DOCUMENT RESUME

ED 397 135 TM 025 310

AUTHOR Kannapel, Patricia J.; And Others
TITLE "I Don't Give a Hoot If Somebody Is Going To Pay me $3600:" Local School District Reactions to Kentucky's High Stakes Accountability Program.
INSTITUTION Appalachia Educational Lab., Charleston, W. Va.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 12 Apr 96
CONTRACT RJ96006001
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; Educational Assessment; Educational Change; Elementary Secondary Education; Interviews; *Performance Based Assessment; *Portfolio Assessment; Portfolios (Background Materials); Program Implementation; Qualitative Research; *Rural Schools; State Programs; *Teacher Attitudes; Teacher Education; Testing Programs
IDENTIFIERS *Kentucky Education Reform Act 1990; *Measurement Driven Instruction; Reform Efforts

ABSTRACT

This paper places Kentucky's accountability program in the national context, and then describes how educators in four rural Kentucky school districts reacted to the program in the initial years of its implementation. The information is drawn from a qualitative study of the implementation of the Kentucky Education Reform Act (KERA) in these four districts, which began in 1990 and was ongoing through the 1994-95 school year. The KERA assessment and accountability program, the Kentucky Instructional Results Information System, uses a performance-based test with open-response questions, performance events, and portfolios. Observation, interviews, and document review were used to study the reactions of educators in the four target districts. Very few of the teachers interviewed in these districts accepted the KERA philosophy that all children can achieve at high levels. However, the study produced evidence that the accountability measures were beginning to drive the reform effort to some extent in that teachers changed their instruction to focus on writing and schools were adding new courses. There was also evidence that at least some teachers were making a stronger effort to reach all students. Most commonly, however, teachers taught to the test because they were forced into it, and they often resented this. The KERA effort might be more successful if the focus was placed more on building teacher capabilities than on the rewards and punishments as they are now conceived. (Contains 24 references.) (SLD)

"I DON'T GIVE A HOOT IF SOMEBODY IS GOING TO PAY ME $3600:"

Local School District Reactions to Kentucky's High Stakes Accountability Program

Paper presented at the Annual Meeting of the

American Educational Research Association

New York, NY

April 12, 1996

Patricia J. Kannapel
Pamelia Coe
Lola Aagaard
Beverly D. Moore

Appalachia Educational Laboratory
P. O. Box 1348
Charleston, WV

The Appalachia Educational Laboratory (AEL), Inc., works with educators in ongoing R & D-based efforts to improve education and educational opportunity. AEL serves as the Regional Educational Laboratory for Kentucky, Tennessee, Virginia, and West Virginia and operates the Eisenhower Math/Science Consortium for these same four states. It also operates the ERIC Clearinghouse on Rural Education and Small Schools.

AEL works to improve professional quality, curriculum and instruction, community support, and opportunity for access to quality education by all children.

Information about AEL projects programs, and services is available by writing or calling AEL, P. O. Box 1348, Charleston, West Virginia 25325-1348, 304/347-0400, 800/624-9120 (toll-free), 304/347-0487 (FAX).

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement, U. S. Department of Education, under contract number RJ96006001. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U. S. government.

AEL is an Affirmative Action/Equal Opportunity Employer.

INTRODUCTION

> All of this rewards business is reinforcing one of the greatest things that's wrong with our society: greed.... I try to teach well because it's the right thing to do. I've got a young kid's mind in my hands.... I don't give a hoot if somebody is going to pay me $3600.... That money isn't going to make the school better.... You can't deal with [schools] as a business and have rewards and sanctions and stuff like that.... They forget the personal side that's attached--that you want these students to learn and that you care about them.
>
> A Kentucky teacher

With these words, a Kentucky middle school teacher summed up the anger and insult that many Kentucky teachers felt in the early stages of Kentucky's new accountability program. Established in 1990 as the centerpiece of the Kentucky Education Reform Act (KERA), the accountability program requires schools to show a specified level of improvement on a new, performance-based assessment or face sanctions that could ultimately result in the dismissal of individual teachers and the voluntary transfer of students to more successful schools. Teachers in schools that improve beyond the goal set by the state are eligible to receive substantial financial rewards.

This paper places Kentucky's accountability program in the national context, and then describes how educators in four, rural Kentucky school districts reacted to the program in the initial years of its implementation.

The information is drawn from a qualitative study of KERA implementation in these four districts, which began in the fall of 1990 and was ongoing through the 1994-95 school year.

KENTUCKY'S ACCOUNTABILITY PROGRAM IN THE NATIONAL CONTEXT

Measurement-Driven Reform

Kentucky's new assessment and accountability program, known as the Kentucky Instructional Results Information System (KIRIS), is the centerpiece of the Kentucky Education Reform Act because the assessment component, when linked with high-stakes accountability measures, is meant to drive instruction (Kifer, 1994). The assessment is linked to KERA goals and academic expectations and is a performance-based test that consists of open-response questions, performance events, and portfolios.

Using assessment as a catalyst for reform became common practice in the United States in the 1970s and 1980s, first with competency testing and then with accountability systems with increased stakes (Linn, 1993). James Popham began to advocate for the use of "measurement-driven instruction" (MDI) in the mid-1980s. Popham defined MDI as occurring "when a high-stakes test of educational achievement, because of the important contingencies associated with the students' performance, influences the instructional program that prepares students for the test" (Popham, 1987, p. 680). Popham identified two types of high-stakes tests: (1) examinations that are associated with important consequences for students, such as tests that qualify students for promotion or graduation; and (2) examinations whose scores are seen as reflections of instructional quality, such as test scores published in the media. Kentucky's KIRIS program is an example of the second type of high-stakes test because educators, not students, experience the consequences of students' test performance.

