

KERA UPDATE

January 1999

#32

What's Inside:

The rise and fall of KIRIS.....Pg 1	
-Scores never made sense	
-Inclusion confusion	
-Math portfolios failed early	
-Performance events failed	
-Key personnel resign	
-Technical test management inept	
-Cheating went up	
-Testing contractor fired	
-Legislature votes major audit	
Kentucky is important (Box).....Pg 2	
What KIRIS showed.....Pg 3	
What other tests showed.....Pg 3	
-ACT	
-NAEP	
-CTBS-5 (Terra Nova)	
-SAT	
-Armed Forces Entrance	
Even Pros mess up with NAEP (Box).Pg 4	
Public school exodus? (Box).....Pg 4	
Remediation rates in Kentucky.....Pg 5	
Show me the money.....Pg 6	
Kentucky dropouts and retentions...Pg 6	
Changing definitions.....Pg 6	
The Replacement: CATS, or CATStastrophe?Pg. 6	
Fooling with figures (Box).....Pg 7	
Lessons for others.....Pg 7	
What quality reports said.....Pg 8	
Articles on cheating.....Pg 12	

Misinformed, Misled, Flawed

The Legacy of KIRIS, Kentucky's First Experiment with OBE Testing

THE RISE AND FALL OF KIRIS

Kentucky's KIRIS assessment is dead. KIRIS, the testbed for the New Standards assessments, the hope of radical reformers for new wave OBE testing, and a possible model for the Clinton "voluntary" national tests, is dead.

After a tremendously troubled 7 years of testing with a program that had an estimated life cycle opportunity cost around a billion dollars, Kentucky's first experiment with radical outcome based assessments has come to a close.

The end was mandated in early 1998 during the Kentucky Legislature's regular session. Liberal lawmakers, overwhelmed with a mountain of evidence, capped by the presentation of a formal audit, finally gave in to pressure from interest groups ranging from parents and teachers to a growing number of conservatives within the legislature itself. All had come to recognize many serious problems with this assessment.

— Scores Never Made Sense

From the outset, KIRIS scores were never credible. After the second year of testing in 1993, many of the best public schools in Kentucky found themselves faced with a very real prospect of takeover by the Kentucky Department of Education. This problem with scoring anomalies would continue throughout the lifetime of KIRIS. The most telling incident occurred in 1996 when Anchorage

Independent, the strongest and highest funded school district in the state, found itself rated in "Decline." Even though Anchorage had the highest KIRIS raw scores of any school system in Kentucky, after the strange formulas of KIRIS were applied, Anchorage found itself rated near the bottom of the heap. Anchorage became a lightning rod for critics of the assessment. The "Anchorage Syndrome" was a frequent topic during legislative deliberations.

—Inclusion Confusion

From the outset, reformers demanded that the Kentucky education program include children with learning disabilities in regular classrooms and in the assessment program. The goal was laudable — preventing these children from being dumped out of the education system.

However, the actual practice of inclusion proved daunting. In a July 1997 report, researchers made some startling revelations. Because they had received extra help ranging from teachers reading questions to actually rephrasing questions (something strictly prohibited for regular students), it turned out that learning disabled children actually outscored regular students on some academic portions of KIRIS, especially science!

The report also raised some strong questions about the amount of help being given to learning disabled students. So much extra help was being given that the report strongly hinted that this program

KENTUCKY IS IMPORTANT

Virtually every idea in education reform has been in progress in Kentucky since 1990. Multi-age elementary schools, inclusion of kids with disabilities in regular classrooms and testing, high stakes accountability, financing equity attempts, site base governance — all the current fads can be found in the Kentucky Education Reform Act of 1990 (KERA).

Radical education reformers constantly cite Kentucky as a positive example to encourage similar changes in virtually every other state in the country. Unfortunately, the material presented by these reformers is usually seriously biased if not completely fallacious. The real impact of these reform attempts is important, but getting at the truth is often frustrated by people with set personal agendas.

The Kentucky Commissioner of Education is highly influential in at least three important national education organizations: (1) The committee developing the Clinton National Tests (just cited again in President Clinton's 1999 State of the Union address), (2) the National Commission for Accreditation of Teacher Education (which is rapidly gaining control of accreditation of all teacher colleges in the country), and (3) the Council of Chief State School Officers (which has received significant federal money to do research and reporting).

In addition, the Governor of Kentucky is a board member of the National Education Goals Panel (Which uses taxpayer dollars to extensively distribute its education reports across the nation).

Thus, there are many pathways for radical ideas from Kentucky, regardless of merit and proved effectiveness, to be pushed into other states.

So, it's important to understand the real status of education in Kentucky. For, while Kentucky's population is small, the impacts of KERA have national implications.

might be a way for schools to jimmy their KIRIS scores. If so, this certainly added to the problems of reliability and accuracy found by other researchers.

{*The Assessment of Students with Disabilities in Kentucky* - See the Reports Section }

— Math Portfolios Failed Early

A highly controversial part of the original KIRIS program was a specialized set of written portfolios in mathematics. These ran into withering criticism from the outset for taking too much time away from more important aspects of math instruction.

In 1996, the Legislature voted Math Portfolios out of the assessment program following extensive protests from parents and math teachers.

Then, in the fall of 1997, a new consortium of college math professors and business leaders from Northern Kentucky issued a strongly worded report about both the math portfolios and the general quality of math training of recent high school graduates. Citing Freshman college math remediation requirements running up to 70% at Kentucky colleges, the group called for a number of important changes in the public school math program. One of the key recommendations, the math portfolios should not be brought back — ever!

— Performance Events Failed

One of the most radical parts of the KIRIS program was the use of performance events for assessment. Performance events involved giving a problem to small groups of students to work together. Each child then wrote an individual report and these were later graded by the testing contractor.

Performance events proved so unreliable that the KY Board of Education was forced to formally vote them out of the program in August, 1996. Performance events died because of very unusual results for 8th Graders — on a scale of 0 to 140, the average 8th Grade score worked out to only 3 or so.

