

KERA UPDATE

April 1997

#15

Are KIRIS Results Reliable for Individual Student Use?

One of the current questions about Kentucky education is whether recent claims that KIRIS is now accurate for individual student evaluation are actually true. Collected below are quotes from some of the various reports from leading education research groups that have examined KIRIS. It remains to be seen whether the KY Department of Education and Advanced Systems in Measurement and Evaluation have been able to make changes that overcome the serious concerns that have been raised in the past over this issue. Editorial comments follow each quote set in full page width blocks.

[Review of the Measurement Quality of the Kentucky Instructional Results Information System 1991-1994](#)

Prepared for the OEA

"...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)

"In most cases, the coefficient alpha reliabilities are lower than those normally obtained for standardized tests used to report individual student scores in major domains such as reading, mathematics, science, or social studies." (Pg 3-24,25)

What KDE has changed to improve KIRIS reliability for school accountability purposes since the OEA Panel report was released in June of 1995? How much evidence has been gathered? The OEA Panel didn't focus on individual reliability, because that was not the proclaimed purpose of KIRIS in 1995. Still the last quote above indicates that the Panel would likely not approve of certifying KIRIS for individual uses without a lot more research.

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

Planning and Assessment Office,
University of Kentucky

"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 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)

This small study by ACT and UK is possibly one of the most relevant to the discussion of whether KIRIS can be used for individual accountability. UK reported, for the high school Class of 1993, that KIRIS did not bear much relationship to college grades. UK declined to continue this study (which involved time and expense) until KIRIS was determined to be more stable. At present, a follow-on study is in progress by ACT and most of the colleges in the state. Pending the review of that study by independent experts, claims that KIRIS can be used for individuals should be treated as premature.

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

Conducted by the Evaluation Center, Western Michigan University

“Overall, there is a considerable amount of measurement error associated with individual student level scores. For this reason, the 1992-1993 Technical Report (Ch 9, P. 10) states that ‘...current reliabilities are not sufficiently high to make student-level decisions without additional information.’” (Pg 37)

This separate group of experts obviously agrees with the OEA Panel.

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

1996

Rand

“...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)

Clearly, Kentucky teachers do not have current confidence in the testing program. What is new that makes KIRIS better for individual students while our teachers continue to have such strong opposite opinions?

Generalizability of New Standards Project 1993 Pilot Study Tasks in Mathematics

Robert L. Linn and Elizabeth Burton

CRESST

“...Sampling variability due to tasks was found to be substantially larger than due to raters.” (Pg 1)

“...Measurement error can be reduced more rapidly and effectively by increasing the number of tasks rather than by increasing the number of raters.” (Pg 7)

“...A substantial number of tasks is needed to achieve a generalizability coefficient as high as 0.70. Even with 15 tasks each rated by two raters, the generalizability coefficient is less than 0.80, a value that some would consider a minimum for purposes of making important decisions about individual pupils.” (Pg 10)

“...The misclassification error rate is quite high with a reliability of 0.80.” (Pg 27)

“...A sizable number of tasks is needed to make dependable decisions about individual students.” (Pg 27)

Kentucky is a New Standards member state, and KIRIS is modeled in the New Standards mold. We have never used close to the minimum number of questions required to even begin to approach acceptable levels of accuracy for individuals on KIRIS. No KIRIS element for 1996-97 will have close to the minimum needed questions, either.

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.

Prepared by American College Testing

“...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)

“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)

This was the first report on KIRIS from a major group of testing experts. Clearly, ACT was unimpressed with KIRIS in 1992. It is significant that KDE never released follow-on studies that ACT did for the Classes of 1993 and 1994. It is even more unfortunate that the ACT studies were totally disbanded after that. Before we even think about accepting KIRIS as valid for individual student use, as with the UK study, we need to get the ACT studies up and running again. Because extensive research, from many sources, has shown KIRIS to date is not individually reliable, it is going to take a very extensive effort over an extended period to develop convincing evidence that KIRIS has changed. With performance events crashing, math portfolios on the wane, and the latest NAEP results once again showing a conflicting picture, it seems like a real stretch to believe that KIRIS has achieved this miracle.