Measurement-driven instruction is often viewed as problematic when norm-referenced tests are used in high-stakes settings. Teaching to these tests tends to lead to a narrowing of the curriculum to the content included on the test, which is often focused around isolated skills (Bracey, 1987; Haertel, 1994; McLaughlin, 1991; Shepard, 1991). When high stakes are attached to norm-referenced tests, teachers often teach in ways that violate their own beliefs and values about how and what students should be taught (Hatch & Freeman, 1988; Urdan & Paris, 1994).

In contrast to norm-referenced, high-stakes assessment, CIRIS is a product of the latest thinking with regard to measurement-driven reform. The new view is that if tests are changed to require higher-order skills, the negative effects of previous MDI efforts will be offset (Haertel, 1994; Linn, 1993; Noble & Smith, 1994; Shepard, 1991). When performance-based examinations are closely linked to a "thinking curriculum," teaching to the test is considered desirable (Linn, 1993; Noble & Smith, 1994; Shepard, 1991). Even with the introduction of performance-based assessments, many people continue to have grave reservations about using an assessment program to drive instructional reform because of possible onerous, unintended consequences of high-stakes testing, most of which would affect students who are not likely to do well on the assessment. Fears have been expressed that schools may retain such students in non-accountable grades (Haertel, 1994), encourage them to drop out of school (Darling-Hammond, 1991), place them in special education so they won't be tested (Darling-Hammond, 1991), or reject them entirely (Darling-Hammond, 1991; McLaughlin, 1991; Shepard, 1991). In addition, school administrators may shift the best teachers to the accountable grade levels (Haertel, 1994), or teachers may try to transfer to schools that have a better chance of doing well

on the test (Darling-Hammond, 1991).

Another problem with externally-mandated tests associated with high-stakes accountability for teachers is that such programs result in teachers feeling "de-skilled" and having a diminished sense of professionalism (Noble & Smith, 1994; Shepard, 1991; Torrance, 1993). In addition, many people believe that teachers must be involved in the development and implementation of the assessment and/or accountability program if they are to buy into it (Darling-Hammond, 1991; Firestone, 1991; McLaughlin, 1991; Shepard, 1991; Urdan & Paris, 1994). Others note that teachers need strong professional development and instructional support if they are to implement new, performance-based curricula and assessments (Almasi et al, 1994; Aschbacher, 1994; Noble & Smith, 1994; Popham, 1987; Shepard, 1991; Soodak & Martin-Kniep, 1994).

In spite of these reservations about measurement-driven reform, educational policymakers continue to search for ways to hold schools accountable for student achievement, and high-stakes assessment programs continue to be the instrument of choice. Because the use of performance-based tests in high-stakes accountability programs is relatively new and limited, there are few studies that document whether or not the positive outcomes envisioned by supporters are beginning to occur. Some early studies in the United States indicate that positive instructional outcomes have accrued from the use of performance-based testing. Koretz, Stecher, & Deibert (1992) found that the Vermont math portfolio assessment program led teachers to spend more time teaching problem-solving, patterns and relationships, and mathematical communication. Almasi, Afflerbach, Guthrie, & Schafer (1994) found that the high-stakes Maryland School Performance Assessment Program resulted in teachers altering

instructional tasks to reflect increased writing opportunities, an emphasis on personal response to reading, student choice in literacy activities, and greater student input and interaction. Studies done in Kentucky report that the strong emphasis on writing in the KIRIS assessment has led teachers to teach a great deal more writing in the classroom (Appalachia Educational Laboratory, 1994; Bridge, Compton-Hall, & Cantrell, 1996; Western Michigan University, 1995).

Two studies of high-stakes, performance-based assessments in Britain, where high-stakes testing has been the norm for a much longer period, produced relatively negative results. Torrance (1993) reported that in the pilot year (1989-90) of implementation of the British National Assessment Program, teachers' workload increased greatly, students were underworked during the study period while teachers helped individuals or small groups stay focused on the assessment tasks, special areas classes (P. E., music, etc.) were frequently cut from the curriculum to give more time for assessment activity, teachers treated the assessment as a separate activity rather than integrating it into their instruction, and teachers felt "deskilled and overly dependent on the packages that had been delivered" (p. 86). Freedman (1994), who studied a British high school exam with high stakes for students found that exam requirements "...inhibited the teachers abilities to build a coherent curriculum with their students and inhibited the amounts and kinds of writing the students did" (p. 4).

The Kentucky Accountability Program

Given the limited research evidence on the use of performance-based assessment and accountability programs, Kentucky is charting new waters in implementing the KIRIS program

and thus, provides an opportunity to study the effectiveness of the new approach to measurement-driven reform. The Kentucky General Assembly mandated development of a performance-based assessment program to ensure school accountability for student achievement of goals set forth in KERA. Students in grades 4/5, 8, and 11/12 are tested annually with an assessment instrument that consists of writing portfolios, a "transitional" test containing multiple-choice and open-response questions, and performance events. The KIRIS test was first administered during the 1991-92 school year, and those results, along with measurement of non-cognitive goals (such as reduction in dropout and retention rate and increase in attendance rate) were used to establish a baseline "accountability index" for all schools in the state. The baseline was used to set an incrementally increased "threshold" or goal score that each school was required to meet by the 1993-94 school year in order to obtain rewards or avoid sanctions. This measurement is ongoing; i. e., a school accountability index is determined biennially and schools are expected to show improvement over their baseline scores. Scores from both years in each biennium, along with measurement of non-cognitive factors, are averaged to determine if schools reached their thresholds. Student performance is reported in terms of four performance standards: novice, apprentice, proficient, distinguished.