The sudden deletion of performance events shook public confidence in KIRIS. It also triggered a lawsuit from a school that

would have scored much higher if the events had counted. That suit was in progress when the actions of the 1998 legislature made it moot. In essence, the school won through legislative relief.

— Key Personnel Resign

The first Kentucky Commissioner of Education resigned in 1995 as the first highly critical reports on KIRIS were being issued. More resignations followed in later years from both the Deputy Commissioner and Associate Commissioner for Assessment and Accountability. Other key personnel within the testing division also left Kentucky. It is noteworthy that the current deputy for testing is not known for technical testing expertise but does have excellent political connections.

— Technical Test Management Was Inept

The company originally hired to create KIRIS clearly wasn't up to the task. From the beginning, the complex problems of creating and grading open response questions were too challenging. Question creation and grading soon became suspect.

In addition, the company was unable to provide reasonable turnaround times. At first, test results took nearly a year to compile. Even in the last year of testing this company handled, schools didn't get test results from the April tests until well into October of the following school year.

This lengthy delay was never improved. In fact, the turn-around for the last year of testing, in 1998, actually decayed again (results were not out until December).

—Cheating Went Up

One of the big hush words with KIRIS is "cheating." There was always lots of whispering, but education reporter Lonnie Harp of the *Lexington Herald-Leader* brought this into the open on July 10th, 1997, in a news article titled "State slow to check school-test cheating charges" (included, with permission of the *Herald-Leader*, at the end of this *Update*). Mr. Harp showed that of 151 allegations of cheating in the first four years of KIRIS, the Department of Education had turned all but 11 back to the local districts for investigation.

Harp's expose opened a fountain of inquires and brought irresistible pressure on the Department of Education to tighten up cheating inquires. By December, another *Herald-Leader* reporter wrote that, in just five months, the new investigation program had looked at 24 more allegations of cheating, and findings of impropriety were made in 11 (See "KIRIS scores to be lowered at 11 schools" also reprinted with permission of the *Herald-Leader*). Thus, the monthly filing rate for cheating allegations at least doubled, and the conviction rate jumped from about one in three to nearly one in two. Clearly, cheating was a problem with KIRIS, and lackadaisical cheating investigation may have added to some of the inaccuracy of KIRIS results.

— **Testing Contractor Fired**

The beginning of the end for KIRIS came when the original contractor made a disastrous error in the 1996 KIRIS scores. The mistake wasn't discovered until April of 1997, a year after the test had been given. At this time, a Kentucky Department of Education contract employee (who also quit recently) discovered that every elementary and middle school in Kentucky had received erroneously low scores.

The discovery came too late. A \$25 million reward pot had already been divided and distributed to schools based on the incorrect scores. It cost the taxpayers \$1.5 million extra to make good rewards to schools and teachers that should have received a higher percentage of the original reward pool. It also cost the technical contractor the richest testing contract in public school history.

— **Legislature Votes Major Audit**

When the news broke about the testing contractor's firing, the Kentucky Legislature was already considering a limited audit of certain aspects of KIRIS following the 1996 collapse of the performance events. Announcement of the latest KIRIS problems added considerable tension to the discussions. The audit was quickly expanded to cover virtually all aspects of KIRIS.

In addition, Department of Education management was added to the list of audit topics.

A professional auditing firm began the work in the Summer of 1997, and the report was released at the start of the legislative session in January 1998.

To no-one's surprise, the audit found serious problems in KIRIS, and the audit team basically stated it would be best to start an assessment program all over again with new baselines. This was necessary because all KIRIS testing to date was loaded with uncorrectable errors.

WHAT KIRIS SHOWED

KIRIS results for the "Third Biennium" were issued in December, 1998. These showed impressive gains for schools. Grade 4 scores rose from 35.7 in 1993 to 48.8 in 1998 (a 37% increase). For the same years, middle schools rose from 37.4 to 51.3 (increases of 23% and 47% respectively). Were these phenomenal score increases credible? A brand new report from RAND looks at that question. The RAND conclusion was that inflation on KIRIS was "considerable" {RAND Report MR1014 (see the report quotes section for ordering information)}.

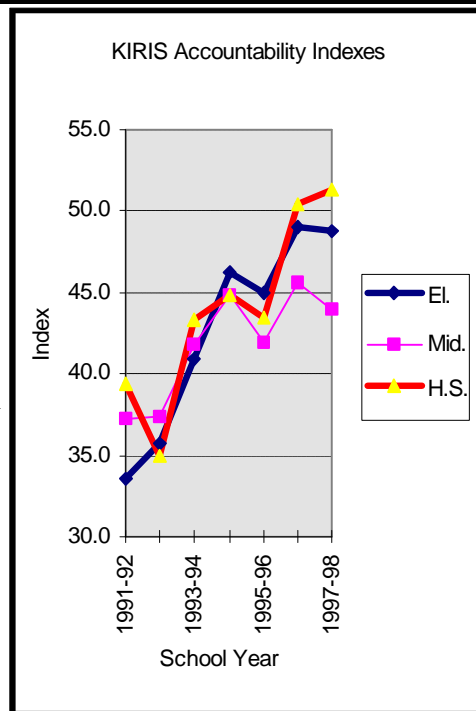
WHAT OTHER TESTS SHOW

— **ACT Testing**

The graph at right shows Kentucky's public high school ACT performance relative to the national average. It is evident that Kentucky's ACT performance has declined slowly since KERA began impacting the classroom around 1992.

