentucky’s public schools. As such, Kentucky’s SBDMs created a profound change in school governance and implemented a vast shift in power in Kentucky’s public schools. Under KERA, teachers – not parents, elected school boards or superintendents – have primary control of virtually every important facet of schools’ operations.

KERA triggered a massive transfer of power from locally elected officials and their chosen professional executive (the superintendent) to the teachers. Essentially, school boards and their appointed education experts lost all meaningful control over schools to SBDMs, including most of the local district’s authority related to decisions involving principal selections, curriculum, building operations and final allocation of funding.

SBDMs are chaired by their school’s principal and by law must have as members a ratio of three teachers, elected by the teachers in the school to every two parents, who are elected by the largest parent organization in the school (normally the PTA/PTO). Because a majority vote carries in the SBDM, KERA handed professional staff the ultimate and wide-ranging control over what happens in schools.

By law, parents do not have enough votes to control SBDM decisions. And non-parent taxpayers have absolutely no say whatsoever about how their tax dollars are spent because even a vote for a local board of education member has no effective impact on what happens within each school’s SBDM.

Some very unsatisfactory impacts on school accountability resulted. Superintendents now rightfully claim they cannot be held accountable for poor performance because they lack authority to move in and fix problems in schools. The superintendent cannot even direct the school to fix curriculum problems or spend tax allocations more intelligently. In reality, SBDMs have most of the authority but little of the responsibility for what happens in Kentucky’s public schools.

Although superintendents technically can still fire principals doing so in practice often is unworkable. Thanks to a court decision favoring teachers against their superintendent, the SBDM can hire the same person right back again as principal even after a firing.

Even if what they do is in the best interest of students, superintendents get into trouble with the OEA or the courts when interfering with SBDMs.

For example, as this report goes to press, the superintendent for the Union County Public School District is in serious trouble after a recent OEA report claimed he committed multiple violations of SBDM laws. What makes this development especially tragic is that Union County schools drastically improved during the first year of the superintendent’s new policies. Between 2008 and 2009, Union County’s elementary school reading performance ranking jumped from 148th to 55th place among Kentucky’s 174 school districts. At the same time, the district’s elementary schools’ math performance made an even more-impressive jump, rising from 156th to 52nd
ranking among Kentucky’s school districts. But instead of placing the focus on the district’s progress, the state’s education bureaucracy is willing to place this improvement in jeopardy because of the superintendent’s willingness to put students’ needs above those of SBDMs.

If a superintendent’s plans on education prevail in Kentucky, it’s only because his SBDMs allow it to happen. Thus, the person who should be the most important education expert in the local school system has been effectively removed from any real authority. But the fact that 461 schools in Kentucky failed to meet NCLB goals in 2009, indicates SBDM policies are not creating acceptable student performance for thousands of Kentucky students.

Additional evidence of the shortcomings in the SBDMs surfaced recently when Kentucky began the process to accept federal School Improvement Grants. During the first step of this grant program, Kentucky identified 10 “Persistently Low-Achieving Schools.” These schools were among the state’s lowest performers on the most recent three years – 2007, 2008 and 2009 – of state math and reading assessments.

The Persistently Low-Achieving Schools each received an audit, formally titled “School Leadership Assessment Reports,” to determine which of four available school turnaround models provided for by state law should be used to try to improve student performance. Among the 10 audits conducted, auditors determined that SBDMs in seven of the Low-Achieving Schools did not have the capacity to turn the school around. The audits recommended that these schools should lose their self-governing authority. As mentioned earlier, an eighth Persistently Low-Achieving School, the Western Middle School in Jefferson County, lost its SBDM authority in 2008.