Schools that score at least one percent above their thresholds and move at least ten percent of students scoring "novice" to a higher performance level receive financial rewards, to be divided according to the wishes of the majority of educators at the school. Schools not achieving their thresholds receive varying levels of assistance and/or sanctions, depending upon how close they come to achieving their thresholds. Levels of sanctions are:

Level 1: Schools that fail to meet their thresholds but maintain the previous proportion of successful students must develop a school improvement plan and are eligible to receive funds from the school improvement fund;

Level 2: Schools in which the proportion of successful students declines by less than five percent: same as Level 1, plus a Distinguished Educator is assigned to the school to assist in implementing the school improvement plan;

Level 3: Schools where the proportion of successful students declines by five percent or more: same as Level 2, plus the school is declared to be a "school in crisis," all certified staff are placed on probation, parents are permitted to transfer students to the nearest successful school, and, after six months, the Distinguished Educator determines which of the certified staff should be retained, dismissed or transferred.

The first round of rewards and sanctions was administered after test scores from the 1992-94 biennium were tabulated. The maximum reward amount was initially set at $3690 per full-time certified staff member in each school, but that amount was reduced to $2602 when more schools than anticipated qualified for rewards. About one-third (479) of Kentucky's 1247 schools earned some level of reward for the first biennium, while 55 schools had declining scores. Recognizing that schools had not had sufficient time to implement all aspects of KERA, the 1994 Kentucky General Assembly delayed imposition of Level 3 sanctions until the end of the 1994-96 biennium.

THE AEL STUDY

The Appalachia Educational Laboratory (AEL) undertook a qualitative study of KERA implementation in four rural school districts beginning with a baseline study in 1990-91 (Coe, Kannapel, & Lutz, 1991) and extending through the 1994-95 school year. The intent of the study was to inform policymakers, practitioners, and researchers about how state-mandated, large-

scale restructuring played out in local school districts. Small, rural districts were the focus of the study because KERA provided an opportunity to study systemic reform in rural schools. The study districts were selected to reflect a range of geographic, economic, and demographic conditions. One district was located in western Kentucky, one was in central Kentucky, and two were in eastern Kentucky. Three were county districts, and one was a small, independent district located within the boundaries of a larger, county district. The districts were assigned pseudonyms to protect their anonymity.

Methods

The AEL research team was an ongoing and regular presence in the four study districts throughout the study period. The chief research techniques were observation, interviews, and document review. The research focused on four key components of KERA, one of which was instruction and assessment changes in Grades 4 through 12. Data on this aspect of KERA were gathered from interviews with district superintendents and key central office staff, school principals, teachers, and parents. In addition, school-based decision-making (SBDM) council meetings and school board meetings were observed at least twice a year. Research staff also reviewed key documents such as local and state newspapers, minutes of all SBDM council and school board meetings, and rewards certification forms from schools receiving rewards. At the state level, researchers attended all meetings of the State Board of Elementary and Secondary Education from 1992-93 through the end or the study period, and reviewed key documents related to the KERA accountability measures.

Research Focus

This paper seeks to answer two key questions:

(1) How did the accountability measures affect school practices?

(2) How did local educators and parents view and respond to the KERA accountability measures?

The first question is concerned with the extent to which the accountability measures produced behavioral change, whether intended or not, among educators in schools and classrooms. The second question deals with attitudes toward and opinions about the accountability measures held by local respondents.

RESEARCH FINDINGS

Effects of Accountability on School Practices

In this section, we consider two aspects of school reform that KIRIS is designed to drive: (1) The KERA mandate that schools must expect a high level of achievement of all students, and (2) The extent of curricular and instructional change designed to help students achieve KERA goals. We also describe some unintended, negative consequences of the accountability program that we witnessed or heard about in the study schools

Expecting high levels of achievement of all students. The basic premise underlying KERA is that all children can achieve at high levels. This issue is addressed at least three times in KERA: (1) The introductory section states that the legislature's intent was to create a system of public education that will allow **all** students to acquire the specified capacities; (2) The section listing KERA goals states that schools must expect a high level of achievement of **all**

students; and (3) The section describing the accountability measures states that the legislature intends for schools to succeed with **all** students (Kentucky Department of Education, 1994).

The KIRIS assessment is designed so that all students must take the test, and so that educators will feel compelled to push all students to achieve at high levels. Schools receive a "novice" score for students who do not take the test. All special education students participate in KIRIS through adaptations (such as having the test read aloud to them) or through an "alternate portfolio." Schools are required to move 10 percent of students scoring "novice" to a higher performance category to receive rewards, and the accountability formula is designed so that schools are expected to move nearly all students to the proficient level by the end of 20 years of reform.

In spite of the strong incentives for schools to push ALL of their students to achieve, very few of the educators we interviewed in the four study districts accepted the philosophy that all children can achieve at high levels. In 1994-95, 85 percent of the 70 or so people who were asked about this philosophy said they did not subscribe to it. Most people felt that some children are simply not capable of achieving at high levels. A high school teacher remarked:

> I think that students should be expected to achieve at the highest level that they are capable of, but I do not believe that all students can achieve at the same level. Every student should be given the opportunity to do their best, to learn as much as they can, to excel, but they're not all going to be proficient writers or proficient mathematicians, much less distinguished. The reality is that we have to treat each other as equal human beings and respect each other in that way, but life is not all an equal playing field as far as ability. We all need to have equal treatment, equal respect, equal opportunity, but to expect all students to perform at the proficient level is not possible in my opinion.... We have to recognize that we have basic differences.