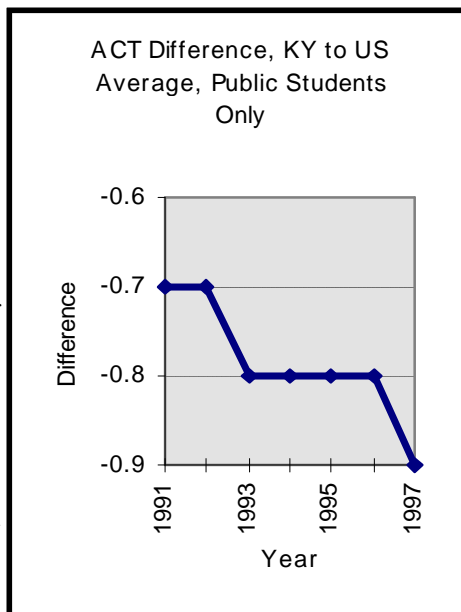
The ACT situation becomes clearer when two subgroups of Kentucky ACT takers are examined. From 1991 to 1998, ACT scores for Kentucky's top students, those who take the full recommended core of college preparatory courses, declined across the board since Kentucky's reform began. Drops varied from a low of -0.1 for math to a more disturbing -0.6 for English and reading.

On the other hand, ACT scores vary widely for kids who didn't take the full core course load but still decided to try for



college. For these children, Math is up 0.6 points; but, in a big surprise for a highly writing-oriented reform, English scores for this group dropped 0.1 points in eight years. Coupled with the very dramatic English drop for the kids who took full core course loads, serious questions are raised about what Kentucky is really accomplishing. {Data Source: *ACT Assessment 1998 Results, Summary Report, Kentucky, & KY School Boards Association Advocate, Oct 1995*}

— **NAEP Testing**



EVEN PRO'S MESS UP WITH NAEP

One of the classic mistakes made with NAEP is to use the "Achievement Level" scores (Below Basic, Basic, Proficient, and Advanced). These are derived as an after-product from the more accurate NAEP "Scale Scores." The problem is that setting the "cut scores" for the achievement levels is highly controversial, so controversial that in the 1996 NAEP Math Report Card, the National Center for Educational Statistics (the NAEP administrator) declared the achievement levels to be only "developmental." NAEP Achievement Levels "...should not be interpreted as statistically conclusive." {NAEP 1996 Mathematics Report Card for the Nation and the States, Pg 42}

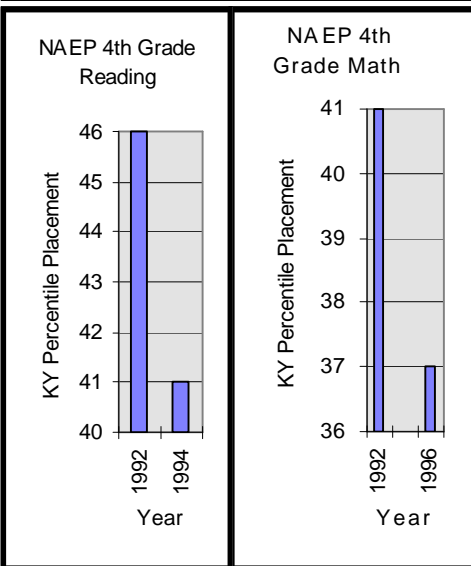
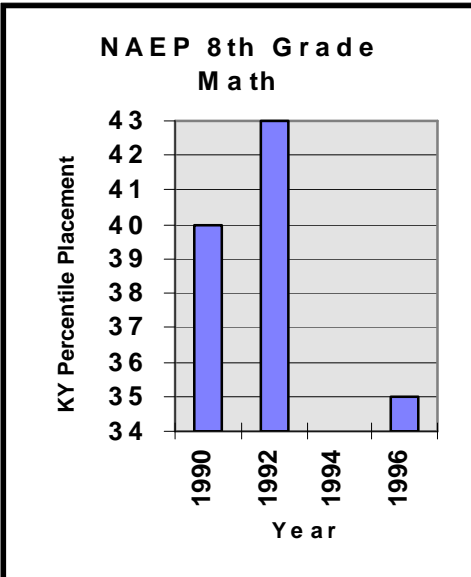
Research based on data that is only developmental and not statistically conclusive is clearly misleading. But, reports stretching from *Quality Counts* to the *National Education Goals Report* continue to make this precise mistake with NAEP data.

There has been a tremendous amount of misleading information published about the national assessment of educational progress. Kentucky most definitely has been part of that skewed picture. And, NAEP is really important, because KIRIS was supposed to be like the NAEP.

But, when you cut through all of the baloney, one of the more interesting ways to examine the NAEP is to simply see how Kentucky stacked up from year to year in the state score lists for NAEP scale score results (Don't use NAEP *achievement level* results for this!!! Even NAEP reports say the *achievement levels* are not statistically conclusive [see box story]).

Using the scale score listings, Kentucky lost percentile rank among the states on every NAEP test that's been given more than once (see graphs above).

The 8th grade NAEP math is interesting for additional reasons. This test was given 3 times. The first two times occurred before Kentucky's reforms had much impact at the classroom level. And,



Kentucky did nicely. By 1996, however, reform was definitely impacting Kentucky classrooms. The results shown on the graph above indicate that these impacts, if NAEP is the measurement, were not good. —CTBS-5 (Terra Nova)

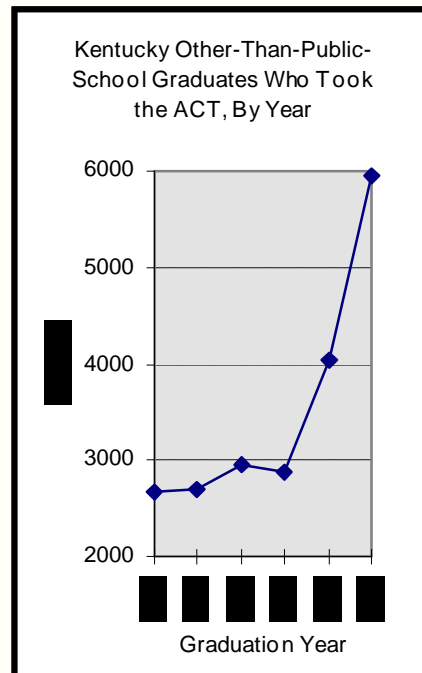
Kentucky was a launch state for the new CTBS-5 test, more commonly called the Terra Nova outside Kentucky. While Terra Nova has been the source of much controversy, it is certainly a better basic skills test than KIRIS.