<table>
<thead>
<tr>
<th>School</th>
<th>Principal Has Capacity to Manage School Recovery?</th>
<th>SBDM Has Capacity to Manage School Recovery?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert Frost Middle School Jefferson County</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Western Middle School, Jefferson County</td>
<td>NO</td>
<td>Lost SBDM Authority Earlier in 2008</td>
</tr>
<tr>
<td>Fern Creek Traditional High School, Jefferson County</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Shawnee High School, Jefferson County</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Western High School, Jefferson County</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Valley High School, Jefferson County</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Leslie County High School</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Caverna High School, Caverna Independent</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>
Despite nearly two decades of operation, SBDMs in 80 percent of these low-performing schools failed to generate acceptable improvements in student performance. Also, six of the principals those 10 SBDMs selected for themselves were judged to be inadequate, providing interesting evidence that the SBDM process does not guarantee competent school leadership.

One thing is certain: The SBDM management model destroyed any possibility of taxpayers and parents holding schools accountable. Formerly, if a parent or taxpayer had an issue, they could bring it before either the superintendent or their locally elected officials. Those former change agents no longer have authority to act.

Certainly, it would not be correct to say that all SBDM councils in Kentucky are dysfunctional. Some schools are run by intelligent teachers, and in those cases there are good collaborative relationships between schools and districts.

However, there also have been appalling cases, like the one cited above in Union County, where significant antagonism has developed between the SBDM and other stakeholders like parents, local board officials and the superintendent. With teachers in firm control, the only guarantee is that the SBDM process can be relied upon to act in the best interest of schools’ adult staff, but not necessarily in the best interests of students and parents.

It’s now evident that the SBDM concept thoroughly destroyed the logical chain of command and accountability in education. This resulted in a chaotic system where, until very recently, only rarely has anyone in schools really been held accountable.

As this report goes to press, discussions continue about ways to change the SBDM program through new legislation that would allow superintendents and local school boards to regain at least some control over the principal hiring process. Unfortunately, this effort is being fiercely opposed by the Kentucky Education Association, and the outlook for this badly needed update to KERA is uncertain, but no less necessary.  

**Over-correction for funding inequity**

When KERA began, districts with low per pupil assessed property wealth generally also had schools with low total per pupil funding. Figure 17 shows the relationship in each Kentucky school district between total per pupil funding and the assessed per pupil property wealth that existed just before KERA’s enactment. The figure excludes the Anchorage Independent School District because it is a strong statistical outlier, providing thousands of dollars more per pupil in local tax revenue than all other districts.
As shown by the heavy black line in Figure 17, the slope of the regression line comparing district per-pupil funding to district property wealth was strongly positive before KERA. In other words, students in wealthier districts tended to have better funded schools.

**Figure 17**

By the 1998-99 school term, the relationship had reversed, as Figure 18 shows. Now, a student in a wealthy district was likely to attend a school that got notably lower per-pupil funding than a school located in a property-wealth-poor district.
Figure 19 shows the negative relationship situation between school funding and local property wealth continues today.
KERA resulted in Kentucky school districts with high assessed property values tending to get notably lower per pupil funding in their schools – a situation virtually unknown anywhere else in the United States.

One interesting example of the impact of this reverse inequity situation is found in Northern Kentucky where students in the high property wealth Boone County Public Schools sometimes now attend classes in crowded, portable buildings while schools in “down state” low property wealth districts now are frequently housed in beautiful, modern buildings.

Another distressing issue related to school funding in Kentucky is that, as the Bluegrass Institute found in its 2005 study “Bang for the Buck,” the relationship between increased spending and academic performance is a negative one.83

For example, more recently the author calculated the correlation between the 2007-08 biennium CATS accountability index scores for each school district and the 2007-08 total per-pupil revenue in each district. That correlation is a negative 0.136.84 In other words, districts where more is spent on education in Kentucky tend to produce lower academic performance. This situation is quite unlike that in virtually every other state and region in the country.

Thus, significantly increased spending in Kentucky’s poorer regions has not equalized educational performance. Despite spending more in its poorer school districts, the testing results still favor Kentucky’s districts with wealthier students.
**What’s coming next?**

Implementing Senate Bill 1 will first require revisiting all of Kentucky’s education standards followed by development of better curriculum and the creation of a new assessment and accountability program based upon those revised standards. Finally, after two decades of operating with an assessment built backwards that created questions before solid standards, Kentucky is taking steps in the proper order in establishing its new testing program. Hopefully, that more rigorous and logical approach will lead to improved assessment accuracy that will, in turn, result in a better academic performance by Kentucky’s students.