Others commented that home background or lack of student motivation would keep some

students from attaining high levels of achievement. A junior high teacher commented:

> No, I don't think all students can achieve at high levels. I think they can achieve at **their** high level. But if you're going to say "This is high level and everybody can reach it," I'm sorry, it just won't work. Not unless you're going to give them to us when they come out of the hospital and we're going to be here 24 hours a day, maybe. But you're fighting a whole other world out there, sometimes; a world that doesn't value education.

Even though it did not appear that most educators believed that all children could achieve at high levels, there were instances when it appeared that student performance on KIRIS was convincing some educators that students were capable of performing at higher levels than they had imagined. The comments of an elementary school principal illustrate how student performance on KIRIS led him to raise his expectations of students:

> We anticipate the movement from "novice" to "apprentice" [at our school] to be about 25 percent; the state asks for 10... We already feel we have five "distinguished" portfolios in the 8th grade... I looked at one piece that was 17 pages in length... This is amazing to me... the way we have allowed these minds not to expand... This is the one thing I think we mention every meeting we have: "Don't forget to expand your expectations of children." The more we see those results, the more we can expand our expectations.

An 8th grade teacher in another district described how one of her students had progressed through working on portfolios:

> I had a kid last year who couldn't write a sentence... not "The dog bit", nothing... I read a short story in his portfolio the other day and it was almost five pages long. In his letter to the reviewer he said, "I never thought I would be able to do this." I look at him and I think, "What if all I had stressed had been grammar skills and punctuation skills but he had never transferred that?" I can tell you he would have failed. He has failed before. He's 16 years old and in the 8th grade... Last year he was the slowest kid I had, and I thought, "I'll never get a portfolio out of him." We had no problem. He's the first one to hand his portfolio in. But I know if I give him anything where he has to have rote memorization, he won't pass it.

We also saw evidence that at least a few teachers who did not profess a belief that all

children were capable of learning at high levels were pushing all to do so because of the accountability measures. One teacher commented:

> I've always felt like I had enough experience with children to know which ones to push, which ones to encourage, which ones to say, "I can't accept this." Now I have to push everyone of them because it's not that they're accountable, I'm accountable. I have to.

One elementary school principal described how the combination of rewards/sanctions and student results led teachers at his school to raise their expectations of students:

> I think we are becoming more aware of all the students than what we have in the past. I think it has been fairly easy in the past to say, "Well, Johnny is just not interested so just let Johnny sit there and the rest of us go on." I think the teachers are more aware now: "Yeah, Johnny isn't interested, but let's see if we can do some things to pull Johnny into it to get him interested." [This shift in attitude occurred] partially because they know that schoolwide we're under the gun to do this.... rewards and sanctions hanging out there over the top of us. Another big part of it is that most teachers are really buying into the fact that, "Yeah, we do need to try to reach these kids, even if it's only a little bit. If we don't, who will?" When I came here, the 7th and 8th grades were absolutely horrible. All we could think about was, "Boy, in two years they'll be gone...." Well, those were the kids that we got these scores with on the 8th grade assessment. It was that group of 8th graders the first year and the second group the next year. Those were all the really bad ones. I pointed out to [the teachers], "You exceeded your threshold in the 8th grade by a little over a point and you did it with those types of students. That's got to tell you something very positive about yourself." To me, that was probably a much bigger motivator than rewards or sanctions was. They realize that they have had a degree of success, where I don't think they really thought before that they could.

In summary, it appeared to us that the majority of educators either did not believe that all children can achieve at high levels, or believed that it was not within their power to ensure that this occurred. However, student achievement on KIRIS made believers out of a very few people, or at least opened their minds to the possibility that they could expect more of students. Also, at least a few teachers were pushing all students to do well on KIRIS in direct response to the

accountability measures. It is too soon to tell if these responses will become a general trend as educators develop a better understanding of what they must do to earn rewards and avoid sanctions. At the time this study concluded, these sorts of reactions were the exception to the rule.

Extent of curricular and instructional changes. We observed a moderate degree of change in school curriculum and instructional practices, most of which appeared to be assessment-driven. The most conspicuous change was a much greater emphasis on writing and the writing process to help students develop writing and mathematics portfolios. Other assessment-driven changes were open-ended problem solving exercises to help students develop mathematics portfolios; practice at answering open-response questions (an important component of KIRIS testing); and incorporation of content that is tested on KIRIS but not previously taught in most rural school districts, such as the arts and humanities.

The increase in the amount of writing taught was in direct response to the strong writing component of KIRIS. KIRIS includes a writing portfolio, a math portfolio with a strong written component, and open-response questions that require students to write. The accountability index is designed so that writing portfolios count for 100 percent of the school's writing score, while open-response questions count for at least 60 percent of the score; 80 percent in some content areas. Thus, there is a strong incentive for teachers to focus on writing, and we saw that they were doing so. Even when teachers made few other changes, most assigned much more writing than previously.