In 1997, Terra Nova was given statewide in only four other states: Indiana, Wisconsin, Nevada, and Maryland. The Maryland scores had processing problems and were not made available. Compared to the other three states, Kentucky received the lowest CTBS-5 scores across the board in all areas tested. This included reading,

PUBLIC SCHOOL EXODUS?

There was a new surprise lurking in the 1998 ACT data — strong evidence that large numbers of college-aspiring students were turning away from public schools in Kentucky. From 1996 to 1998 the number of ACT test takers who had not graduated from a Kentucky public high school positively skyrocketed, up more than 100%! This raised a host of questions about possible loss of faith in the KERA-influenced public schools and the credibility of enrollment figures for homeschool and private school students.

To put this in perspective, the official KDE non-public school senior class enrollment for 1998 was only 3,600 students. But, nearly twice as many took the test in a status of other-than-public school. Where did the extra 2,400 kids come from?



language arts, and math and is true regardless of grade level.

Kentucky even scored lower than Nevada, a state not known for strong education programs!

A second year of Terra Nova testing was completed in 1998. Usually, the second year of tests like Terra Nova produce a phenomenon known as the "Lake Wobeggon Effect," named for the legendary place of Garrison Keellor fame

where all students were above average. With Lake Wobeggon effect, the scores should rise regardless of real education progress. In Kentucky's case, the scores stayed virtually flat. Third and sixth grade scores went up a scant 0.1 NCE, while high school actually declined 0.1 points.

—SAT Testing

You have to be careful about drawing too many conclusions about Kentucky from the SAT because only the top 12 percent of Kentucky students take this test. Still, the results are interesting, and show that on SAT, Kentucky's performance has not been terribly exciting. Compared to the national average, Kentucky's math SAT fell 2 points from 1992 to 1997, and the Verbal declined an identical 2 points.

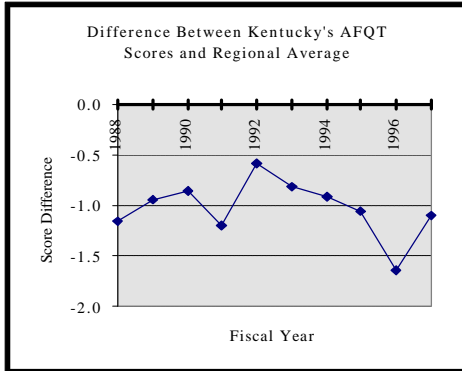
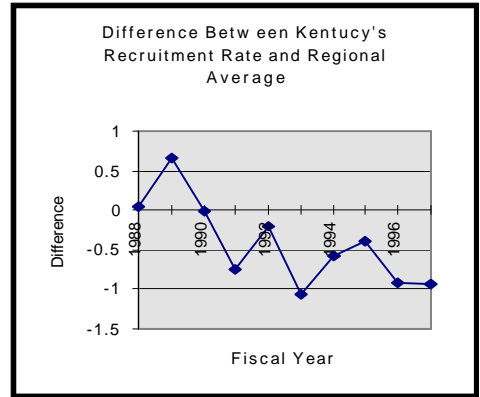
—Armed Forces Testing/Recruiting

The armed forces test applicants, and this data is also available annually. Here, the picture for Kentucky kids, like the case of the ACT, isn't good. As shown below, Kentucky's AFQT scores have fallen since 1992 compared to the regional average score. Recruiting is also down.

Keep in mind that this is a true employment test. So far, private employers in Kentucky have not shared results of their tests. However, newspaper articles indicate employers are having trouble finding qualified applicants, at least

in the "Golden Triangle" area of the state in Northern Kentucky. {Data Source: *Dept. of Defense*}

REMEDIATION RATES IN KENTUCKY The past year brought further indications that the graduates of Kentucky's Ungraded Primary (a multi-age implementation of former grades K through 3) could be arriving in 4th Grade with inadequate skills. Following a trend that began in 1993-94, one of two Kentucky 4th Graders needed remedial/



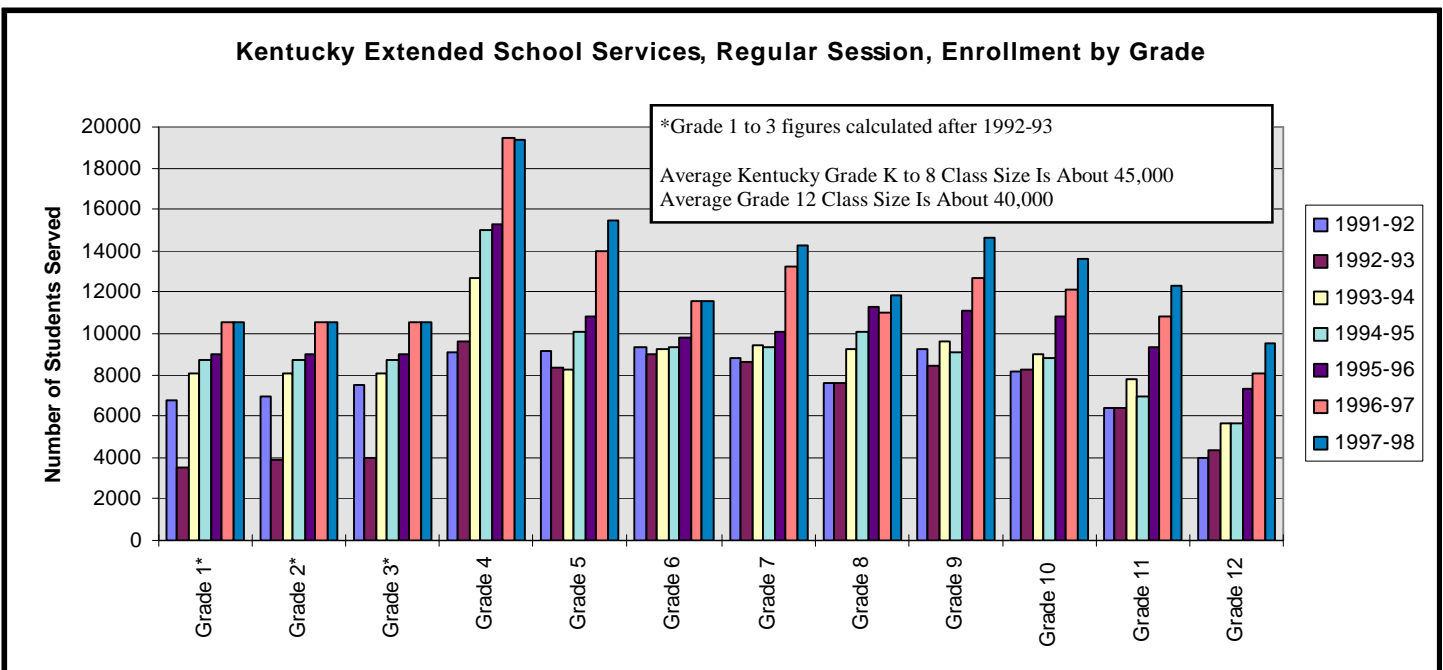
can match Kentucky's enormous rates of tutorial and remedial help. {Data Source: *KY Dept. of Ed.*}