**What we need to guard against**

There is no question that occurrences during the past two decades offer important lessons for Kentuckians and the nation as a whole.

First, it is clear that even today, as Columbia Teacher’s College past president Arthur Levine warns, most education policy is not driven by solid, scientific research, due, in no small part, to the fact that there isn’t very much scientific research being conducted. Consequentially, fad ideas still predominate in education as too many educators simply don’t understand the poor quality of the research in which they place so much faith.

Any educator involved with curriculum development needs to read the reports from Levine and absolutely must be familiar with the scientific-research-oriented What Works Clearinghouse that was established by the US Department of Education several years ago. It’s unacceptable for schools to continue using programs that the clearinghouse has investigated and finds only marginally effective, or worse.

Second, instincts of many parents concerning things like fuzzy math and whole language reading have proved accurate. Given the general absence of high quality, scientific research, parents and legislators should not hesitate to question the education community’s recommendations that fail the common sense test. While educators claim “research shows …” the history of KERA tells us that education research often doesn’t show anything accurate.

A third problem to watch for is the possibility that yet another KIRIS/CATS assessment may emerge despite good faith efforts of legislators to finally get an accurate assessment program. For sure, if Kentucky’s education community tries to continue matrixed test models so they can continue to use excessive numbers of open response writing questions, then the public and legislature must cry “foul.” We will never get valid and reliable scores for individual students from a matrixed assessment.

Fourth, the education community needs to understand that its credibility has been damaged by past egregious mistakes. Educators continued to defend both KIRIS and then CATS long after ample evidence accrued showing that both programs were providing hopelessly inflated and untrustworthy data. Continually defending failing reading and math programs accomplishes
nothing more than to further erode public confidence in Kentucky’s public education system and ultimately winds up hurting children. It simply cannot be allowed to continue.

There are hopeful signs that some educators, including those in some teachers’ unions, are getting the message. Recently, the American Federation of Teachers agreed to significant changes to Colorado’s laws on teacher tenure and evaluation. The Los Angeles Times writes of this change:

“The law ties teachers' job reviews to the performances of their students on achievement tests.”

Others simply don’t want to change. During the 2010 session of the Kentucky General Assembly, the state’s teachers’ union, which is a branch of the National Education Association, fought bitterly to block a badly needed bill to implement charter schools. Aside from blocking students and parents from the educational enhancements that charter schools can provide, this selfish action also undermined Kentucky’s attempt to win $175 million in federal Race to the Top funds.

Continued spin in public releases of education data won’t help restore credibility to the system. With a public now much more attuned to what’s happening in schools, “stretching” data is not in the true best interests of public schools.

Fifth, it’s obvious that placing primary reliance on a testing program as a way to force change is ineffective. The NAEP shows our proficiency rates have edged up only slightly since KERA began. In no case do we see much more than about one in three students scoring proficient or more on these federal tests. Plus, even after 20 years of KERA, we continue to see similar very small proportions of our students scoring at or above the ACT benchmark scores that signal suitable preparation for college and careers. Kentucky needs to start looking at other options to spur educational change.

Allowing Kentucky’s teachers to continue with inadequate preparation is unacceptable. It is evident that Kentucky’s teachers more professional development in programs that do work, or at least show great promise of working, for students. Those who create curricula and select instructional programs need to be introduced to tools such as the What Works Clearinghouse that are starting to assemble credible research to assist and better inform that process. The new state education standards and the new tests will also require significant teacher development activity.