It is unlikely that many teachers would have emphasized writing to the degree that they did had they not been forced into it, yet many teachers began to see the benefits and reported that

they would continue to emphasize writing even if KERA were not in effect. This was especially true of teachers who served as writing portfolio "cluster leaders"--teachers who received regional and state-level training and information about portfolios, which they passed along to their colleagues. This suggests that professional development combined with student results led to philosophical and behavioral change for teachers. An eighth grade writing portfolio cluster leader explained why she had become a portfolio enthusiast:

> For years we have been so concerned with [teaching] English in part. You did a unit and then you went to another one and you never brought it all together. Somehow the kids never understood, "Why am I doing this?" So to me, actually seeing that they can communicate and use these skills is great.

However, teachers reported that writing consumed so much classroom time that they did not have time to teach skills or cover the amount of content they felt they should. Teachers of content areas other than writing and math were especially resentful of having to spend so much time on writing. An elementary physical education teacher commented:

> My number one concern, since I teach P. E., is that we have 80 percent of the American public overweight, and now they're trying to get us to write and do all this other stuff in our P. E. time. We only have 30 minutes three times a week [with each class]. If we go to junior high, they're only [in P. E.] for six weeks and don't have any the rest of the year. And now they're taking what little time we have and trying to get us to do other things. I'm sure that's well and good, but if they're going to get overweight and fat and die at 30, it's not going to do them very much good to be educated.

Another assessment-driven change was the use of open-response questions on tests and in classroom assignments. Because of the heavy weighting given to open-response questions in the accountability formula, teachers in all schools were teaching students to answer these questions. Many schools had all teachers post a list of the steps required to answer an open-response question, and some schools held mini-courses in answering open-response questions. In some

respects, this was a superficial change because the skill of answering open-response questions was sometimes taught out of context of the subject matter being taught. However, most teachers we talked to supported the use of open-response questions in general and many said they had always included them on tests. Thus, the use of open-response questions in and of themselves was something teachers could easily accept, but the pressure of the accountability measures often compelled schools to teach open-response questions in an unnatural way.

A curriculum change that appeared to be driven almost entirely by the assessment program was expansion of the curriculum to include arts and humanities courses. This content is not traditionally covered in any depth in rural schools, but two high schools in our study added arts and humanities courses to the curriculum when art and humanities questions on KIRIS began to count in the accountability formula. While teachers may have supported the addition of arts and humanities courses, it appeared that the main impetus for offering the courses was to prepare students for KIRIS. There had been no indications prior to this time that the schools were considering adding such courses, for their own sake, to the curriculum.

It should be noted that the developers of KIRIS envisioned a system in which assessments were embedded into classroom instruction. In fact, KIRIS itself contains not only an accountability component but a continuous assessment component that consists of instructionally-embedded assessments (Kifer, 1994). In practice, however, the accountability testing has taken on much greater importance than the continuous assessment (Kifer, 1994). For the most part, it did not appear that many teachers were integrating instruction and assessment. In general, teachers tended to tack writing activities onto their existing curriculum or supplant other content in order to squeeze in writing. This suggests that coverage of traditional content

may have suffered. For instance, when teachers assigned portfolio pieces, they often saw such assignments as interruptions to the curriculum. One of the students we interviewed complained that she had difficulty learning pre-calculus concepts when instruction was interrupted so the class could work on their mathematics portfolios. Because portfolios and other parts of KIRIS (such as open-ended responses) were frequently perceived as "add-ons" rather than integral to classroom instruction, some teachers eventually cut back on innovations they had begun to implement, such as use of manipulatives and hands-on activities in order to spend time "teaching to the test."

Unintended consequences. As discussed in an earlier section, high-stakes accountability systems can produce adverse consequences along with the more positive ones. Some of the adverse consequences that might have occurred have been avoided through the KIRIS requirements that all students be tested, and that schools move a certain percentage of low scoring students to higher categories. Even so, we have seen and heard evidence of unintended side effects that resulted from the pressure imposed by the accountability measures.

Haertel (1994) predicted one unintended outcome of KIRIS that apparently did occur at one school: A primary teacher reported in 1993-94 that the principal forced teachers to retain a disproportionate number of students in the primary program to prevent them from being tested in the fourth grade the following year. The principal agreed that this was the case, but said that these students would have likely been retained in the early grades in the absence of an ungraded primary program.

While this was the most onerous consequence of the accountability measures we heard about, we heard other comments suggesting that pressure to do well on the assessment was

forcing teachers and principals to focus exclusively on that goal. For instance, a principal expressed concern about the amount of time eighth-grade students devoted to the band program and athletics:

> Like it or not, band's not going to do anything for us as far as us reaching our threshold. But we've got half of our eighth grade out today to a band camp when we need to be in our classrooms getting ready for our KERA testing... Band directors and coaches don't see anything other than their little program. I told our basketball coach the other day, "I'm sorry, but your 18 wins a year in junior high school is not going to help me one darn bit to keep my job."

The same principal expressed the view that the extended school program, which is designed to assist students who need extra time to meet KERA goals, should be offered to only those students who have the potential to move from the "novice" to the "apprentice" performance category. He felt that the extended program was wasted on students who, in his view, would never score higher than "novice."

An eighth-grade teacher in another district spoke of the temptation to focus her efforts only on students who were capable of advancing to the next performance category:

> I have kids who are novice students, and that's all they are ever going to be. But I worked probably harder with those novice students, and they worked as hard as any of the other levels to come up with their finished [portfolio] product... but yet, at the novice level, they only score zeroes. All they'll ever score is a zero. We don't get any credit at all for the improvement that they made--and they made a lot of improvement, but they're just never going to be at that next level. I just feel like, as a teacher, this kid is a novice, whatever I do with him isn't going to help that score. Just forget him, let me work with these. And I think that's going to happen to a lot of kids. There should be some way to measure the improvement that they have made.

