

Innovative Assessment in Kentucky's KIRIS System: Political Considerations

Brian Gong
Center for Assessment

Presentation at the Workshop on "Best Practices in State Assessment"
Sponsored by the National Academy of Science, Board on Testing and Assessment
Session on "What is the Status of Innovative Assessments, Political Considerations"
December 11, 2009 Washington, DC

Presentation Overview

- n Why and how were innovative assessment & accountability systems adopted in Kentucky?
- n What was innovative?
- n Implementation challenges and progress
- n Evaluation
- n What happened eventually and why?
- n Lessons learned?
- n Disclaimer: this is mostly a view informed by my personal experience 1994-98; necessarily just highlights



Motivation for Assessment and Accountability System

- n What was the motivation for the assessment? Why was it considered? Who wanted it? Who wasn't in favor of it?
- n What was involved in getting the assessment adopted? What, if any, obstacles were encountered? How were they overcome?



Context for Adoption - 1

- n 1989 Kentucky Supreme Court decision declared entire Kentucky educational system unconstitutional – state culture was this reflected financial disparities but also broad resolution to “get Kentucky off the bottom, from next to Mississippi”
- n 1990 Kentucky General Assembly passed KERA (Kentucky Education Reform Act) – broad bi-partisan vote and support (e.g., Senate vote 30-8; Prichard Committee)



KERA legislative context

- n KERA one of first state systemic education reform bills; whole atmosphere one of innovation to improve student learning and school functioning. Some examples:
 - n Historic tax increase to fund reform (over \$1B)
 - n Create state school board and state Commissioner; reconstituted SDE
 - n Site-based governance (school councils)
 - n Anti-nepotism law for schools, districts, school boards
 - n Integrated school and community health service centers
 - n Ungraded Primary (K-3)
 - n 10 paid professional development days/year for teachers
 - n State standards, assessment, & accountability
- n Broadly supported with some compromise (e.g., KY school boards association didn't like school councils)



What was innovative?

- n Goal: All schools and students Proficient, on average, within 10 years – common standards-based target for all students
- n Use: School accountability targets set on test performance (plus other indicators); consequences of rewards, support, and sanctions
- n Assessment: Use standards-based assessments rather than NRTs – general enthusiasm for “authentic assessments” that would involve HOTS, performances, be contextualized like “good classroom instruction”



Innovative assessment types

- n **Assessment types: primarily performance-based (not MC)**
 - n **CR: Very short (phrase or sentence), short (half-page) and long (page to unlimited; assumed content knowledge + HOTS); scored to public (general) rubric**
 - n **Writing portfolio: collection of “best work” from classroom curriculum; 6 pieces representing specified sample of genre; locally scored**
 - n **Performance events: on-demand, one-hour, involved “manipulatives” or investigation by group; individual written responses scored by rubrics**



Innovative assessment design

- n Designed for instructional improvement
 - n Performance-based to model desired instruction
 - n Seven content areas included: math, reading, writing, science, social studies, arts & humanities, vocational studies & practical living
- n Matrix-sampled to inform school accountability
 - n In R, M, Sci., SS students took common + matrix
 - n In A&H, PLVS students took matrix only
- n Forms used matrix CR to equate
- n Testing in grades 4, 8, and 11; then split content to 4/5, 7/8, and 10/11
- n Alternate portfolio for students with moderate-to-severe cognitive disabilities



Innovative accountability design

- n Require all students be included in accountability, including mobile and SWD
- n Required that all schools be included in school accountability, including schools without any testing and alternative schools.
- n Multiple indicators weighted into overall performance index (tests, non-cognitive)
- n Each school judged on progress towards closing the gap between baseline and goal
- n \$5 million distributed each accountability cycle to successful schools' faculty to vote on how to distribute; amount of money proportional to distance above target (for reliability)
- n \$25 million used for support of low-performing schools, including "Distinguished Educators" who had governance power to recommend firing teachers



Implementation Challenges & Progress

- n What was involved in developing the assessment? What, if any, obstacles were encountered? How were they overcome?
- n What issues were encountered with implementation of the assessment? What, if any, obstacles were encountered? How were they overcome?



General Implementation

- n KDE awarded contract to Advanced Systems in 1991; reawarded in 1996; Advanced Systems contract terminated in 1998; KIRIS replaced by CATS in 1998 by legislature; CATS dissolved by legislature in 2008 to be replaced by new system by 2010
- n Rapid initial development of materials, policies, and training 1990-98; again in 1998 for CATS; again in 2007
 - n For example, 1992: what could schools post on classroom walls? Policies developed in conjunction with professional association of district assessment coordinators. What were KIRIS skills; “power verbs” and instruction developed by School Support group
- n Evolution of policies by state board and KDE (lots of input by legislature, KEA, others)
 - n For example, move math to grade 5 due to writing portfolio in grade 4; what to do about schools that change grade-level configurations during accountability?