SHOW ME THE MONEY Say what you will about the various parts of KERA, one thing is certain — it is expensive! A recent study by Dr. Lawrence O. Picus indicated Kentucky's total "opportunity costs" for this test run over \$120 million per year. The cost per student was an astronomical \$848 per year! And, that was the *low* estimate. Dr. Picus' high estimate was over twice as large, \$1,792! For reference, testing with nationally-normed standardized achievement tests runs about \$5 per student. {*Alternative Assessment Programs: The True Costs* - See Reports section }

tutorial help in Kentucky's Extended School Services program during 1997-98! The products of Ungraded Primary must be questioned when they need so much remediation in the year after they leave this very controversial program.

The graph below gives a quick idea about where the help from this program was directed, and redirected, during recent years. It is doubtful that any other state

Overall, the Kentucky education bill rose from \$2 billion to \$3 billion in just 5 years. {Data Source: *KY Office of Education Accountability*}



KENTUCKY DROPOUTS AND RETENTIONS

One of the most important “non-cognitive” indicators of school performance is the dropout situation. Unfortunately, the Kentucky Department of Education has changed the dropout formula not once, but twice, since KERA was enacted. This destroyed any possibility of making easy comparisons to the pre-KERA period. Still, it is obvious that the Kentucky dropout picture was improving before KERA began in 1990. Claims of success for KERA must be carefully weighed when this type of hard data is examined. {Data Source: *KY Dept. of Ed.*}

Another declining situation in 1995-96 was the number of high school children retained in grade (those not promoted or graduated). There was a very rapid

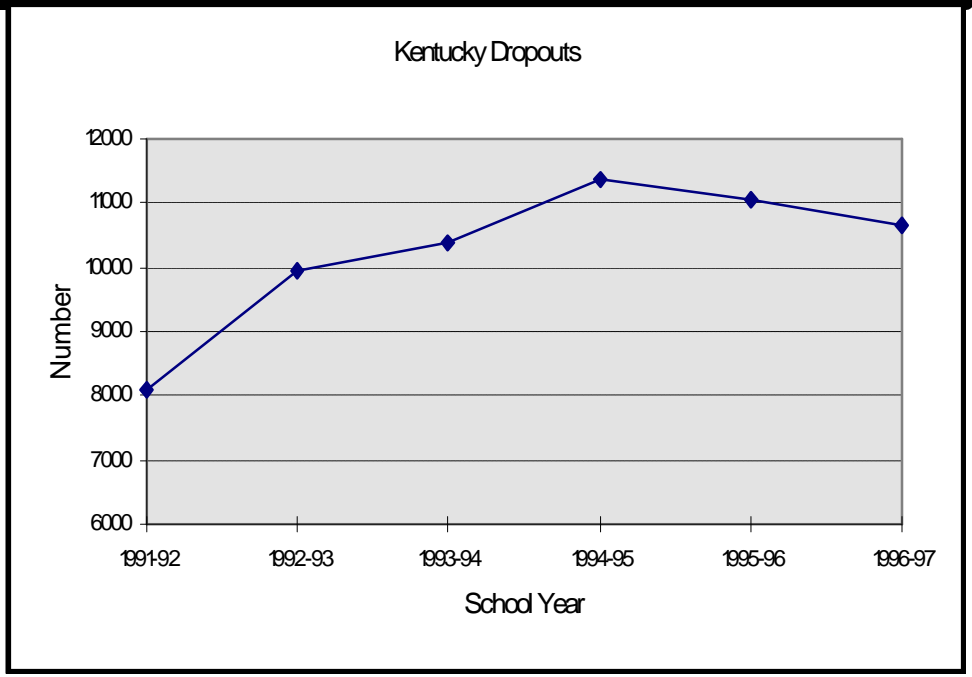
Kentucky’s education bill rose 50% in just 5 years!!!! This “STUFF” isn’t cheap!!!!!!

increase in retention in 9th Grade. In fact 9th Grade retention is the highest ever in more than a decade. Retention in all other grades is also running at or near a high since the reform act was passed in 1990.

One added note: The dropout graph on this page stops at the 1996-97 school year because the KY Department of Education now delays reporting these figures for a year. The delay allows KDE to report KIRIS scores earlier, but it is obvious the timeliness of dropout reporting actually got much worse.

CHANGING DEFINITIONS

Kentucky’s radical education reformers change definitions on a moments notice whenever it suits their purposes. A good example of this occurred during the development of KIRIS. When it became obvious to all that the KIRIS performance events were a failure, the Department simply changed the definition of a performance based test. The original definition had stressed real performances on some sort of project; but, the new definition simply says that anything that involves open response writing is a performance event.