Haertel (1994) identifies another possible adverse consequence of high-stakes assessment programs: that the test scores themselves become the goal of education. He states that "something is lost when teachers and students work for grades themselves instead of the

intellectual attainments those grades are meant to represent" (p. 70). While KIRIS is designed such that scoring well on the assessment should represent actual educational achievement, we heard evidence suggesting that the pressure of the accountability program was causing some teachers to focus more on assessment scores--and the resulting salvation or loss of their jobs--than on the welfare of their students.

At one school, parents reacted negatively to the SBDM council's plan to increase the number of credits required in core subject areas in order to improve KIRIS scores. Parents feared students would not be able to take needed elective courses. Discussion of the proposed plan at a council meeting revealed that teachers were focused almost entirely on test scores while parents focused on student needs. A parent council member stated:

> I have a concern because so much of what I'm hearing about curriculum is an echo of a test score rather than our children's current needs or future dreams.

Teachers agreed they were focused on the test. A teacher commented:

> The only thing that counts under KERA is test scores. That's one of the flaws of KERA... We end up doing exactly what they tell us to do because we don't want to be [sanctioned], and that's the only way you can go. And that chokes off stuff that is extremely valuable, like what [the parents] are saying.

At another school where special activities were planned prior to KIRIS testing to encourage students to do their best on the test, a parent expressed concern that the emphasis on doing well on the test was for the benefit of teachers more so than students:

> We're asking parents to... get [their children] excited and do the things that will help their child score better on the test. If they're doing these things and their child scores better, the school looks better and we get a reward, are they doing it just so the teachers will get more money? .... Are we as council members going to encourage parents to get their kids ready to be tested so the teachers can get more money, or is that money going to be put back into the system as a whole to help the whole school?

A fourth-grade teacher spoke of how her concern about improving assessment results almost caused her to lose sight of student welfare:

> At the beginning of the year when I first got the test scores, I was really worried and I was trying to push. And then one day I talked to a mother [whose] husband had kicked her and five children out of their home. She was taking these children to different places every night to live. She and her daughter were living in a car. And I said, "I don't care what that child does on the test; I want him to know that I care about him. And I want him to know that school is a safe place for him to be." That put me back on track for knowing what the kids need.... I believe they all will do the best they can, but as for me putting the pressure on them to succeed, I'm not going to, and I'm going to try to stop putting it on myself. In the classroom, I'll do all the activities I can that will help prepare them for the test and if they do well, they do well, and if they don't, then I know they've tried. It's one day out of their life...

Local Responses to the Accountability Measures

In the preceding section, we described the extent to which the accountability system forced change to happen in the four districts we studied. Many of these changes may have a positive effect over time if teachers come to see the value of them and implement the changes for their own sake. At the time our study concluded, however, educators' attitudes toward the accountability measures seemed to act as a barrier to their acceptance of the assessment-driven changes as worthwhile. As a rule, the accountability program was strongly resented by educators in the four districts.

Almost across the board, the educators we spoke to about the accountability program resented the measures as an insult to their professionalism. They felt that educators are motivated by intrinsic factors and that extrinsic rewards will not transform an incompetent, uncaring teacher into a competent, conscientious one. Many teachers said they try to do their best because that's their job. When asked if she was motivated by the promise of rewards or

threat of sanctions, a high school teacher remarked:

> I haven't really thought that much about it either way. I think that I work the way I work and work as hard as I work because of who I am; that's just me; but [also] because this is what my job is. If these are the goals of the school and these are objectives and we have this school transformation plan to go by and I'm expected to do that, we're supposed to be trying to implement the strategies that will help our school do well, and I think that's very important for our school, our students, and our community, then I guess I'm just a conscientious person. So that's more my motivation than thinking about sanctions or rewards. Obviously, I don't want to be a school in crisis, but I don't think I think about that so much as simply doing what I'm supposed to do.

Many teachers said they were not motivated by the promise of rewards because they did not believe the rewards would actually be conferred. A large number of teachers recalled legislation from the 1980s that would have provided teachers who successfully completed an evaluation process with a $300 bonus. Funding for that legislation never came through. When asked if she had been motivated by the promise of rewards, an elementary school teacher replied:

> No. We laugh about that. We think it's so funny. Of course, you've heard about the $300 we were supposed to get several years ago; did you ever see yours? They said this year, "You all will get rewards..." and we all just died laughing. We thought that was the biggest joke we'd ever heard. I said, "When they peel off the hundred dollar bills, I'll believe it. But until then, forget it." It's a standing joke around here: "Yeah, right, we're going to get more money. In your dreams."

While the majority of teachers who spoke to this issue said they had not been motivated by the promise of rewards, over one-third of the educators we spoke to on this issue in 1994-95 admitted that the threat of sanctions had motivated them to make changes in instruction and to push students to the limits of their capabilities. Thus, many educators found the threat of sanctions motivating in the negative sense. A high school teacher explained:

> The thought of losing your job is far more threatening, far more intimidating than receiving rewards. First you want to make sure you keep your job and keep the school open. You hear intimidating statements like if a school is a "school in crisis," you're going to have these master teachers or whatever they're called, Distinguished Educators come in and they're going to have dictatorial powers to close down the school... When the powers-that-be make a change and they say you've got to do something, then you've got to do it because they have the power to make life miserable for you if you don't.

Educators in the study districts also resented the use of an unproven assessment instrument to determine school success. They were skeptical about many aspects of the assessment, including the comparison of two different groups of students to determine school progress, the reliability of scoring, and the validity of the test for individual students. Some teachers questioned the validity of KIRIS because individual student scores were often inconsistent with student performance in the classroom. In addition, they wanted a test that provided useful student information; they did not understand how KIRIS could be reliable for schools but not for individual students.

Many teachers questioned the validity of KIRIS because they often could not identify what, if anything, they had done to cause the school's scores to go up or down. When asked to explain their scores, the educators we talked to usually attributed the scores to something outside the school's control--most commonly, having an unusually bright or slow class in the accountable grades that year (e.g., the "good class/bad class" theory), or the general make-up of the student body (e.g., having deprived vs. stimulating home backgrounds). Educators from schools that failed to meet their thresholds often attributed their lack of success to having a high baseline, which implies that they subscribed to the "good class/bad class" theory. A junior high teacher at a school where the high school scores fell below the baseline described the reaction of

high school teachers to the assessment results:

> They don't know why they dropped.... They're all thinking it's the composition of the kids that made this difference here.... We've scrambled around.... Nobody knows quite what to do, so everybody is coming up with ideas, some of which get tried, some of which have a longer life span than others, but there's not a cohesive sort of plan.... We don't know what the hell is going on, all right? And so [teachers] say, "One thing I know is that [the state is] weak legally." Really, this is what goes through people's heads.... They're long since over, "What can we do in here as a group to make these better?" and they're into, "How can I protect my..." They're overwhelmed big time. They're operating on reflexes because they don't know what to do.

We also saw scant evidence that the accountability measures motivated educators from different schools in the same district to work together for districtwide improvement. When results from the first biennium were announced and schools identified for rewards, there was a great deal of almost mean-spirited competitiveness among educators. In a district where it happened that the most innovative school was the only one that was not successful, several teachers in successful schools stated with some measure of satisfaction that they had done better because they had eschewed much of KERA. Sources in three of the four districts privately accused districts or schools that had been rewarded of previewing the test, setting artificially low baseline scores, or otherwise "cheating" on the test. (In only one case did we find convincing evidence of a questionable practice: one school administrator reported that teachers typed students' portfolio entries and corrected them; it was his impression that "everyone" did this.) Some educators in non-rewards schools in two districts were so incensed about the whole business that they reported grossly inaccurate comparisons between their own schools' scores and those of rewards schools (for instance, reporting that a school that didn't earn rewards had the highest absolute scores in the district when this was not the case). Generally, there was very

little willingness to attribute a successful school's performance to anything positive the teachers had done.

DISCUSSION

In this paper, we provided evidence that the accountability measures have begun to drive the reform effort to some extent in that teachers changed their instruction to focus on writing and schools were adding new courses to the curriculum. In addition, there was evidence that at least a few teachers were making a stronger effort to reach all students. In some instances, educators seemed to be internalizing the philosophy behind the changes they made because they saw the positive effects on students. More commonly, however, educators taught to the test because they were forced into it, and they often resented this.

The designers of KIRIS hoped that the test would mirror "good" instruction to such an extent that teaching to the test would be a worthwhile endeavor, and perhaps they were right. Many educators and parents reported to us that students' writing abilities improved tremendously as a result of the increased emphasis on writing, and that students had become better thinkers through practice in analyzing and explaining their work. Also, the addition of courses such as arts and humanities to the curriculum is likely to have a positive effect on students.

It is disheartening, however, to find educators more preoccupied with saving their jobs than with helping students achieve. It is likely that the framers of KERA envisioned an educational system in which everyone worked together toward the ultimate goal of helping all students achieve at high levels, with the assessment program serving as an instrument for measuring whether that occurred. What has happened so far is that there is a movement toward

everyone working to help students do well on KIRIS as an end in itself. While it is hoped that KIRIS mirrors the right kinds of instruction and high levels of achievement, a single assessment instrument can never measure everything that should be taught in school. The temptation for Kentucky teachers is to concentrate on the changes likeliest to raise KIRIS scores in the short term.

Others who have studied measurement-driven reform have also been troubled by the wrong-headed focus that results when assessment drives instruction. Noble & Smith (1994) charged that tying performance-based assessment with high-stakes accountability for teachers is philosophically inconsistent because, while the assessment promotes a constructivist view of student learning, the accountability that is tied to the assessment is based on a behaviorist view of teacher learning:

> Reform mandated through measurement flies in the face of what cognitive scientists refer to as intentional learning, that is, learning desired and controlled by the learner. Individual construction of knowledge depends strongly on the sense of being in charge of their learning. When teachers do not feel in control of their own professional lives, they act passively, become compliant, and act automatically without reflecting on their own beliefs (p. 132).

Torrance (1993), who studied measurement-driven reform in Britain, does not believe that assessment should drive instruction:

> Developing innovative approaches to assessment should certainly proceed in tandem with attempts to improve instruction, but they cannot drive it. The conditions under which good teaching may develop can be elucidated, but cannot be so easily captured and manifested in a single teaching and assessment package. Ultimately, if our definition of high-quality education includes flexibility, creativity, and the capacity to deal with the unexpected--for teachers and students alike--then such characteristics must be allowed in any system that is designed and within which teachers and students are expected to work...(p. 88)

In Kentucky, the questionable validity and reliability of the assessment instrument makes

its use as a lever for reform especially problematic. Two recent evaluations of the KIRIS assessment and accountability program recommended that its reliability and validity be more strongly established (Hambleton et al, 1995; Western Michigan University, 1995). As for our own analysis, our study sample was too small to provide a basis for studying the validity of KIRIS as a measure of school success. However, the longitudinal, qualitative nature of the study gave us enough information to at least conjecture about why the 20 schools in our study performed as they did on KIRIS. To our surprise, we could perceive no pattern to how the schools performed. In one district, the schools that had done the least to implement KERA--all of which happened to be outlying, "community" schools with the lowest enrollments--earned rewards, while the two central schools, one of which had implemented KERA wholeheartedly, failed to earn rewards. In this district, KIRIS results led some educators to believe that the way to earn rewards was to stick with traditional methods as much as possible. This pattern did not hold out in any of the other districts. In fact, the opposite appeared to occur in one district, where the most innovative school earned the largest rewards in the district. Also, the pattern of outlying and/or small schools earning rewards did not hold out in all four districts. Schools with SBDM councils did not perform better than schools without councils. Schools that had engaged in schoolwide improvement efforts did not appear, as a group, to perform any better on KIRIS than schools where teachers were left to their own devices. If this lack of pattern holds up over time and for a larger number of school districts, the validity of KIRIS for making judgments about school success will be called into question.

The upshot of all of this is that, to date, KERA has resulted in many changes, some of them substantive. Systemic reform, however, has not yet occurred in the sense that all the pieces

of the reform effort are working in concert toward the goal of improved student achievement. While some would argue that it is too early for this to happen and that the accountability program will eventually force systemic reform, it is our view that the accountability measures may actually impede the philosophical reform that is required to make the kinds of changes envisioned in KERA. If all students are to achieve at high levels, then the focus must be on students--not on an instrument that is supposed to measure their achievement. KERA is a step in the right direction in that it provides substantial money and resources to helping teachers focus on meeting the needs of individual students. The change effort might be even more successful if the focus was placed more singularly on building teacher capacities than on rewards and punishments as they are now conceived.

References

Almasi, J. F., Afflerbach, P. P., Guthrie, J. T., & Schafer, W. D. (1994, April). The impact of a statewide performance assessment program on classroom instructional practice in literacy. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Appalachia Educational Laboratory (1994, December). Instruction and assessment in accountable and non-accountable grades. Notes From the Field, 4(1). Charleston, WV: Appalachia Educational Laboratory.

Aschbacher, P. (1994). Helping educators develop and use alternative assessments: Barriers and facilitators. Educational Policy, 8(2), 202-223.

Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delta Kappan, 68(9), 683-686.

Bridge, C. A., Compton-Hall, M., & Cantrell, S. C. (1996, January). Classroom writing practices revisited: The effects of statewide reform on classroom writing practices. Lexington, KY: Institute on Education Reform, University of Kentucky.

Coe, P., Kannapel, P., & Lutz, P. (1991, February). Initial reactions to the Kentucky Education Reform Act in six rural school districts. Charleston, WV: Appalachia Educational Laboratory.

Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. Phi Delta Kappan, 73(3), 220-225.

Firestone, W. A. (1991). Merit pay and job enlargement as reforms: Incentives, implementation, and teacher response. Educational Evaluation and Policy Analysis, 13(3), 269-288.

Freedman, S. W. (1994, June). School reform through examinations: Lessons from the British experience. Occasional Paper No. 38, National Center for the Study of Writing. Berkeley, CA: University of California.

Haertel, E. H. (1994). Theoretical and practical implications. In T. R. Guskey (Ed.), High stakes performance assessment: Perspectives on Kentucky's educational reform (pp. 65-75). Thousand Oaks, CA: Corwin Press.

Hambleton, R. K., Jaeger, R. M., Koretz, D., Linn, R. L., Millman, J., & Phillips, S. E. (1995, June). Review of the measurement quality of the Kentucky Instructional Results Information System, 1991-1994. Frankfort, KY: Kentucky General Assembly, Office of Education Accountability.

Hatch, J. A. & Freeman, E. B. (1988). Who's pushing whom? Stress and kindergarten. Phi Delta Kappan, 70(2), 145-147.

Kentucky Department of Education (1994). 1994 Kentucky school laws. Frankfort, KY: Kentucky Department of Education.

Kifer, E. (1994). Development of the Kentucky Instructional Results Information System (KIRIS). In T. R. Guskey (Ed.), High stakes performance assessment: Perspectives on Kentucky's educational reform (pp. 7-18). Thousand Oaks, CA: Corwin Press.

Koretz, D., Stecher, B., & Deibert, E. (1992, August). The Vermont portfolio assessment program: Interim report on implementation and impact, 1991-92 school year. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education, UCLA.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. Educational Evaluation and Policy Analysis, 15(1), 1-16.

McLaughlin, M. W. (1991). Test-based accountability as a reform strategy. Phi Delta Kappan, 73(3), 248-251.

Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." Educational Policy, 8(2), 111-136.

Popham, W. J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68(9), 679-682.

Shepard, L. A. (1991). Will national tests improve student learning? Phi Delta Kappan, 73(3), 232-238.

Soodak, L. C. & Martin-Kniep, G. (1994). Authentic assessment and curriculum integration: Natural partners in need of thoughtful policy. Educational Policy, 8(2), 183-201.

Torrance, H. (1993). Combining measurement-driven instruction with authentic assessment: Some initial observations of national assessment in England and Wales. Educational Evaluation and Policy Analysis, 15(1), 81-90.

Urdan, T. C. & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. Educational policy, 8(2), 137-156.

Western Michigan University, The Evaluation Center. (1995, January). An independent evaluation of the Kentucky Instructional Results Information System (KIRIS). Frankfort, KY: Kentucky Institute for Education Research.