Technical challenges of innovative KY assessments

n Construct definition

- n 7 essential areas in court decision, 6 school outcomes; 57 “valued outcomes,” Academic Expectations, “Core Content for Assessment” – in the end, people relied on the assessment to define valued content, degree of expected expertise, and evidence – but assessment alone couldn’t provide sufficient information to inform teaching and action

n Creating tasks that validly elicited construct(s)

n Creating scoring that captured performance on construct and had enough information to be competitive in terms of efficiency and reliability; generalization across task domains

n Standard setting with complex, multiple evidence

n Equating to make scores from test forms comparable from year to year

- n Memorable prompts/tasks, not repeated year-to-year (e.g., writing, performance events, extended CR) – tried equating through out-of-state administrations; judgmental equating; equating through MC; no equating



Technical challenges – 2

- n Scorer accuracy and reliability
 - n CR year-to-year on standard (rigor) – instituted year-to-year drift monitoring and correction
 - n Portfolio – instituted writing portfolio audit (over-ride local scores); scoring analysis (feedback) and training and qualification – high scorer accuracy and reliability
- n Score reliability
- n Portfolio validity: whose work?
 - n instituted clearer guidelines and training about feedback and support about writing instruction



Technical challenges – 3

- n Construct definition in cross-discipline/ integrated tasks (e.g., math/social studies or reading/arts & humanities) – developed task development guidelines; eventually dropped in favor of single subject assessment
- n Construct definition and scoring rubric for alternate portfolio – wide range of students; locally chosen tasks
- n Construct definition and scoring rubric for writing and mathematics portfolios – holistic; holistic with comments; analytical



Operational challenges

- n Develop and field test quality CR
 - n E.g., if form had 12 CR, and released 8 common, and yield was $\frac{2}{3}$, then would need to field-test 12 CR yearly; 12 field-test forms with one FT item each (15-30 minutes for FT of CR only)
- n Scoring and turn-around time
 - n 8.1 million responses to score/administration = 4 main subjects x 12 CR/student + 2 smaller subjects x 3 CR/student = 54 CR/student x 50,000 students/grade x 3 grades
 - n Writing portfolios: 20-60 minutes each x 2 (double-scoring at local level) x approx. 20 students/class (150,000 total in grades 4, 8, 12)
- n Accountability reports returned in mid-fall



Operational challenges – 2

- n Sufficient staff with assessment expertise
 - n Built up during 1991-96 to around 5
 - n Decreased starting 1996 (Kingston left; Reidy asked to resign; 3 psychometricians, Gong left)
 - n One less engaged psychometric staff person by 1999
 - n No psychometric staff by 2004
 - n No state assessment director for extended periods 2004-2008; hired external consultant to help part-time



Political challenges

- n School accountability quickly unpopular with many
 - n Seen as too aspirational
 - n Do not believe “ All students can learn”
 - n Wanted student stakes for motivation in high school
 - n Rewards divisive (local votes on how to distribute)
 - n Too complicated
 - n Weak indicators of school quality (e.g., “success after graduation”)



Political challenges – 2

- n Opposed philosophically by many powerful state groups and individuals
 - n “Valued outcomes” seen as intruding on parental prerogatives and personal privacy
 - n “Outcomes based assessment” portrayed as government conspiracy to produce compliant workers
 - n Math and literacy wars: basic skills accuracy and fluency vs. greater emphasis on problem solving or writing organization & development (scoring weights)
 - n Unresponsive big government; arrogant and heavy-handed governance



Political challenges – 3

- n Program caught in political election battles
 - n Senate went 1990-98 from bi-partisan support of KERA under solid Democrat control to Republican control 18-17-1 independent who sided w/R – education & testing issues became large election issue
 - n Governor races made KIRIS a lightning rod
 - n Some school motivation for students to perform well went down when 1996 gubernatorial candidate in tight race announced that he would abolish KIRIS if elected
 - n KY Republican state legislature advocated for Voluntary National Test to replace KIRIS (when most Republicans at federal level were decrying VNT)



Political challenges – 4

- n Legislative leaders and state Board changed over time from “old guard” when KERA was passed
- n Commissioners changed (Boysen, Cody, Wilhoit selected and charged differently in reform life cycle)
- n Adversarial relationship between KDE and legislative oversight group (OEA) through 2000
 - n TAC established by legislature through OEA; no KDE TAC (OEA TAC’s contracts not renewed 2008; new TAC formed 2009)