Resistance to really effective reforms cannot continue. More, and better, accountability is needed throughout the system. In part, this can be accomplished with school choice options, including charter schools, which create competition for students. Competition has spurred lots of creativity in the private sector, and educators should not remain close-minded to this idea that is now adopted in four out of every five states. Competition stimulates effective innovation – something that did not occur in Kentucky after KERA’s passage.
In closing

While KERA certainly was bold when it was enacted in 1990, it has not turned out to be brilliant. As it closes its second decade, some of Kentucky’s radical education reform’s main accomplishments have been to highlight a great number of fads that failed.

To be sure, KERA has sent us some messages. It is now up to all of us to learn from those lessons. Kentucky’s education and political leaders need to:

- Be strong and honest enough to step back and do major rethinking about what works best for children,
- Do more to develop the commonwealth’s teaching corps,
- Develop better assessments, and
- Create, utilize and credit better research to inform the entire process.

Our students need – and deserve – better.
Endnotes

1. Brothers’ comments made at the October 8, 2009 Kentucky Board of Education Meeting in Frankfort, Kentucky. An audiovisual recording of this meeting is online at: http://www.schoolfunding.info/states/ky/ROSEvCBE.PDF


4. See National Center for Education Statistics, The Nation’s Report Card: Writing 2007, (NCES 2008-468), Institute of Education Sciences, U.S. Department of Education Washington, DC., 2008. In this assessment, California’s whites significantly outscored Kentucky’s whites with proficiency rates of 38 percent versus 27 percent. California’s blacks statistically tied Kentucky’s blacks with proficiency rates of 13 percent versus 14 percent (Page 58). Thus, if California had the same demographics as Kentucky did on this assessment (virtually no racial groups except whites and blacks), the Golden State would have significantly outscored Kentucky for writing. Also see the text discussion titled “Writing portfolios for accountability -- Failed assessment element during the CATS era” in this paper for more details on California’s and Kentucky’s writing performance on this federal assessment.


10. The input data used to create Figure 1 include NAEP scale scores assembled from the Main NAEP Data Explorer, online at: http://nces.ed.gov/nationsreportcard/naepdata/. The state average per pupil funding comes from the Kentucky Department of Education’s “Annual Receipts and Expenditures Reports,” online at: http://www.education.ky.gov/KDE/Administrative+Resources/Finance+and+Funding/School+Finance/FInancial+Information/Annual+Financial+Reports.htm. The per pupil funding amounts were adjusted using the US Census Bureau’s CPI-U index, available online at: ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt.

11. The same data sources used to create Figure 1 supplied the information to create Figure 2.

12. The ACT Composite Scores used to create Figure 3 were computed by the author from annual data reports of Kentucky public high school ACT scores provided to the Kentucky Office of Education Accountability by the ACT, Inc. Available online at: http://www.freedomkentucky.org/index.php?title=ACT_Scores_in_Kentucky_by_Year.


15. Transformations: Kentucky’s Curriculum Framework, Volume I, was not issued until after the KIRIS tests were already underway.


It would be several more years before Kentucky published the first *Core Content for Assessment* documents. While these certainly were an improvement over the earlier, totally inadequate goals and outcomes material, they still had major deficiencies and would initially received very low rankings in analyses of state standards by the Thomas Fordham Foundation and the American Federation of Teachers.


Kentucky Department of Education, “Kentucky School and District Accountability Results, Accountability Cycle 3 (1994-95 to 1997-98), Background Information,” December 3, 1998, Frankfort, Kentucky. See page headed “KIRIS Accountability Cycle 3 – Performance Judgments.” Note that while schools in the lowest two categories under the “Growth Index” column can request Highly Skilled Educational Assistance, it isn’t mandatory. No takeover of the school is described, either.


Public school only ACT math scores were computed by the author from Excel data files obtained from the Kentucky Office of Education Accountability, which list the average ACT scores for each public high school in Kentucky.


Figure 6 from the Kentucky Department of Education, Frankfort, Kentucky. No longer on line.