How convenient, and misleading. You see, the Kentucky Education Reform Act stressed performances for assessment. In fact, KIRIS was to be primarily “performance-based.”

When the performance events failed, something had to be done, and it was. Never mind that the law said the Kentucky assessment must be primarily performance-based. Just change the definition.

CATS or CATStastrophe?

THE PENDING

REPLACEMENT As this *KERA Update* was going to press, the Kentucky Board of Education was hard at work developing a replacement for KIRIS, to be known as the Commonwealth Assessment and Testing System, or CATS. While the CATS configuration has not been finalized, the major question is going to be if CATS will be different enough from KIRIS to avoid the many problems that plagued the older assessment.

CATS continues to use tricky open response questions, and portfolios will still be a significant grading element, as well.

Confusing the issue, however, CATS bows to strong political pressure and will also add multiple-choice elements to the mix (despite strong resistance within KDE). Nationally-normed standardized testing also continues, but with the controversial

Terra Nova. As more states come to use the Terra Nova, a better idea of Kentucky’s real progress may never-the-less become possible to gauge.

Only time will tell if CATS can overcome the problems with KIRIS. However, with the same ideology and mindsets still firmly in place in KDE, the prognosis must be guarded, at best.

FOOLING WITH FIGURES

Early in the program, the Kentucky Department of Education tried to claim that KIRIS “correlated” well to other tests such as the NAEP and the ACT. In the August 1994 *Kentucky Teacher*, a Department employee published correlation data that he claimed showed good relationships with other tests. The “correlation” factor used, Pearson’s correlation coefficient, has been frequently used by educators to claim a great variety of things about tests. Readers should be aware, however, that the correlation coefficient, often simply called the “correlation,” can be a very misleading number, often misused to imply something far more important than is actually the case.

Here is a quick example of how the correlation coefficient can mislead. Take a look at Table 1 below. This shows that 10 students took 5 different tests. They got identical scores on Test 1 and Test 2a. Then, student performance uniformly falls

on Test 2b, and it falls uniformly again on Test 2c and 2d. The average scores on the test 2 series also fall rather sharply — well over 2 full points from a 5.5 on Test 2a to just 3.1 on Test 2d.

Now, look at the “correlations” for Test 1 as compared to Tests 2a through 2d. Notice there is hardly any change. That is because the best student is still the best, the weakest still the weakest (although there are increasingly more students with a 1 point score on test 2b to 2d).

In fact, for social sciences, these correlation figures (all well above +0.9) would be considered very impressive, even the figure for Test 2d.

Never-the-less, while the correlations remained high, it is very apparent that overall student performance has fallen dramatically from Test 2a to Test 2d.

So, while high correlations mean that the best students do the best on both tests, citing only correlations in a report about these tests would hide the fact that actual student performance is falling dramatically on the Test 2 series.

The Kentucky Department of Education’s report on KIRIS correlations made precisely this sort of misrepresentation. In addition, a number of basic principles and

cautions for use of the correlation coefficient, which can be found in any introductory text on statistics (such as development of scatter plots), were also omitted in the *Kentucky Teacher* analysis.

Fortunately, KIRIS’ true “correlation” to other tests has now been extensively explored by highly qualified researchers in *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)* by Daniel M. Koretz and Sheila L. Barron of RAND (see the reports section for ordering information). This very thorough report makes it perfectly clear that KIRIS was grossly inflated, and that the scores didn’t track other tests at all.

So much for wacky correlation coefficients. Now, you know better.

LESSONS FOR OTHERS

There are a tremendous number of lessons in the Kentucky KIRIS experience.

It is clear that Kentucky's first experiment with radical OBE testing is a failure. KIRIS didn't match other test results from ACT, SAT, Armed Forces Testing, and even the NAEP. While Kentucky academic indicators stayed flat, at best, remediation rates positively skyrocketed. Dropouts and retentions also jumped above pre-reform levels. In addition, there is new evidence of a decline in college interest in public high school students and an exodus of college aspirants to other school systems.

The more radical assessment ideas like math portfolios and performance events failed, but this information continues to be misrepresented outside the Bluegrass State. These programs have been completely rejected in Kentucky, but they are still being touted elsewhere.

Cheating on high stakes tests seems inevitable. It is important that strong safeguards be established to identify cheating on these tests, and prosecution must be diligent. The credibility of the entire test is destroyed if the Department of Education just winks its eye at cheaters.

OBE reforms are hugely expensive. Even with an enormous increase in funding, Kentucky's program is still mostly promises; but, there's no real proof of significant progress.

Table 1 — A sample of the correlation coefficient

Student	Test 1	Test 2a	Test 2b	Test 2c	Test 2d
1	1	1	1	1	1
2	2	2	1	1	1
3	3	3	2	1	1
4	4	4	3	2	1
5	5	5	4	3	2
6	6	6	5	4	3
7	7	7	6	5	4
8	8	8	7	6	5
9	9	9	8	7	6
10	10	10	9	8	7
Average	5.5	5.5	4.6	3.8	3.1
Correlation to Test 1		1	0.995591	0.98193	0.956541

WHAT QUALITY REPORTS SAY ABOUT KIRIS

A Study of Core Course-taking Patterns for Kentucky ACT-tested Graduates of 1991-1993 and An Investigation of the Relationship between Kentucky's Performance-based Assessment Results and ACT-tested Kentucky Graduates of 1992.

1994

Prepared by American College Testing for the Kentucky Department of Education, Frankfort, KY

For Copies: Phone (800) KDE-KERA

REGARDING CORE ACADEMIC COURSES: "...Kentucky has the **lowest** percent of ACT-tested graduates completing the colleges and ACT recommended core sequence of any state in the nation." (Pg 5)

"...it is apparent that the performance categories of **Novice** and **Apprentice** are not adequately distinguishing pupil performance relative to their above average performance on the respective ACT tests. A similar concern is warranted for student classified as **Proficient** or **Distinguished** who have ACT scores in the bottom half of the Kentucky quartile distribution." (Pg 6)

"In essence the Kentucky performance tests are as susceptible to the claims of apparent bias as are traditional testing methods." (Pg 7)

"ACT strongly advises that no judgments regarding individual student decisions can or should be made, at this time, on the basis of the present Kentucky performance test results." (Pg 8)

An Independent Evaluation of the Kentucky Instructional Results Information System (KIRIS)

1995

Conducted by the Evaluation Center, Western Michigan University for the Kentucky Institute for Education Research

For Copies: Phone (502) 227-8976

"Neither education and testing agencies nor the measurement profession has solved the many technical and operational problems with large-scale performance-based assessments." (Summary Pg 2)

"KDE and ASME postponed much of the needed research and development of assessment questions and implemented the legislatively mandated performance-based system at a very fast pace." (Summary Pg 2)

"...we think it would be a further mistake if KDE were to eliminate the multiple-choice questions altogether." (Summary Pg 8)

"...it is clear that taken by themselves two of the three components of the KIRIS accountability index are not sufficiently reliable to be used in a high stakes assessment. These two components are the writing portfolio and the performance events." (Summary Pg 9)

Perceived Effects of the Kentucky Instructional Results Information System (KIRIS)

1996

Rand Santa Monica, CA

For Copies, Phone (310) 451-7002

"Most teachers reported that KIRIS has put teachers under 'undue' pressure. Most teachers reported that teacher morale in their schools is low and has been harmed by KIRIS, and about half reported that KIRIS has reduced their own job satisfaction. A sizable minority reported that KIRIS has also decreased the morale of their students." (Pg 51)

"...very few (teachers and principals) considered the long-term goal of reaching an accountability goal of 100 to be realistic." (Pg 51)

"...about half the teachers strongly agreed that scoring standards for KIRIS are inconsistent over time, and a similar percentage strongly agreed that the curriculum content for assessments is not defined well enough for them to prepare students adequately." (Pg 51)

"About half the teachers reported that the emphasis on writing in KIRIS makes it hard to judge the mathematics achievement of some students." (Pg 52)

"...a majority of teachers also reported that at least one part of the assessment had a moderate or great deal of negative impact on instruction." (Pg 52)

"Most teachers (Nearly 90%) agreed that KIRIS has caused teachers to deemphasize or neglect untested subject areas." (Pg 53)

"Some teachers reported deemphasis on language mechanics, number facts, computation, and mathematical algorithms." (Pg 52)

Chester E. Finn, Et. Al.
Fordham Foundation
1015 18th St. NW
Suite 300
Washington, DC 20036
July 1998
1-888-TBF-7474

History — “F”
Geography — “F”
Math — “D”
Science — “F”
English — Not even rated due to very low score in American Federation of Teachers report
Making Standards Matter

Overall grade — “F”, ranked 10 from bottom of 50 states plus Washington, DC.

Brian M. Stecher et. al.
1998
CRESST
UCLA
Los Angeles, CA 90095
310-206-1532

“We found no consistent differences between teaching practices in high gain (on KIRIS) and low-gain schools based on second biennium KIRIS gains. (Pg 81)

“...the only difference associated with KIRIS gains is self-reported preparation; a greater percentage of teachers in high-gain than low-gain schools feel well-prepared to teach the more advanced mathematics topics, including algebraic ideas and statistics and probability.” (Pg 82)

“More eighth grade mathematics teachers in high-gain schools than low-gain schools did the following...had students practice computation on a daily basis.” (Pg 82)

“There was one difference in fifth grade in the other direction; a higher percentage of teachers in low-gain schools than high-gain schools increased coverage of probability and statistics in the past three years.” (Pg 82)

“Teachers in low-gain schools were more likely than teachers in high-gain schools to report that the Kentucky Writing Program had a great deal of influence on what they taught in writing.” (Pg 84)

“Policy decisions were made before there was any evidence that there were adequate technical procedures to support the policies.” (Pg 7)

“To date, policy decisions seem to have been made with the assumption that some psychometric procedure existed or could be devised to support any policy decision.” (Pg 7)

Prof. James S. Catterall, et. al.
Kentucky Legislative
Research Commission
1998
Capitol Annex
Frankfort KY 40601
502-564-8167

“Given the long chain of adjustments and equating links that holds the KIRIS program together, it would seem timely and prudent to begin the process of reestablishing a baseline year and re-setting standards as soon as possible.” (Pg 7)

“The 1997 KIRIS scores for all grades within and thus for elementary and middle schools almost certainly include an unknown inflation factor due to grade shifting.” (Pg 9)

“The 61 percent reading change could have been prompted by a reading achievement change of less than 5 percent!” (Pg 13)

“KIRIS thus appears to have a tendency to promote unrealistic impressions of achievement gains by its central and almost exclusive focus on its artificial school index score.” (Pg 13)

“KIRIS is thus set up for accusations of inflated score changes under the very best of circumstances. Only changes in score interpretation and primary reporting will fix this problem.” (Pg 13)

“...Currently allocated rewards and sanctions appear to have an indefensible spurious quality.” (Pg 13)

“Staying the present course with KIRIS would perpetuate the dissemination of indefensible information about school and student performance.” (Pg 15)

“Auditing in general seems an area where KIRIS has been unresponsive to the OEA panel report....(Pg 20)

“Between 1992 and 1995, the KIRIS gains of high school students in reading and mathematics were not reflected at all in their ACT scores.” (Pg xv)

“...the validity of KIRIS gains during the first four years is not a trivial concern — these gains have been widely presented as evidence of the program’s success and have been used to allocate tens of millions of dollars in rewards to schools.” (Pg 3)

The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)

Daniel M. Koretz & Sheila L. Barron

1998

RAND

Rand Santa Monica, CA

For Copies, Phone (310) 451-7002

“— until the publication of *Core Content* — the frameworks provided by KDE offered teachers little concrete guidance about the specific content and skills students would need to do well on the KIRIS assessments....the level of detail is still far less than one would have with a curriculum-based examination system....” (Pg 9)

“...Kentucky’s change on NAEP is not reliably different from that of the nation as a whole or of most other states participating in the Trial State Assessment.” (Pg 47)

“The comparison between NAEP and KIRIS in reading is unambiguous: while fourth grade KIRIS scores increased by a remarkable three-fourths of a standard deviation during the first two years of the program, NAEP scores remained unchanged.” (Pg 111)

Kentucky’s NAEP mathematics scores did increase over the first four years of the program, but far less than the scores on KIRIS. The KIRIS gains were roughly three and a half times as large as the NAEP gains in the fourth grade and four times as large in the eighth grade.” (Pg 111)

“...trend disparities this large suggest appreciable inflation of gains on KIRIS.” (Pg xv)

“Finally, it is important to place the results of this study in context....KIRIS has embodied themes common in educational reform programs nationwide, and the experience with KIRIS holds lessons for other programs, as well.” (Pg xviii)

An Evaluation of the Kentucky Education Reform Assessment System (KIRIS) and Academic Performance at the University of Kentucky

1995

Planning and Assessment Office, University of Kentucky

For Copies: Phone (606) 257-1633

“The major purpose of the study was to determine whether the KERA assessment program scores are sufficiently strong predictors of college level work that they could be used as an admissions criterion for the University System.” (Pg 1)

“The results of the regression analysis performed by ACT and confirmed by the Lexington Campus Office of Planning and Assessment indicated that 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.” (Pg 4)

“The relationship between the KIRIS assessment scores, whether we are talking about the four performance levels, the raw scores on subject area tests, or the combined percentile rank calculated by the Kentucky Department of Education, do not exhibit strong correlations with grades earned in introductory courses at the University.” (Pg 5)

Alternative Assessment Programs: What Are The True Costs?

(Cost Estimates From Table 13, Pg 47)

		Low	High
Lawrence O. Picus, et. al.	Total Costs Kentucky Assessment Program (1995-96)	\$120,463,573	\$254,413,039
April 1997	Total Cost per Student Tested	\$848.59	\$1,791.96
CRESST TR 441	Total Cost per Test Administered	\$141.41	\$ 298.66

CRESST UCLA Los Angeles, CA 90095-1522

For Copies: Phone (310) 206-1532

“Koretz, Madaus, Haertel, and Beaton (1992) estimate that Advanced Placement exams, which are similar to some types of performance assessment, cost \$65 per subject test, while commercial standardized tests cost from \$2 to \$5 per subject test.” (Pg 64)

Review of the
Measurement Quality of
the Kentucky Instructional
Results Information
System 1991-1994

1995

Prepared for the Office of
Educational
Accountability, Kentucky
General Assembly,
Frankfort, KY

For Copies: Phone (800)
242-0520

"After reviewing large numbers of curriculum and technical documents and assessments, conducting a number of relevant analyses, and carefully considering our findings, the Panel is in unanimous agreement that KIRIS is seriously flawed and needs to be substantially revised. The Panel is not suggesting that the education reform movement in Kentucky is a failure or that educators are not working hard to implement the goals of KERA." (Pg 1)

"As KIRIS is currently designed and implemented, it is flawed to the extend that, in the Panel's judgment, the present form of KIRIS cannot support its accountability and assessment goals and objectives. Also, as a result of the serious measurement flaws, (1) the public is being misinformed about the extend to which student achievement has improved state-wide; and (2) the public is being misled by being given information about the accomplishments of individual students that may be inaccurate." (Pg 5)

"Testing students for eight or more hours is certainly a long enough time to permit more extensive student reporting than was done in 1991-92 and 1992-93." (Pg 7-12)

"...the scoring of portfolios remains too flawed for use in a high-stakes system." (Pg 6)

"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 as represented in KIRIS." (Pg 10)

"...considerably more progress is needed to establish KIRIS as a technically sound accountability and assessment system." (Pg 10)

"...the average scores of fourth-grade learning-disabled students with certain combinations of accommodations were well above those students without disabilities." (Pg vi)

Despite KDE's seemingly restrictive guidelines on the use of accommodations, the large majority of students with disabilities tested with the regular KIRIS assessment were given accommodations, and many were given more than one. This pattern was most striking in Grade 4, in which 81% of students with disabilities were given at least one accommodation, and 66% were given more than one." (Pg 12)

*The Assessment of Students
with Disabilities in
Kentucky*

Daniel Koretz

July 1997

CRESST TR 431

CRESST
UCLA
Los Angeles, CA 90095-
1522

For Copies: Phone (310)
206-1532

"Thus in science, the mean student with a learning disability scored at about the 58th percentile on the distribution of nondisabled students." (Pg 19)

"Learning-disabled fourth graders receiving dictation in combination with oral presentation scored above the average for nondisabled students in every subject but mathematics, and they scored almost half a standard deviation above the average in science. Learning-disabled fourth graders provided with paraphrasing in addition to these two accommodations scored even higher...." (Pg 25)

"In other words, the average scores of learning-disabled students receiving these three accommodations together ranged roughly from the 57th percentile on the distribution of nondisabled students in mathematics to the 71st percentile in science." (Pg 25)

"...the sheer pervasiveness of accommodations is reason for concern. Over 80% of students with disabilities in the 4th grade and two thirds or more in the 8th and 11th grades were assessed with accommodations, and most of these were provided with two or more accommodations. This frequency appears inconsistent with the detailed and seemingly restrictive guidelines for the use of accommodations issued by the Kentucky Department of Education." (Pg 60)

"...these findings suggest that particularly in the fourth grade, accommodations may be used in part to offset poor reading and writing skills rather than disabilities, *Per Se*." (Pg 61)

"The striking associations between accommodations and test scores also raise doubts about the quality of of assessment results for some students with disabilities." (Pg 61)