Political challenges – 5

- n Politicized and divided legislature turned to their own advocates and “experts”
 - n Very influential individuals persistently and persuasively criticized KDE, programs, and policies (e.g., Richard Innes, Cindy Rausch, Eagle Forum)
 - n Mixed quality scholarship; much innuendo, persistent
- n Prolonged battles wore down KERA/KIRIS “defenders”
 - n “You have to win every battle; I have to only win one.” -- Republican state senator to KDE Assoc. Commissioner Neal Kingston
- n KDE somewhat defensive



Advocacy's interpretation

KERAgate Baffles Kentucky Parents and Taxpayers

Louisville, KY --The Kentucky Education Reform Act (KERA), which was passed by the legislature in early 1990, has grown more controversial with each passing year. Parents are so unhappy with KERA that they are pulling their children out of public schools in record numbers. Kentucky is now the number-one state in declining public school enrollment...

In June came the results of a study done by a national, hand-picked panel, a "dream team" of experts, paid for with the taxpayers' money. Their report confirmed earlier reports that the test was not valid or reliable, calling it "seriously flawed." It also said that the Department of Education, the State School Board, and our testing company, Advanced Systems of New Hampshire, had been overstating, exaggerating, misleading and misinforming the public regarding student learning progress.

Though the panel members were flown to Kentucky to testify before the Education Committee's Subcommittee on Accountability, they were allowed a mere one hour of a two-hour meeting to both report and answer questions. The explanation was that the Department of Education was also present and deserved equal time. In the brief time for questions, the panel revealed that the testing data gathered over the last four years -- costing Kentuckians \$100 million -- would be of little or no use.

The experts also acknowledged that there is now no way to determine if Kentucky students are doing better or worse than before KERA, nor is there any way to compare their achievement with that of students in other states. ...

Prior to the meeting of the full Education Committee, Eagle Forum held a news conference calling for an investigation. Other education groups participated in the conference, along with a senator and representative who are members of the Education Committee. In the subsequent committee meeting, the motion for an investigation failed. Instead, the committee voted to extend the contract with Advanced Systems, the testing company that had already spent \$100 million, for another year at the cost of another \$7.5 million of the people's money.

When the bid request went out to testing companies for a new contract for 1997, an open records request revealed memos from members of the panel of experts saying that it looked like it would be "business as usual" in the Kentucky tests. Their recommendations, which the people had paid to receive, were ignored.

On January 17, 1996, a newspaper headline read "Only 1 Bidder Pursues State's School-test Contract." The lone bidder was none other than Advanced Systems of New Hampshire, a company that was small and relatively unknown in 1991 when it was awarded the original contract over much larger and more experienced testing companies. ..

Kentuckians who are paying attention are asking for a legislative investigation of what has become widely known as "KERAgate." -- *Reported by Donna Shedd* <http://www.eagleforum.org/educate/1996/feb96/ker.html>



Evaluation

- n How well did the assessment work (or if currently underway, how well has it worked)?
- n Kentucky one of the most studied “programs” in recent history, but no definitive agreement on program history



Evaluations

- n Two key summative program evaluations of assessment and accountability systems
 - n OEA Panel, 1996: “fundamentally flawed”
 - n Catterall Group, 1998: “good progress but have not implemented fully”
- n Research
 - n KDE
 - n HumRRO – program implementation, consequential validity
 - n KY university-sponsored groups (Pankratz, Petrosko, Kifer)
 - n RAND
 - n CPRE
 - n Many others



Lessons Learned and Recommendations

- n Be very clear about the constructs, purposes, and uses. No assessment can compensate for poor quality learning targets/content standards/expressions of expected expertise.
- n Design assessment and key uses (e.g., accountability) together so they can be coherent and efficient.



Lessons Learned and Recommendations

- n Those proposing technical evaluations, and those likely to be subject to evaluations, should consider the intended and unintended consequences, particularly in politically charged contexts, as illustrated by Kentucky's experience. How will program evaluation be done that helps improve innovative assessments and programs using innovative assessments?



Lessons Learned and Recommendations

- n Those proposing innovative assessments now for large-scale use (e.g., testimony for USED “Common assessment RFP , 11/09) should show how they would meet technical and operational challenges better than did KDE in the 1990’s.
- n The technical criteria based on current psychometric models are inconsistent with views of content and cognition and should be applied modestly to innovative assessments; the field should invest in developing better models and criteria (cf. Mislevy, 1998, “Foundations of a New Test Theory”; Shepard, 1993, “Evaluating Test Validity”)



For more information:

Center for Assessment

www.nciea.org



Brian Gong

bgong@nciea.org