This question can be viewed online at: [www.eddatafrominnes.com](http://www.eddatafrominnes.com). It is also available in Appendix C in a report from the Evaluation Center at Western Michigan University titled “An Independent Evaluation of the Kentucky Instructional Results Information System (KIRIS), Kentucky Institute for Educational Research, Frankfort, KY January 1995.


The Evaluation Center at Western Michigan University, *An Independent Evaluation of the Kentucky Instructional Results Information System (KIRIS)*, Kentucky Institute for Education Research, Frankfort, Kentucky, January 1995.
37 Garrett, Robert T, and Schaver, Mark, “KERA test study has questions, no shocks,” The Courier-Journal, Louisville, Kentucky, February 17, 1995. (Can you fix this indenting problem?)
38 Hogan, Roseann, et. al., An Evaluation of The Kentucky Education Reform Assessment System (KIRIS) and Academic Performance at the University of Kentucky, University of Kentucky, Lexington Campus, March 1, 1995.
KIRIS proficiency rates calculated from the sum of the percent of students rated Proficient and Distinguished in Kentucky School and District Accountability Results, Accountability Cycle 3, (1994-95 to 1997-98), State and Regional Scores by Subject, Kentucky Department of Education, Frankfort, KY 1998.
48 Graph developed from Kentucky Council on Postsecondary Education data provided to the author. Data scheduled to be available on line in the Council on Postsecondary Education web site by the time this report is published. http://cpe.ky.gov/.
The Bluegrass Institute created a short video, using examples from Kentucky Department of Education training videos, to demonstrate how writing portfolio rules overly constrained effective writing instruction. That video is online at: http://www.youtube.com/watch?v=R7maF1AhyJM.

Table 1 was assembled with the NAEP Data Explorer. On line at: http://nces.ed.gov/nationsreportcard/naepdata/.


KIRIS GRADE 11 Science, 1995-96 Released Item, was formerly online at the Kentucky Department of Education.

Public school only ACT English scores were computed by the author as weighted average scores from individual high school score in Excel data files obtained from the Kentucky Office of Education Accountability, which list the average ACT scores for each public high school in Kentucky.


Cook, Cathy, “The Significance of the NCTM Standards to the Pathways Critical Issues in Mathematics,” North Central Regional Educational Laboratory. Online at: http://www.ncrel.org/sdrs/areas/issues/content/cntareas/math/pa0.htm


This figure was assembled using the NAEP State Comparisons Tool. Online at: http://nces.ed.gov/nationsreportcard/statecomparisons/


Public school only ACT reading scores were computed by the author from Excel data files obtained from the Kentucky Office of Education Accountability, which list the average ACT scores for each public high school in Kentucky.


For an expanded discussion of the School Improvement Grant process, see the Kentucky Department of Education’s web site article, “LEADERSHIP ASSESSMENT,” online at: http://www.education.ky.gov/KDE/Administrative+Resources/School+Improvement/Leadership+Assessment/.

Data in Table 2 was assembled from the individual Leadership Assessment Reports. Access to those reports with a link available at: http://www.education.ky.gov/KDE/Administrative+Resources/School+Improvement/Leadership+Assessment/.

A discussion of a highly contentious legislative session where the principal hiring process was debated can be found in “Legislative unhappiness over negotiation process a big factor in defeat of principal hiring bill,” Kentucky School Boards Association eNews Service, Frankfort, Kentucky, March 9, 2010. Temporarily online at: http://www.ksba.org/news/article/legislative-unhappiness-over-negotiation-process-a-big-factor-in-defeat-of-

Data sources used to create Figure 17: Per Pupil Property Wealth by District and Total Revenues By District from Kentucky Office of Education Accountability, *2000 Annual Report*, Frankfort, Kentucky.

Data sources used to create Figure 18: Per Pupil Property Wealth by District and Total Revenues By District from Kentucky Office of Education Accountability, *2000 Annual Report*, Frankfort, Kentucky.


Data sources used to calculate the correlation include: Per Pupil Revenue from the Kentucky Department of Education’s “2007-08 Receipts and Expenditures Report” Excel Spreadsheet on line at: