

PERFORMANCE AUDIT REPORT

Reviewing the Use of State Assessment Tests In Kansas

**A Report to the Legislative Post Audit Committee
By the Legislative Division of Post Audit
State of Kansas
June 1996**

Legislative Post Audit Committee

Legislative Division of Post Audit

THE LEGISLATIVE POST Audit Committee and its audit agency, the Legislative Division of Post Audit, are the audit arm of Kansas government. The programs and activities of State government now cost about \$7 billion a year. As legislators and administrators try increasingly to allocate tax dollars effectively and make government work more efficiently, they need information to evaluate the work of governmental agencies. The audit work performed by Legislative Post Audit helps provide that information.

We conduct our audit work in accordance with applicable government auditing standards set forth by the U.S. General Accounting Office. These standards pertain to the auditor's professional qualifications, the quality of the audit work, and the characteristics of professional and meaningful reports. The standards also have been endorsed by the American Institute of Certified Public Accountants and adopted by the Legislative Post Audit Committee.

The Legislative Post Audit Committee is a bipartisan committee comprising five senators and five representatives. Of the Senate members, three are appointed by the President of the Senate and two are appointed by the Senate Minority Leader. Of the Representatives, three are appointed by the Speaker of the House and two are appointed by the Minority Leader.

Audits are performed at the direction of the Legislative Post Audit Committee. Legislators or

committees should make their requests for performance audits through the Chairman or any other member of the Committee. Copies of all completed performance audits are available from the Division's office.

LEGISLATIVE POST AUDIT COMMITTEE

Senator Lana Oleen, Chair
Senator Anthony Hensley
Senator Phil Martin
Senator Alicia L. Salisbury
Senator Don Steffes

Representative James E. Lowther, Vice-Chair
Representative Tom Bradley
Representative Duane Goossen
Representative Sheila Hochhauser
Representative Ed McKechnie

LEGISLATIVE DIVISION OF POST AUDIT

800 SW Jackson
Suite 1200
Topeka, Kansas 66612-2212
Telephone (913) 296-3792
FAX (913) 296-4482

The Legislative Division of Post Audit supports full access to the services of State government for all citizens. Upon request, Legislative Post Audit can provide its audit reports in large print, audio, or other appropriate alternative format to accommodate persons with visual impairments. Persons with hearing or speech disabilities may reach us through the Kansas Relay Center at 1-800-766-3777. Our office hours are 8:00 a.m. to 5:00 p.m., Monday through Friday.



LEGISLATURE OF KANSAS
LEGISLATIVE DIVISION OF POST AUDIT

MERCANTILE BANK TOWER
800 SOUTHWEST JACKSON STREET, SUITE 1200
TOPEKA, KANSAS 66612-2212
TELEPHONE (913) 296-3792
FAX (913) 296-4482

June 6, 1996

To: Members, Legislative Post Audit Committee

Senator Lana Oleen, Chair
Senator Anthony Hensley
Senator Phil Martin
Senator Alicia Salisbury
Senator Don Steffes

Representative Jim Lowther, Vice Chair
Representative Tom Bradley
Representative Duane Goossen
Representative Sheila Hochhauser
Representative Ed McKechnie

This report contains the findings, conclusions, and recommendations from our completed performance audit, *Reviewing the Use of State Assessment Tests in Kansas*.

The report also contains appendices showing the results of surveys of school superintendents, principals, and teachers concerning their use of the State assessment tests, and a description of the steps taken to ensure that the assessment tests were reliable and valid.

This report includes several recommendations for ensuring that the assessment results are used in appropriate ways, and for ensuring that steps are taken to continue to monitor the reliability and validity of the assessment tests. We would be happy to discuss these recommendations or any other items in the report with any legislative committees, individual legislators, or other State officials.


Barbara J. Hinton
Legislative Post Auditor



Reviewing the Use of State Assessment Tests in Kansas

EXECUTIVE SUMMARY LEGISLATIVE DIVISION OF POST AUDIT

Question 1: Are There Limitations as to How the Information Obtained From the State Assessment Tests Should Be Used, and How Is It Being Used?

In a broad sense, the type of assessment tests selected and the way they were developed imposes some overall limits on the use of assessment test results. page 12
Because the assessment tests specifically measure students' progress toward meeting the Kansas curriculum standards, no national comparisons can be made. Also, because each assessment test goes through a pilot development stage, results from early versions of the test can't be compared with results from later versions.

The assessment tests were designed so that, by themselves, the test scores would be reliable and valid when used to measure students' average performance at individual schools. page 15
Concerns had been raised about the reliability and validity of the assessment tests. The University of Kansas' Center for Educational Testing and Evaluation, the contractor that developed the tests, followed the steps required by professional standards to ensure the assessment tests were valid and reliable for measuring students' average performance at individual schools. But, officials from the Department of Education and the Center said the assessment test results shouldn't be used as the sole factor in comparing individual student performance or in making comparisons between schools or districts, because of the many other factors that can affect such comparisons.

In general, school administrators indicated they are using State assessment test scores as designed. page 18
They are using the results to see whether their students are being taught what they're expected to know, to identify changes that may be needed in what or how they teach, and to try to improve those test results over time.

Because of confusion about how test results should be used, assessment test results sometimes are being used for purposes they weren't designed for. page 18
A significant number of school administrators and teachers said they're using the assessment test results to compare individual students' achievement. Some superintendents and principals said they were using the assessment test results to make comparisons between schools and districts. A small number said they had or were considering tying students' graduation requirements to their assessment test scores, or were using the results to evaluate teacher performance. Recent actions by the Legislature— including mandating a school building report card and student performance expectations—will either require or encourage assessment test results to be used for purposes they weren't designed for.

School officials and others we talked with or surveyed during the audit had both positive and negative opinions about the value of the assessment tests. *Those who liked the assessment tests were most positive about how the results could improve education in Kansas. Those who didn't like the tests generally voiced concerns about the usefulness of the information they provided, the consistency of scores, and the amount of time it took to administer them.* page 21

Members of the public also expressed praise and concern about the State assessment tests. *Although some people we spoke with said they thought the assessment tests were beneficial and valuable, others expressed concerns that the assessment tests included inappropriate attitude-type questions and had reading selections that contained inappropriate subject matter. We found that the attitude questions no longer are asked. Also, the Department tries to minimize controversy over the reading passages through its selection process.* page 23

Conclusion page 25

Recommendations page 26

Question 2: Question 2: How Do Kansas' State Assessment Tests— and Its Methods of Measuring Student Performance— Compare With Other States?

Kansas is further along than more than half the states we contacted in the development of its assessment tests. page 27
Five of the 13 states we reviewed—Illinois, Kentucky, Maine, Maryland, and Oklahoma—are using assessment tests, which generally are similar to those being used in Kansas. The main differences we noted were that the other states tended to have centralized scoring of their tests, and also required other standardized tests to be administered. Two other states, Arizona and California, used assessment tests in the past, but dropped them because the results were being used for purposes they weren't designed for. Both states are developing new assessment tests.

APPENDIX A: Comparing the State Assessment Tests With the Iowa Test of Basic Skills and the National Assessment of Educational Progress page 33

APPENDIX B: Evaluating Reliability and Validity page 37

APPENDIX C: Summary of Interviews with School District Superintendents page 41

APPENDIX D: Summary of Survey Responses from Principals page 45

APPENDIX E: Summary of Survey Responses from Teachers page 51

APPENDIX F: Agency Response page 57

This audit was conducted by Ellyn Sipp, Scott Brunner, and Trish Pfannenstiel. If you need any additional information about the audit's findings, please contact Ms. Sipp at the Division's offices. Our address is: Legislative Division of Post Audit, 800 SW Jackson Street, Suite 1200, Topeka, Kansas 66612. You also may call (913) 296-3792, or contact us via the Internet at: LPA@PostAudit.ksleg.state.ks.us.



Reviewing the Use of State Assessment Tests in Kansas

Quality Performance Accreditation, mandated by the State Board of Education in 1991, is the process the Board uses to accredit schools. This process bases accreditation decisions on outcomes that show how well schools are teaching students what they need to learn, instead of on such things as the number of library books a school has, or the number of students per teacher.

The State Board of Education also has established *Kansas Curriculum Standards*. These standards represent the knowledge, skills, and abilities that Kansas children should achieve in the major academic subject areas of mathematics, reading, writing, science, and social studies.

To measure progress toward meeting those standards, the State Board contracted with the University of Kansas' Center for Educational Testing and Evaluation to develop State assessment tests in each of the major academic subject areas. The Board mandated that those tests be administered by all schools in Kansas, as part of the accreditation process to help provide data to use to improve schools and show schools' improvement.

Legislators and others have raised concerns about a number of aspects of the State assessment tests. Some questions center on the selection, purpose, and use of these tests—for example, why the State Board chose these tests as opposed to other tests or indicators that could be used. Other concerns center on the frequent changes made to the tests, which precludes most year-to-year comparisons. Still other questions relate to how Kansas' performance measurement methods and assessment tests compare with other states. This performance audit addresses the following questions:

- 1. Are there limitations as to how the information obtained from the State assessment tests should be used, and how is it being used?**
- 2. How do Kansas' State assessment tests—and its methods of measuring student performance—compare with other states?**

To answer these questions, we interviewed officials from the Department of Education and the University of Kansas. We spoke with experts about the steps needed to evaluate the adequacy of assessment tests, and reviewed studies that described the steps that had been taken. We surveyed a sample of public and private school superintendents, principals, and teachers, and interviewed other interested people about the assessment tests. Finally, we contacted officials in a number of other states about their programs to measure student progress.

In conducting this audit, we followed all applicable government auditing standards set forth by the U. S. General Accounting Office.

In general, we found that the type of assessment tests the Board of Education has chosen—criterion-referenced tests tied to Kansas curriculum standards—prevents any comparisons of test scores with students in other states. Because the tests have had to be developed from scratch, changes that have had to be made during the pilot phase of development also have temporarily limited year-to-year comparisons. The

assessment tests were designed so that, by themselves, the test scores would be reliable and valid when used to measure students' average performance at individual schools. Officials from the Department of Education and the University of Kansas' Center for Educational Testing and Evaluation said the assessment test results shouldn't be used as the sole factor in comparing individual student performance or in making comparisons between schools or districts, because of the many other factors that can affect such comparisons.

School administrators generally indicated they are using State assessment test scores as designed. However, because of confusion about how test results should be used, they sometimes are being used for purposes they weren't designed for. Recent actions by the Legislature will either require or encourage the assessment test results to be used for such purposes. School officials and others had both positive and negative opinions about the value of the assessment tests. Specific concerns were that the tests don't provide any new information, scoring of performance components is inconsistent, and the tests take too long to administer. Teachers also expressed both positive and negative opinions about the performance components—they generally liked the performance component of the writing assessment test, but disliked that aspect of the social studies and mathematics tests. Members of the public had some concerns that attitude questions were being asked on the assessment tests; we found those types of questions no longer are included.

We also found that because of federal initiatives, other states also are developing assessment tests. Although Kansas is further along than about half the 13 states we contacted, at least five other states have assessment test programs in place. Like Kansas, these states give students at the elementary, junior high, and high school levels criterion-referenced assessment tests that measure students' progress toward each states' curriculum standards, goals, or expectations. Like Kansas, these states use the results to measure students' improvement in individual schools over time. The tests in these other states include a variety of types of questions, although the emphasis is on open-ended and performance-based questions. Finally, school districts in these states also continue to give standardized tests to their students. In two states, that is a state requirement. In the other three states, as in Kansas, giving standardized tests is a local school district option.

Overview of the State Assessment Tests

In the late 1980s, the State Board of Education initiated efforts to improve Kansas education. This effort was in response to concerns expressed by business leaders that students were graduating without the skills needed to be successful later in life. Business leaders were especially concerned that Kansas students didn't have basic problem-solving, reasoning, critical-thinking, and communication skills, and that they were unable to assimilate new and unfamiliar information and work cooperatively in teams.

In 1989, the Board of Education Began Developing Curriculum Standards and Assessment Tests That Would Provide a Way of Assessing Students' Achievement of Those Standards

These efforts involved a major restructuring of Kansas schools and what they taught. Instead of ensuring that Kansas students could meet some minimum level of competency, schools now would be expected to ensure that students could meet challenging academic standards and develop higher-order thinking skills, as reflected in the Kansas curriculum standards.

These Kansas curriculum standards shouldn't be confused with the notion of a "standard curriculum," which might specify the courses to be taught and the books to be used, or say that a district needs to offer four units of English and five units of mathematics. Instead, the Kansas curriculum standards define what "outcomes" (or knowledge, skills, and abilities) Kansas students should possess by the time they have completed certain grade levels. The focus of those standards also shifted. For example, the standards for mathematics de-emphasize mathematical computations, and emphasize problem-solving and estimation skills. Similarly, the standards for social science focus on the process of understanding how to pull information together and apply it to real-world situations, rather than emphasizing specific facts or dates of historical significance.

According to Department of Education officials, the Kansas curriculum standards were designed to be "world-class." This meant they would be at a level equal to the best standards of any other country, and high enough so that Kansas students could compete in the world economy.

In 1990, the Board of Education approved the first curriculum standards in the area of mathematics. Over the next two-to-three years, standards also were developed for reading, writing, science, and social studies. In developing the initial draft standards, Department staff convened advisory committees made up of education officials and teachers who were familiar with national standards in their subject areas. These individuals wrote the initial draft standards.

For each subject area, the draft standards were then sent out for review to teachers, higher education faculty, professional organizations, business representatives, and community members. In addition, the standards were reviewed by a group of experts involved in setting national curriculum standards to ensure they were

“world-class.” The accompanying profile provides examples of the curriculum standards developed for Kansas students in mathematics, reading, and writing.

Beginning in late 1990, the Board of Education contracted with the University of Kansas’ Center for Educational Testing and Evaluation to develop assessment tests to measure schools’ progress toward meeting the new curriculum standards. The first assessment test developed was for mathematics. In developing this test, the Center contacted school principals and asked them to nominate teachers to be trained and to write potential assessment test questions for mathematics. The Center then used these questions to develop a pilot mathematics assessment test that was first given to students in Spring 1991.

Since then, assessment tests also have been developed for the four other subject areas—reading, writing, science, and social studies. For each subject area tested, teachers have submitted potential questions. Because the focus of these assessment tests has been on higher-order thinking skills, a variety of formats has been used, including multiple-choice questions, multiple-mark questions, open-ended questions, written narratives, group projects, and extended projects. The profile at right lists some sample assessment test questions designed to measure specific curriculum standards.

The grade levels at which the State assessment tests are administered, and the frequency of those tests, are shown in the accompanying table.

State Assessment Test Schedule

<u>Subject Area</u>	<u>Grades Tested</u>	<u>Frequency</u>
Mathematics	4,7,10	every year
Reading	3,7,10	every year
Writing	5,8,10	every other year
Science	5,8,11(a)	every other year
Social Studies	5,8,11	every other year

(a) The science test will be given in grade 10 in the 1996-97 school year.

According to Department officials, developing tests to assess whether students were being taught what they were expected to know under the Kansas curriculum standards was viewed as one of the quickest ways of ensuring that those standards were adopted. If schools were tested against particular standards, they would be more likely to adopt them. Developing assessment tests to measure progress toward the State’s curriculum standards also helped ensure that teaching materials and methods would be changed, if needed, to improve what students were taught, and how they could apply that knowledge to “real-world” situations.

In 1992, the Legislature passed a law that piggybacked on what the Board was already doing in this area, and that required the Board to establish an accreditation process. That process—quality performance accreditation—was to be based in part on curriculum standards that specified what Kansas school children



What Are Kids in Kansas Expected to Know?

MATHEMATICS

Grade 4

Curriculum Standards or Outcomes:

- the student should be able to explain, model, and perform operations with whole numbers to solve problem situations
- the student should explore operations with fractions, mixed numbers, and decimals

Sample Problem:

Luis wants to buy a dirt bike that costs \$180 (not including tax). He earns \$2.00 an hour babysitting. He has saved \$65.
About how many more hours will he need to work?

READING

Grade 3

Curriculum Standards or Outcomes:

- the student should be able to read, listen, and view for understanding and interpretation
- the student should be able to summarize and interpret a message and purpose intended by the presenters



Sample Problem:

This was the big day. Sue and her father were going to build a doghouse. They had drawn the plans for the house. Her mother helped her buy the wood. They bought some nails and paint. Sue was sure they had all they would need. She could not wait for breakfast to be over. Then they could get started. It had taken so long for this day to come.

Who was going to build a doghouse? Mark all that are correct.

- | | |
|------------------|-----------------|
| a) Sue's sister | d) Sue's father |
| b) Sue's brother | e) Sue's mother |
| c) Sue | f) a friend |

WRITING

Grade 5

Curriculum Standards or Outcomes:

- students should be able to write and speak with clarity, power, and purpose
- students should be able to write for audiences outside their classroom

Sample Problem:

Students are asked to write about a subject listed on the assessment test or choose a topic of their own. Once they have chosen a topic the student will:

- think about what he/she wants to write
- write a rough draft
- revise and edit the rough draft making any changes he/she wants
- recopy the paper and proofread the final draft



should know by different grade levels. It also was to include an assessment component to assess students' achievement of those standards. Under the law, standards and tests for assessing them were to be developed and implemented at three grade levels—elementary, middle, and high school level.

As shown in the graphic at right, assessment test results were only one of the measures of school improvement required as part of the school accreditation process. Other factors included attendance rates, drop-out rates, graduation rates, classroom grades, and other test scores. Also, the accreditation process covered many areas besides school improvement, including the development of a school mission statement, a school needs assessment, community involvement, and staff development.

What Do the Following Terms Mean?

Types of Tests:

Assessment such tests measure a student's knowledge, skills, and abilities through a variety of testing methods and techniques, such as multiple-choice, multiple-mark, open ended, and performance-based questions

Areas Measured:

Curriculum Standard what students should know and be able to do by the end of a particular grade

Standards of Excellence expected level of performance in selected subject areas, adopted by the Board of Education in 1994 according to legislative mandate

Measurement Quality:

Reliability indicates whether or not a specific test, when administered on different occasions, will produce the same test results

Validity indicates whether the test is really measuring what it was designed to measure

Types of Questions:

Multiple-Choice the student must select the correct answer from a list of possible responses—there is only one correct response

Multiple-Mark a form of multiple-choice question where there may be more than one correct response—all correct responses must be selected for full credit

Open-Ended the student must provide an answer; no list of possible responses is offered for selection. The student must construct his/her own response, demonstrating his/her understanding of the problem and subject area. This type of question examines what a student knows and how he/she uses that knowledge to solve "real-world" type problems.

Performance-Based includes portfolios, group projects, open-ended questions, interviews, and the like which focus on higher-order thinking skills

Where Do the State Assessment Tests Fit Into QPA?

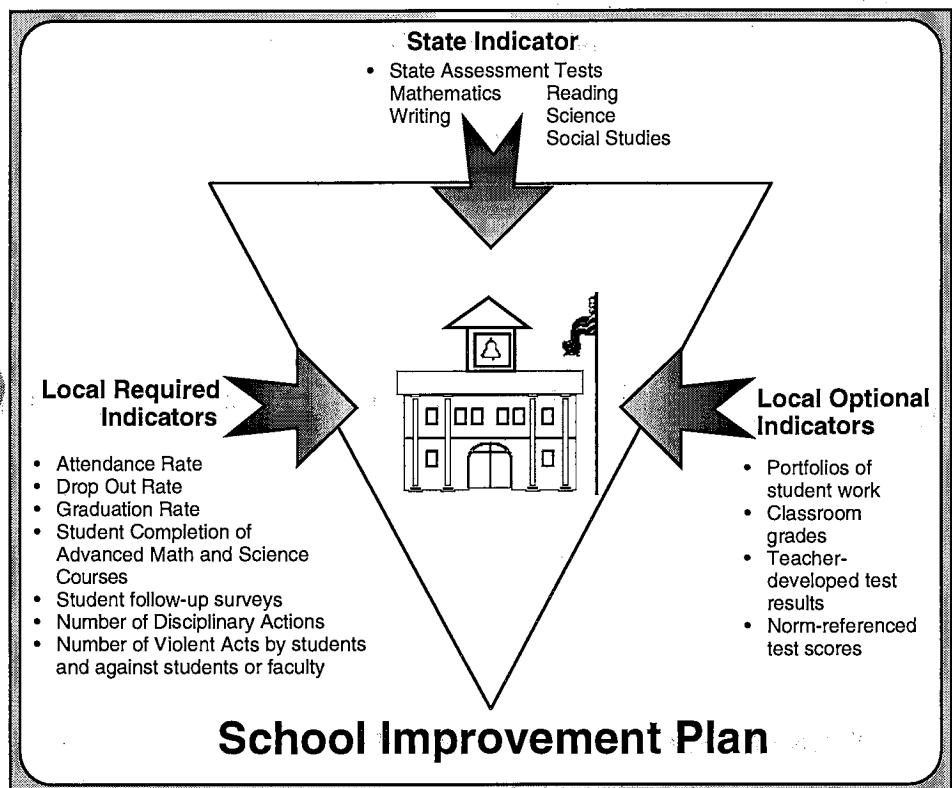
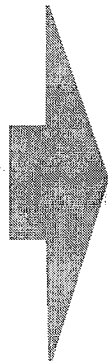
In Kansas, schools are accredited by the State Board of Education under an accreditation program known as Quality Performance Accreditation, which focuses on five areas:

Areas of
Focus
Under
Quality
Performance
Accreditation

- **community involvement**—includes a site council comprised of school administrators, teachers, parents, and local business people that develop the school's mission statement and determine the need for community programs, such as preschool and adult education
- **staff development**—includes on-going staff training related to a school's mission statement, academic focus, and school improvement plan
- **high academic performance**—requires implementing a school improvement plan based on challenging "world-class" standards (in Kansas these are known as *Kansas Curriculum Standards*) that outline steps to help students attain expected skills, knowledge, and abilities
- **integrated curriculum**—schools should incorporate such things as reading, writing, problem-solving, group projects, and the like in classroom instruction to encourage higher-order thinking skills by students
- **school improvement**—schools are to establish and maintain high expectations for students, and continuously monitor students' achievement towards the *Kansas Curriculum Standards*

As shown in the graphic below, the State assessment test results are only one of the indicators used under Quality Performance Accreditation in determining school improvement.

- Community Involvement
- Staff Development
- High Academic Performance
- Integrated Curriculum
- School Improvement



Many of these areas are developed in the first three years of the four-year accreditation cycle. At the end of the fourth year, a school must be able to show improvement towards quality performance accreditation outcomes and goals set by the school. An on-site accreditation team visits the school, reviews a variety of indicators that show school improvements, and makes a recommendation to the Board regarding accreditation.

The 1995 Legislature amended the law to require that the school performance accreditation system be based, in part, upon improvement in performance on the curriculum standards.

Information from the assessment tests is supposed to be used to help schools determine their progress at meeting the Kansas curriculum standards. It also is supposed to be used to identify areas of weaknesses, so that schools can make appropriate changes in their teaching materials, instruction methods, or other areas to help them improve. Such changes, it's hoped, will increase students' average scores, over time.

The Board of Education has specified "standards of excellence" that identify the average score students in each school are expected to achieve on the assessment tests. When students take assessment tests, their scores are averaged, by grade and by subject area, for each school. For instance, each school will get an average score for the 4th-grade mathematics assessment tests. For this test, the intent is to see whether schools have taught students the things Kansas curriculum standards say they'll need to know by the time they're in the 4th grade.

The Board has adopted a standard of excellence for each subject area, according to legislative mandate. For instance, the standard of excellence for the 4th-grade mathematics assessment is 75%. That means if a school's average score on the 4th-grade mathematics assessment is 70%, the school is performing below the standard of excellence. If the school's average score is 80%, it's performing above the standard of excellence.

To-date, most schools are performing below the specified standard of excellence. The charts on pages 9-11 show the most recent Statewide average scores for the mathematics, reading, and writing assessments, and the related standards of excellence.

There's Been a Similar Shift Nationwide in the Development of Curriculum Standards and Assessment Tests

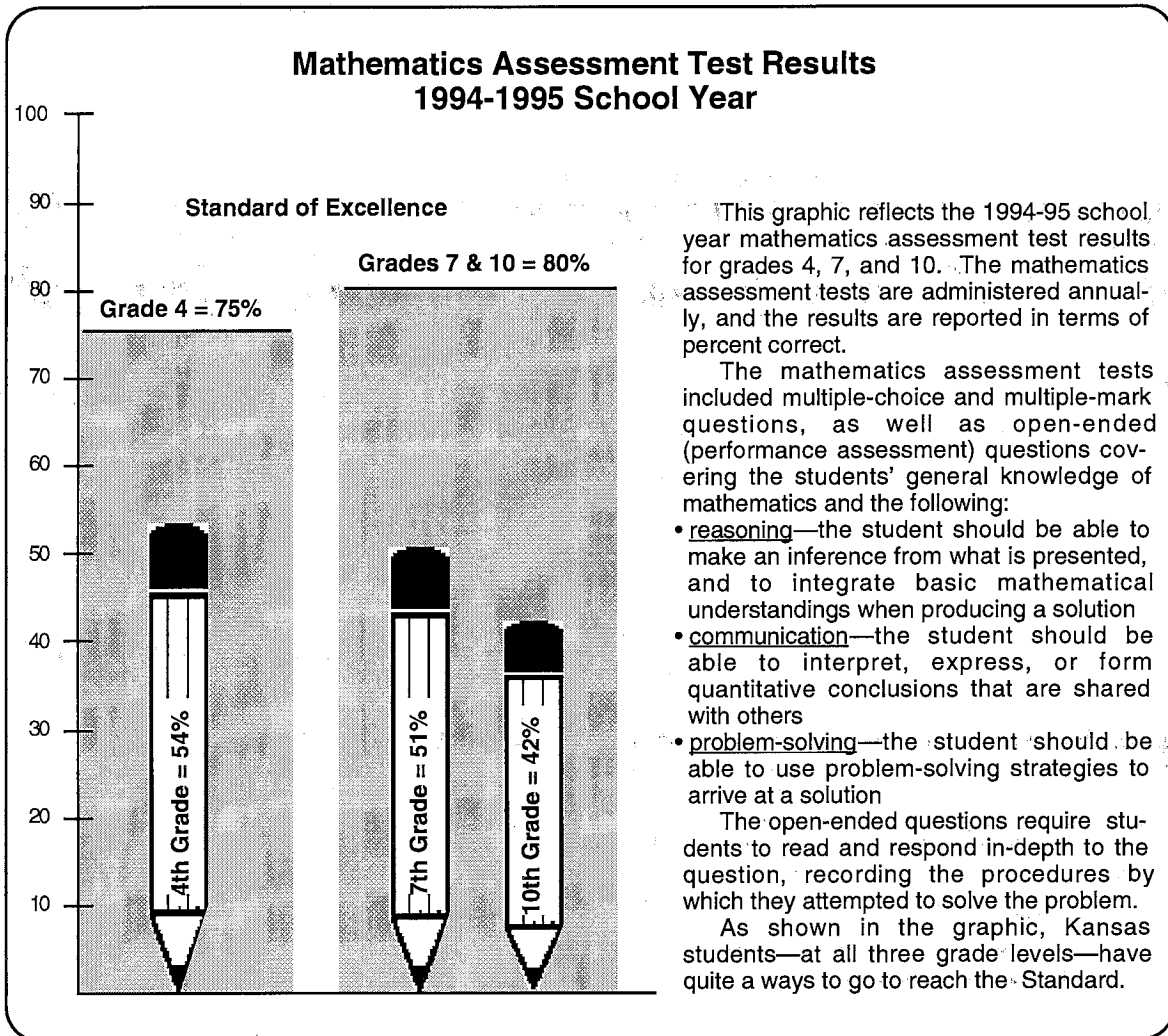
For example, at the federal level, Congress enacted the Goals 2000: Educate America Act in 1994. That Act outlined a series of educational goals, including the following:

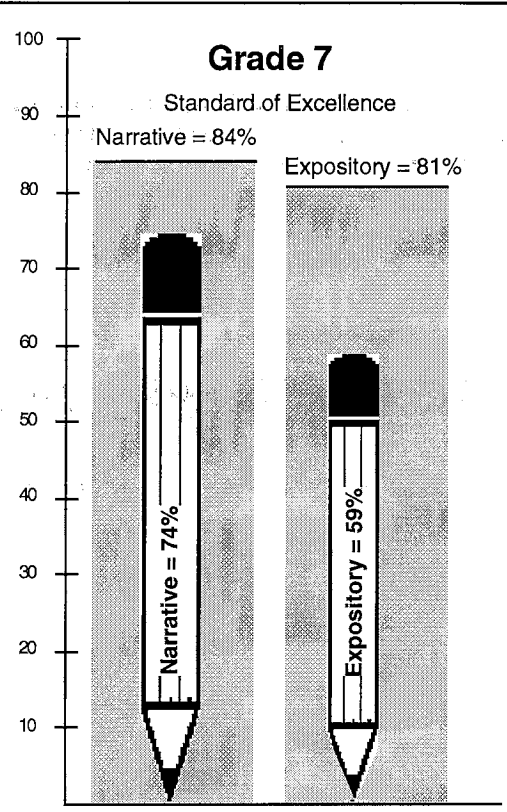
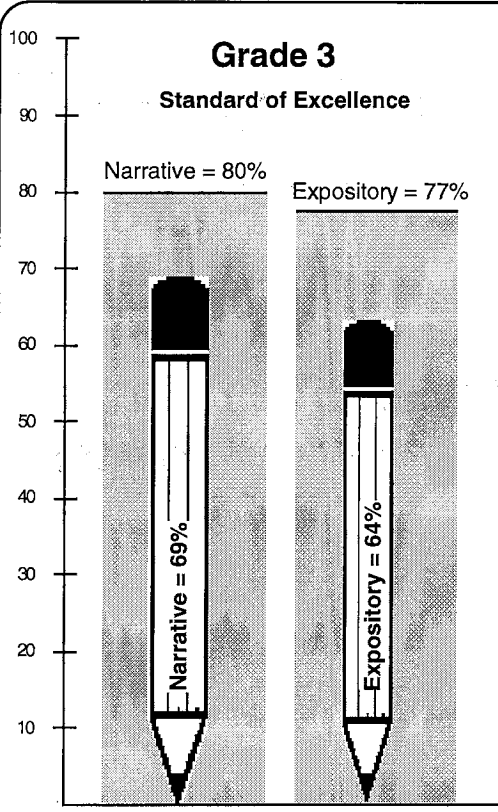
- all students will demonstrate competency over challenging subject matter
- United States students will be first in the world in mathematics and science
- the high school graduation rate will increase to at least 90%

To receive federal funding under the Act, states are required to develop a set of high-quality student assessments that test each state's standards. The assessment tests must be used as the primary means of determining the yearly performance of each school. Finally, the assessment tests must be given at the elementary, junior high, and high school levels.

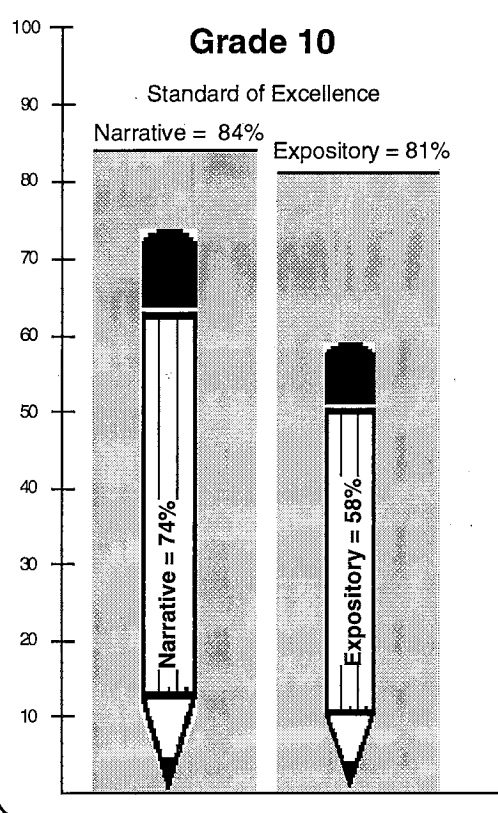
A related federal law, Title I of the Improving America's Schools Act, provides funds for "at-risk" students (students who are more likely to drop out of school, or have academic problems because of low family income or other disadvantages). To be eligible for these funds, schools also have to develop state assessment tests at three grade levels. These have to be designed to show how at-risk students in a school are doing, and to measure improvements in students' scores at that school.

Other states also have been involved in developing curriculum standards and assessment tests, although Kansas was further along in the development of its assessment tests than more than half the states we contacted during this audit. More information about other states' programs will be discussed in question two of this audit.





Reading Assessment Test Results 1994-1995 School Year

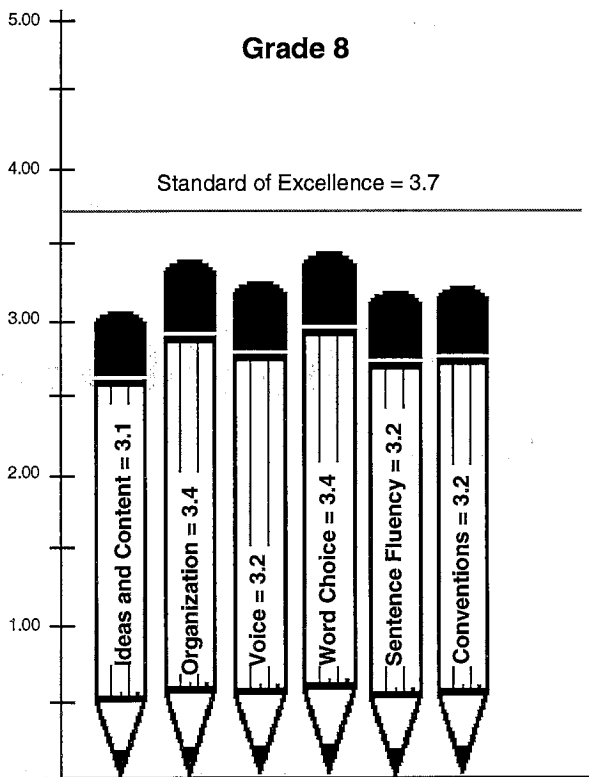
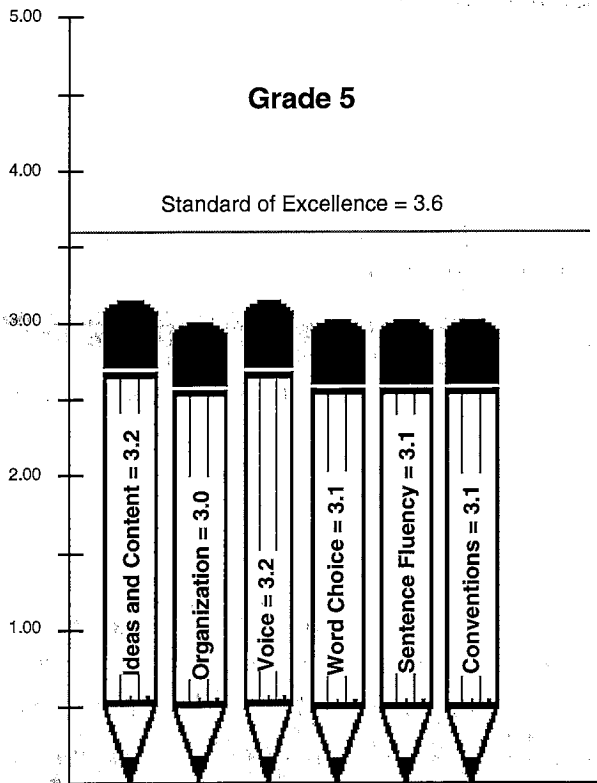


The graphic on this page reflects the 1994-95 school year reading assessment test results for grades 3, 7, and 10. Reading comprehension was the major focus of the test. The reading assessments tests are administered annually, and the results are reported in terms of percent correct.

The students at each grade level tested read and answered questions about two different kinds of selections. One selection was a narrative passage which told a story. The other selection was an "expository" passage which provided information to the student similar to that found in a "how-to" manual or an encyclopedia. Multiple-mark (more than one right answer) and written-response questions were used to determine the students' understanding of the passages. The written response questions were scored locally by teachers using a scoring guide.

As shown in the graphic, Kansas students do relatively well on the narrative, but appear to have problems with the expository selections.

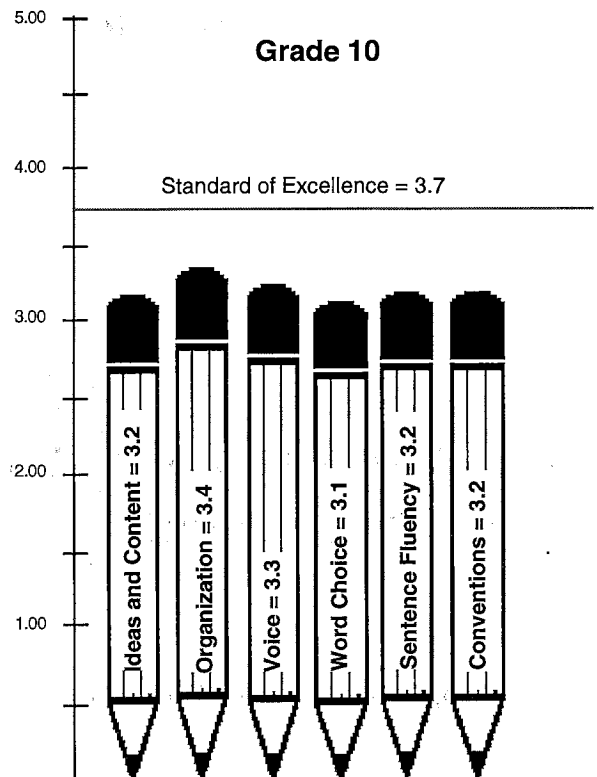
Writing Assessment Test Results 1993-94 School Year



The graphic displays the 1993-94 school year writing assessment test results for grades 5, 8, and 10. The writing assessment tests are administered every other year, and the results are reported based on a five-point scale (5=strong, 4=maturing, 3=developing, 2=emerging, 1=beginning). The students at each grade level are scored on the six-trait analytic model:

- **ideas and content**--the writing should be focused and clear, have a controlling idea, and have enough detail. The student should be selective, show insight, and write from experience. There should be evidence of integrity (wholeness) in the writing.
- **organization**--the student should open with a real lead, and have effective sequencing and good pacing as well as smooth transitions. The writing should build to a high point and end with a sense of resolution.
- **voice**--the writing should give the sense of the person behind the words and facilitate a writer-reader interaction. There should be evidence of audience awareness, commitment, involvement, and conviction. The text should be lively, personal, individual, and expressive.
- **word choice**- the student should show evidence of a strong vocabulary. Writing should be natural, with energetic verbs and precise nouns and modifiers.
- **sentence fluency**--writing should have a rhythmic sound, natural, and easy-on-the-ear. The phrasing should be poetic or musical, making it easy to read aloud. Sentences should be powerful, clear, and graceful, with a variety in length and structure.
- **conventions**--there should be appropriate spelling, punctuation, grammar, capitalization. There should be consistency.

As shown in the graphic, students at all grade levels are scoring slightly above the "developing" level, but below the standard of excellence.



Are There Limitations as to How the Information Obtained From State Assessment Tests Should Be Used, And How Is It Being Used?

The type of assessment tests the Board of Education has chosen—criterion-referenced tests tied to Kansas curriculum standards—prevents any comparisons of test scores with students in other states. Because the tests have had to be developed from scratch, changes that have had to be made during the pilot phase of development also have temporarily limited year-to-year comparisons. The assessment tests were designed so that, by themselves, the test scores would be reliable and valid when used to measure students' average performance at individual schools. Officials from the Department of Education and the University of Kansas' Center for Educational Testing and Evaluation said the assessment test results shouldn't be used as the sole factor in comparing individual student performance or in making comparisons between schools or districts, because of the many other factors that can affect such comparisons.

School administrators generally indicated they are using State assessment test scores as designed—to see whether their students are being taught what they're expected to know, and to identify changes that might be needed in what or how they teach. However, because of confusion about how test results should be used, they sometimes are being used for purposes they weren't designed for, including comparing individual students' achievement and comparing schools or school districts. Recent actions by the Legislature will either require or encourage the assessment test results to be used for such purposes. School officials and others had both positive and negative opinions about the value of the assessment tests. Specific concerns were that the tests don't provide any new information, scoring of performance components is inconsistent, and the tests take too long to administer. Teachers also expressed both positive and negative opinions about the performance components. They generally liked the performance component of the writing assessment test, but disliked that aspect of the social studies and mathematics tests. Finally, members of the public had some additional concerns about attitude questions that were being asked on the assessment tests; we found those types of questions no longer are included.

In a Broad Sense, the Type of Assessment Tests Selected and The Way They Were Developed Imposes Some Overall Limits On the Use of Assessment Test Results

During this audit, legislative questions were raised about why the Board of Education didn't use an "off-the-shelf" standardized test, such as the Iowa Test of Basic Skills, or the National Assessment of Educational Progress (described in the profile on the next page), that could be used to compare Kansas students with students nationwide. Other concerns centered on the changes made to the tests that limited year-to-year comparisons. The following sections of the audit discuss these limitations and why they exist.

Because the assessment tests specifically measure students' progress toward meeting the Kansas curriculum standards, no national comparisons can be made. Some standardized tests, such as the Iowa Test of Basic Skills, are developed

National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) is a Congressionally-funded project of the National Center for Education Statistics of the U.S. Department of Education. NAEP has collected and reported information for nearly 25 years on what American students know and what they can do. It is the nation's only ongoing, comparable, and representative assessment of student achievement.

Since 1988, the NAEP assessment tests have been given to a scientifically selected sample of students attending both private and public schools. The assessment tests are administered to entire classrooms of students in grades four, eight, and 12 every other year in even-numbered years. Current legislation requires assessment tests in reading and mathematics at least every two years, in science and writing at least every four years, and in history or geography and other subjects selected by the National Assessment Governing Board at least every six years.

In recent years, NAEP has developed assessment procedures that rely more heavily on open-ended questions. These questions examine how readily students can think for themselves, reflect on and evaluate their understanding of the subject, explain what they know, and apply their knowledge in different situations. Another recent change in NAEP is the gathering of information to report to individual states—state report cards, if you will—that allow state-to-nation and state-to-state comparisons.

In 1994, 39 states participated in the NAEP assessment tests. Kansas was one of 11 states that didn't participate. According to officials at the State Board of Education, Kansas doesn't participate in NAEP because:

- the mathematics component of the assessment test aren't based on the *Curriculum and Evaluation Standards for School Mathematics*
- there are differences in the testing philosophy between NAEP and the State Board of Education
- it would cost an estimated \$60,000 to test a sample of 2,000 of the State's 430,000 students
- there's a time lapse between the administration of the assessment tests and the reporting of the test results

to measure students' basic knowledge in a wide range of academic subjects. Because these tests don't specifically measure any particular state's curriculum but, rather, a more "generic" set of basic knowledge, the results can be used to compare students nationwide.

Department officials told us they were interested in an assessment test that specifically measured Kansas' curriculum standards—no off-the-shelf test could do that. Because the assessment tests would have to be developed from scratch, the Department also determined that contracting with the University of Kansas, instead of using a private test development company, would be the most cost-effective alternative.

The Department also chose to develop a "criterion-referenced" test instead of a "norm-referenced" test. In a criterion-referenced test, students are measured against specific criteria; in this case, the Kansas curriculum standards. A student's score is based on how well he or she did on the assessment test. For instance, a student may have gotten 85% of the answers correct.

In a norm-referenced test, such as the Iowa Test of Basic Skills, a student is measured against other students who have taken the test. The student's score doesn't tell how many correct answers he or she got, but is expressed as a "percentile" that gives the student's relative ranking. For instance, a student who scores in the 90th percentile has scored better than 90% of those other students who took the test. This doesn't mean the student got 90% of the answers correct. In fact, a student could do poorly on a norm-referenced test, but still score high if all the other students did even

worse. Likewise, a student could do very well on a norm-referenced test, but still score low if all the norm group of students did well.

School districts in Kansas haven't given up on the idea of comparing their students with other students nationwide, however. As the accompanying box shows, nearly all districts continue to give their students a standardized, norm-referenced test, such as the Iowa Test of Basic Skills. Appendix A compares the Kansas assessment tests with the Iowa Test of Basic Skills and the National Assessment of Educational Progress.

School Districts Continue to Give Standardized Tests

During this audit, we reviewed survey information gathered by the Center for Educational Testing and Evaluation at the University of Kansas that listed the standardized, norm-referenced tests each school district offered in the 1994-95 school year. The survey results showed that only five school districts didn't give a standardized test that year. We contacted a sample of 63 of Kansas' 304 public school districts, including all five that had indicated they didn't give standardized tests, to verify the information on the survey and to find out if the school districts had changed the tests they gave during the 1995-96 school year.

Officials in only three of the districts we contacted said they didn't give a standardized test during the 1995-96 school year. In some cases, officials said they had reduced the frequency of testing; for instance, by not testing those grades that would be tested by the State assessment tests. However, it appears that the vast majority of school districts in Kansas continue to give both standardized tests and the State assessment tests.

Because each assessment test goes through a pilot development stage, results from early versions of the test can't be compared with results from later versions. Professional standards call for assessment tests to go through a pilot development period of one or more years. In Kansas, by design, the mathematics assessment test went through three years of pilot tests, the first in 1991. Multiple forms of different mathematics assessment tests were developed and given randomly to students across the State.

Officials from the University of Kansas' Center for Educational Testing and Evaluation said they wanted educators to understand the types of questions that could be asked, and the types of skills and abilities their students would have to learn or be taught. After the initial pilot assessment, the results were analyzed and teacher comments were reviewed so that changes could be made as necessary. Similar pilot assessments were given in reading and writing, and similar assessment development steps were taken.

Center officials told us that the assessment test results from the 1994-95 school year established a baseline of school information for the mathematics and reading assessments, and the results from the 1995-96 school year established the baseline for the writing assessments. According to Center officials, these three tests will remain relatively unchanged for the next few years so that year-to-year comparisons can be made. The Department hopes the tests will remain unchanged for at least the next four years, although no firm decision has been made yet.

The other assessment tests aren't as far along in the development process. The science assessment was first given as a pilot test during the 1993-94 school year. After revisions were made, it was given a second time during the 1994-95 school

year. The results from the 1996-97 school year will provide the baseline for this assessment test. The social studies assessment was given as a pilot test during the 1994-95 school year, and will not be given again until the 1996-97 school year. Because the social studies assessment test is still being finalized, the Department doesn't know what year will be the baseline for this subject area.

**The Assessment Tests Were Designed So That,
By Themselves, the Test Scores Would be Reliable and Valid
When Used To Measure Students' Average Performance
At Individual Schools**

In developing educational tests, the purpose of the test and the way in which the results are to be used are critical factors in how those tests are designed and structured. These factors can influence the number and types of questions asked. They also influence the types of evaluations needed to ensure that the tests are reliable and valid for the purpose for which they were designed.

According to officials from the Department and the Center, the purpose of Kansas' assessment tests was to measure whether students were being taught what they were expected to know, and the extent to which they could apply the knowledge and skills they'd learned in school. It also was intended to provide feedback to schools about whether their students were being taught what they were expected to know, based on the Kansas curriculum standards, and whether their students were improving in these areas over time.

Given this purpose, the assessment tests were designed so that an average test score would be developed for each grade tested in each school. Each school's average scores in one year then could be compared to its scores in later years to see if improvement had taken place.

Results were designed to be used at the school level because that's where any changes would have to take place to try to improve students' scores. For example, school administrators and teachers may decide to change the teaching materials being used (such as textbooks) or they may decide to change their teaching methods (for example, involving students in more group projects, asking more open-ended questions on tests, emphasizing "estimation" skills, and the like.).

At the start of this audit, concerns were raised by legislators and others about the reliability and validity of Kansas' assessment tests as measures of students' average performance in a school. Our review showed that the Center for Educational Testing and Evaluation followed the steps required by professional standards and good practices to ensure that Kansas' assessment tests were valid and reliable for the purposes intended. In addition, several on-going monitoring efforts are planned to ensure that these assessment tests continue to be valid and reliable. More information about this area is presented in Appendix B.

It is important to remember that Kansas' assessment tests were designed to be reliable and valid when student scores are averaged, by grade tested and by subject area, at the school level, and are compared for that individual school over time. The results also can be used to compare individual school districts or the State over time. However, problems arise if the tests are used for other purposes.

Department and Center officials said the assessment test results shouldn't be used as the sole factor in comparing individual student performance or in making comparisons between schools or districts, because of the many other factors that can affect such comparisons. According to Department and Center officials, Kansas' assessment tests weren't designed to measure or compare individual students' performance. If test results are used that way, the scores are not very accurate because the "margin of error" for any individual student's score is too large. This issue is explained in the chart at right.

As the chart shows, when students' scores are averaged together, which was how the results were designed to be used, the error rate goes way down. Department and Center officials told us that, if someone wanted to view an individual student's score, that score should be used as just one indicator of a student's performance, in conjunction with such other indicators as grades, teacher tests and observations, and standardized test scores.

In addition, Department and Center officials told us that trying to compare assessment test scores between schools or school districts could result in comparisons that are invalid. A school's average score on an assessment test doesn't provide any details about why the students at that school scored the way they did. A number of other factors could be causing differences in scores, including the following:

- schools' attendance, drop-out rates, or graduation rates (which may relate to students' motivation to perform well on the tests)
- students' socio-economic status
- when the tests are given at each school in relation to what's been taught

These types of factors tend to be relatively constant over time within a school, or even within a district. However, if comparisons are made between schools or districts, these and other pieces of information are needed to understand why one school's (or district's) average score was higher or lower than another's. Comments from two principals help explain the problems with comparing schools. One said:

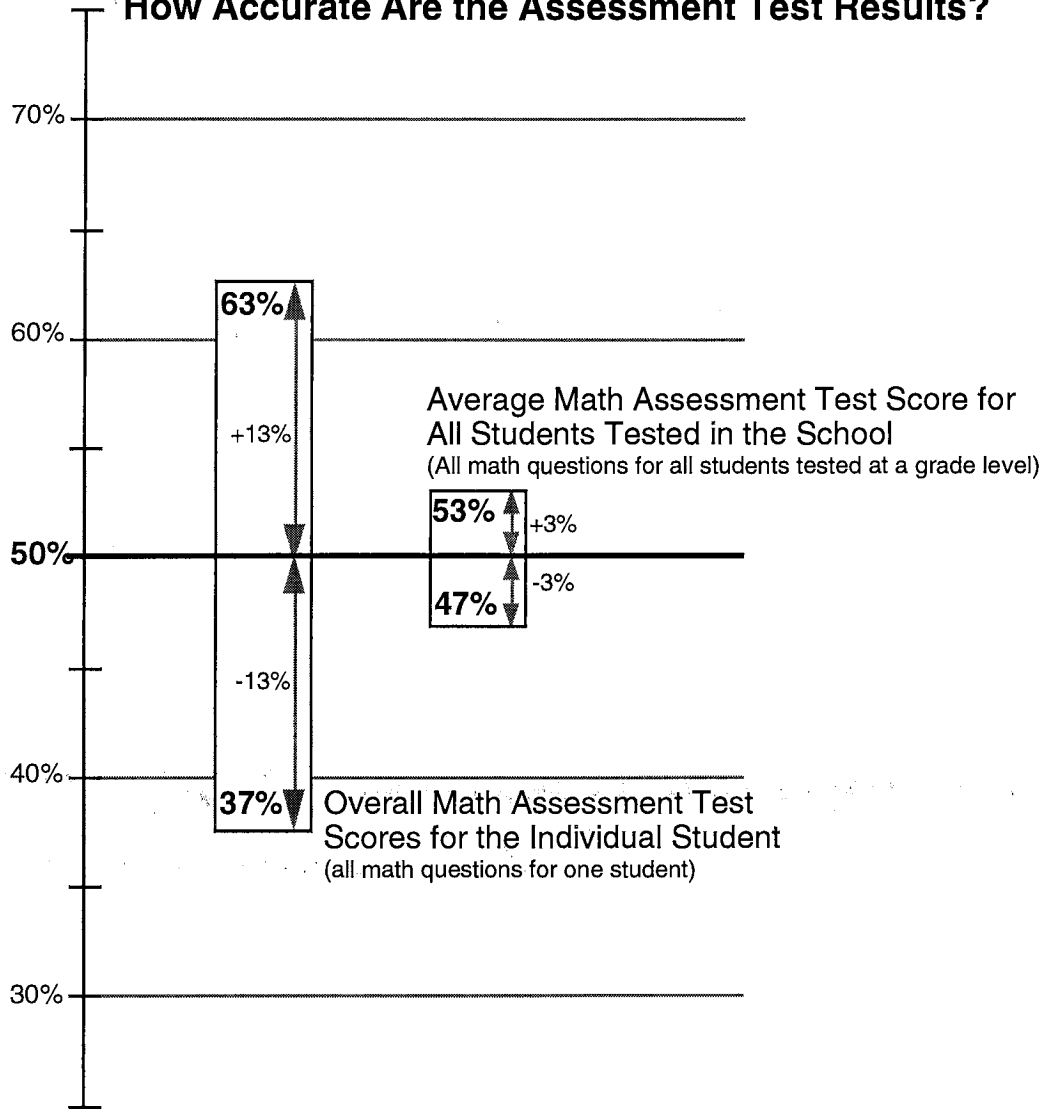
It is unfair to use the results of the tests to compare districts, because each district decides when certain information is introduced to students at a particular grade. If a district does poorly on part of the test, it may not reflect poor teacher performance or student performance, it may just be that information has not yet been taught.

Another said:

To use this test to evaluate teachers or schools is very unfair. Some populations have a majority of students that do not speak English or speak it well. Some populations are very poor. Even with coding, these students appear to do poorly on the test when compared with other districts.

Department and Center officials told us if parents, educators, policymakers, or others wanted to use assessment results for these other purposes, the Board would have to design an entirely different test, and would have to evaluate the reliability and validity of the test for each purpose.

How Accurate Are the Assessment Test Results?



Based on the 1995 assessment test results, the University of Kansas' Center for Educational Testing and Evaluation calculated the "measurement error" for two types of scores from the math assessment test using a 95% confidence band (which means the score will be within the range 95% of the time).

As the chart shows, the smaller the number of math questions included in the calculation of a score, the greater the measurement error. Thus, if a student got an overall score of 50% correct on the math assessment test, because of measurement error that student's score could be as high as 63% or as low as 37%. Described another way, that is an error rate of plus or minus 13%. That's not very accurate.

The average score for all students tested in a school at a grade level is based on the largest number of questions and students, making it the most reliable measure of student achievement in math.

In General, School Administrators Indicated They Are Using State Assessment Test Scores as Designed

We interviewed officials from the Department of Education, and surveyed a sample of public and private school superintendents, principals, and teachers from across the State to find out how they were using the assessment test results. We found that many school officials are using the assessment results to see whether their students are being taught what they're expected to learn, to identify changes that may be needed in what or how they teach, and to try to improve those test results over time. For instance, 37 of the 66 superintendents we surveyed (56%) said they had used the assessment test results to change their teaching materials or coursework, such as adding specific subject areas in estimation or science experimentation. A total of 46 superintendents (or 70%), said they had used the results to change teaching methods, such as focusing more efforts on cooperative learning or open-ended problems.

Nearly all the superintendents said they also used the results as part of the quality performance accreditation process.

A total of 181 principals (54%) said they were using the assessment test results to change how they teach, and 253 principals (or 75%), said they were using the results to change what they teach. In separate surveys conducted by the Center, more than half the teachers indicated they had, or planned to, make changes in how and what they teach as a result of the assessment tests.

Even though most survey respondents said they were using the results for their intended purposes, some reported they weren't using the assessment test results at all. About one-fourth of the superintendents and principals reported they weren't using the results to change what their districts taught. In addition, about 15% of the superintendents said they didn't use the assessment test results to make instructional changes, and nearly half the principals said the same thing.

Because of Confusion About How Test Results Should Be Used, Assessment Test Results Sometimes Are Being Used For Purposes They Weren't Designed For

Professional standards require test developers to provide easily understood information about how test results can be interpreted. In our surveys of superintendents, principals, and teachers, we asked what information had been sent to them on how to interpret the assessment results. Most of the communication from the Department goes to the superintendents and principals, or to a central testing coordinator or administrator. Nearly all the superintendents said they had received manuals and other materials from the Department and had attended or sent staff to assessment test workshops.

We reviewed the information the Department or the Center sent to these individuals and found it did talk specifically about how the test results should and should not be used. Three-fourths of the superintendents who responded to the question told us the information the Department or Center sent them was useful or usable, although about half of these said they still had some questions.

However, we also found that the assessment test results sometimes were being used for purposes they weren't designed for, including comparing students, buildings, and districts. There also is some evidence that the results are being used for, or being considered for use as, graduation requirements or to evaluate teachers. These areas are discussed below.

- **A significant number of school administrators and teachers are using the assessment test results to compare individual students' achievements.** A total of 159 principals (47% of those who responded to our survey), said they use the results to compare individual student achievement. Also, 94 of the 204 teachers who responded to the question (46%), said they had been told they could use the assessment test results to evaluate individual students.

As noted earlier, the problem with using assessment test results to compare individual students' performance is the large measurement error of plus-or-minus 13%. Thus, if one student's score of 50% is compared to another student's score of 60%, it's not possible to say with certainty which student actually knew more.

- **In all, 11 of the 66 superintendents we surveyed (17%), and 84 principals (25% of the respondents), also said they were using the assessment test results to make comparisons between schools or districts.** Again, the problem with using the assessment test scores to make these types of comparisons is that they don't include other relevant information—socio-economic indicators, information about what is taught, attendance and drop-out rates, and the like—that would help explain why the differences in test scores existed. Without this additional information, comparisons between schools and districts are not valid.

- **A relatively small number of superintendents and principals said they either had or were considering tying students' graduation requirements to their assessment test scores, or were using the results to evaluate teacher performance.** One superintendent said his district was considering using the results as a graduation requirement. Three of the 66 superintendents (5%), said they used the assessment test results to measure teacher performance. A total of six principals (2%) said they used the assessment tests as graduation requirements, although from the limited comments they provided, we don't know how much weight is given to the test scores in determining whether a student graduates. Six principals also said they used them for evaluating teachers.

The assessment tests were not designed to be used for these purposes. They may be used as one indicator of how well a teacher is doing, but a teacher's performance is affected by a lot of factors outside his or her control. For example, a teacher may have an exceptionally strong class one year, and a weak class the next. Even though the teacher may teach the students the same things in the same way each year, the class scores may be very different. In addition, the teacher has little control over what the students were taught by previous teachers, even though that can influence how well students perform on the assessment tests.

According to the professional literature and Center officials, using the assessment test results as a basis for graduation imposes a whole new level of reliability

and validity requirements, as well as legal requirements. Kansas' assessment tests were never designed to be used for this purpose, and likely would be challenged in court if a district tried to do so.

In three other states we contacted, assessment test results were used to report individual student achievement or serve as a graduation requirement when they weren't designed for that purpose. In Arizona and California, the assessment tests ultimately were dropped completely because the results were being used for inappropriate purposes. In both states, new assessment tests are being designed that will allow for these uses of the results. In Illinois, test developers addressed the problem of people using test results to report individual students' performance by doubling the number of test questions to improve the reliability of individual student scores.

Recent actions by the Legislature will either require or encourage assessment test results to be used for purposes they weren't designed for. Beginning in January 1997, the Legislature has mandated that the Board prepare a building report card for each school that will show, among other things, the school's assessment test results over time, the average scores for the district in which the school is located, and the average scores for all schools Statewide. Some demographic information and local indicators of student progress also will be included as part of the report card.

School officials expressed a number of concerns about the building report card on our surveys. One principal commented:

I also feel that the Kansas assessment was a tool to help assess if we're meeting State standards, but with the proposed report card, I see it as the only tool legislators and patrons will use to make judgments about our schools.

Another said,

The reason that I am dissatisfied simply related to the development and use of the results. I think tests such as these should not be used to compare classes, buildings, or districts. From the beginning, there were assurances that this was not going to happen...yet (the new building report cards) seem to encourage this very thing.

Because it's likely that these report cards will be used to compare assessment results between schools and districts, there's a significant risk that these comparisons will be invalid. To minimize this likelihood, it will be critical that additional information about school and district make-ups be provided as part of the building report cards.

The 1995 Legislature also mandated that the Board develop student performance expectations, which essentially would be the scores students would be expected to achieve on the assessment tests. Department officials told us they don't know what they'll be doing in this area yet, although they have begun meeting to discuss individual performance standards. However, the existing assessment tests weren't designed to measure individual student performance.

Center officials told us this requirement may mean that the assessment tests would have to be redesigned so that individual student scores would be more reliable. As examples of possible changes, they said the tests may have to include more questions, or may have to rely more heavily on multiple-choice questions instead of the open-ended or performance-based components.

School Officials and Others We Talked With or Surveyed During This Audit Expressed Both Positive and Negative Opinions About the Value of the Assessment Tests

Superintendents were most positive about the assessment tests: 62% of those who responded to the question said they liked the State assessment tests overall. Principals were less positive, with 56% of the respondents saying they were satisfied overall. Teachers were less positive about the assessment tests. More than half of those who responded to the question (60%), said they were dissatisfied with them overall. A sample of comments is shown below.

Some Teachers and Principals Find the State Assessment Tests To Be Useful

The following are examples of comments Kansas teachers and principals made on our survey that show some schools find the State assessment tests useful in helping them make improvements.

- I think the tests are valuable for teachers and students to see where improvement could be made.
- The test can help me to improve areas of weakness in my math program. I feel that the test can be used to improve the instruction that the class receives. This will better prepare them for the future.
- The State assessment tests results have brought more data to assist schools in their school improvement plans. While school improvement has been a goal for all of us in education, the tests' results are another tool in helping schools reach their educational goals.
- These tests . . . give us a better picture of our students and of our curriculum. These assessments have also reshaped some of our instructional methods to a more hands-on, research-based mode which is more realistic for our students' needs.

Some Teachers and Principals Have Concerns About the Assessment Tests

The following are examples of negative comments we received about the State assessment tests in response to our survey of teachers and principals across the State.

- The State assessments were more of a frustrating experience than learning. The instructions were vague...Most students had difficulty because they just didn't know what was expected.
- When these results get into the press and district scores are compared with other area school districts, the headaches will begin. No one will want to be compared unfavorably and I see everyone taking one of two options: (1) Don't grade the tests hard or (2) Teach to the test...I personally feel we are wasting tax money on these assessments.
- Variations in a particular class group's ability can make for significant variations. When these scores are used to represent a school's performance, I feel it is a misrepresentation.
- When adding all of the State assessments to our regular achievement tests, it's too much pressure on our students. I'm not sure it's a true measure of our students' abilities.

Those who liked the assessment tests were most positive about how the results could improve education in Kansas. Although respondents generally tended to cite problems with the assessment tests, about 8% of them did give some positive comments. These respondents said that the assessment tests were an excellent way to help students learn. They also said that the assessment tests were better measures of what schools are teaching than standardized, norm-referenced tests, and that the tests helped to identify areas where changes needed to be made in what and how they teach their students.

About half the teachers indicated the assessment tests provided them with feedback on changing how they teach and even more said the assessment tests gave them feedback on changing what they teach. In addition, six teachers specifically said that the assessment tests had helped them re-evaluate their teaching.

Those who didn't like the assessment tests generally voiced concerns about the usefulness of the information they provided, the consistency of scores, and the amount of time it took to administer them. Some of the specific concerns about the usefulness of the testing information were that the tests don't provide any new information beyond what is already available from teacher-made tests, standardized tests, or student grades, and that the assessment tests aren't good measures of what is taught in class or of the curriculum standards.

Concerns also were expressed that local teachers might score their students' tests higher than they should be scored. One principal said:

I have a problem with the scoring of the open-ended questions on the reading and also the teacher scoring of the writing assessment. Whenever you leave the scoring of a test to human judgment, you have a discrepancy in the scoring. Some teachers score high and some score low. Even with the third reader, the tests could be scored too high or too low in some area schools. If this happens, it makes a marked difference in the results that are published for the public to see and question!

A teacher commented:

Since each school grades its own tests, what's to guarantee that some graders aren't just putting down high scores rather than honestly evaluating according to the (rules) — no standardization is possible.

Center and Department officials said they encourage local teachers to score the assessment tests because it is an invaluable professional development activity. Officials we spoke with from the company that develops the Iowa Test of Basic Skills said they also encourage local districts to score the performance components of assessment tests. However, in the other states we contacted, the scoring was done centrally. As noted in Appendix B, the Center is doing studies to determine whether a problem with inflated local scores exists.

Complaints about time centered around the class time required for students to take the test, and also the time it takes local teachers to score the tests. In general, the open-ended and performance components of the assessment tests are scored by a local teacher. The writing assessment is scored by two local teachers and, if there is a

large difference in scores, a third teacher also scores the test. In addition, as a quality control check, 10% of the assessment tests are sent to the Center for central scoring.

Teachers had both positive and negative opinions about the performance components of the assessment tests. Each of the assessment tests includes a performance component. For instance, the science assessment tests include a group science project at the 5th-grade level and an individual science project in grade eight. The Department has indicated that the performance components are "state-of-the-art" testing, and that they are the best way to measure the higher-order thinking skills that have never been measured before. Other states we contacted also generally had some performance pieces to their assessment tests, although many states limited their performance activities to open-ended questions, instead of the group or individual projects used in Kansas.

However, testing literature has identified potential problems with performance tests. For example, the tests are more subjective than multiple-choice question tests, are difficult to standardize, and can be more subject to variations in scoring. If there's a group project, it can be difficult to measure individual students' contributions. Also, there are concerns that these types of tests are very expensive in terms of time, but don't yield that much more additional information than other types of tests. Finally, a performance component focuses on the process of doing something; for example, a science experiment. As a result, it's more limited in the subject matter it covers.

The Department has addressed some of these concerns by not reporting a separate "performance" score, but including the performance results in a student's overall score for a particular test. Nonetheless, concerns still exist about this aspect of the assessment tests. For instance, one teacher expressed some concerns about the group projects:

If anything, the State assessments should be used for individual work, not for group work like projects done in groups of three or four as in social studies and science! Not everyone in groups can or should be given the same group score. They don't all work, participate, or have the same cognitive skills.

We found the level of satisfaction relating to the performance component varied by subject. Teachers who gave the mathematics or social studies assessment tests were more likely to be dissatisfied with the performance components. On the other hand, writing teachers thought the performance component was very useful.

Some of the specific comments about the performance component, by subject area, are listed in the profile at the top of the next page. Complete survey results can be found in Appendices C, D, and E.

Members of the Public Also Expressed Praise and Concern About the State Assessment Tests

We talked to several members of the public who had given testimony before the House or Senate Education Committees in favor of or in opposition to Quality Performance Accreditation. We asked them for any comments they had about State

Teachers and Principals Expressed Both Positive and Negative Opinions About Particular State Assessment Tests

Teachers and principals expressed opinions on our survey about particular State assessment tests. Some comments raised concerns about the performance-based questions, particularly questions with open-ended responses or extended writing exercises. For example:

Science The performance parts of the science test (11th grade) depends a lot on the students' ability to write. Students that don't like to write may not give very good answers, even if they have the knowledge.

My sections of 8th graders were especially interested. They enjoyed designing their own experiments and testing them.

Social Studies The results provided do not compliment the instruction the students receive in these subject areas. For example, the social studies assessment requires English, grammar, and research skills as prerequisites. A student's understanding of American history...is ignored, distorted, or misunderstood in the students who lack the prerequisites.

Reading The reading assessment questions are vague considering the desired responses are specific regarding word choice.

I find the reading and writing assessments authentic...I think the reading assessment is so good that I will use it as a model for some of my assessments.

Writing The writing assessment is the best we've had for this area. The six-trait model works because kids seem to understand it. Writing has been improving! But, as a sole source, I feel it is unreliable because it is subjective—as writing will be.

Math Although our math curriculum goes along with the State assessment test in math, most children can't score with a high score because of all the reading involved with each problem on the test.

assessment tests. Some of the people we spoke with said they thought the State assessment tests were beneficial, necessary, and valuable. However, other people expressed two major concerns about them:

- the assessment tests were asking students questions that, in their opinion, were none of the State's business or were inappropriate attitude-type questions
- the assessment tests were using reading selections that contained inappropriate subject matter

We found that in the pilot assessment tests, questions about students' attitudes toward the subject matter were asked, such as 'how well do you like reading', or 'do you read for fun.' As described earlier, these types of questions are used to help establish the validity of the assessment tests. (Similar questions often are asked on standardized tests as well to help establish the validity of those tests.) These attitude questions no longer are asked.

Concerns about the inappropriate reading selections focused on the passage given to third graders in 1993. In particular, people we spoke with said they thought

the passage dealt with the issues of suicide and death. Center and Department officials told us that particular reading passage is no longer being used. These officials also noted that there often is controversy over the reading passages, but that they try to minimize any problems by their selection process, as described below:

- librarians nominate reading selections for each grade
- teachers review these nominated selections
- reading advisory committee reviews these selections
- the Board makes the final selection

Conclusion

Kansas assessment tests were designed to be used for comparisons over time of an individual school, individual district, or the State as a whole, to see if students were mastering the things they are supposed to learn under the State's curriculum standards. Average scores were not intended to be used to make comparisons between students and between school districts. Making these types of comparisons requires significantly more information than can be gleaned from the test scores alone.

Responses to our surveys show that most educators are using the State assessment tests for appropriate purposes. A few indicated they were using them for things they weren't designed for. Generally, superintendents seemed to have a better understanding of acceptable uses than did principals and teachers. This may indicate that the districts need to improve communication to make sure information reaches all levels. A legislative requirement to produce building report cards could increase the risk that test scores will be used inappropriately. The Department will need to take appropriate precautions when it issues building report cards to help ensure that test scores are not misinterpreted or misused. Another legislative requirement to set expectations for student performance could require that the tests be redesigned to make them better measures of individual student performance.

The performance components of the assessment tests (those parts that require students to write open-ended responses or complete group or individual projects) generated a lot of positive and negative comments. Much concern was expressed about the time involved and the cost of administering and grading these parts of the tests. Clearly, these types of questions represent a new way of testing whether students are mastering the higher-order critical thinking skills required by the curriculum standards. After they have been in use an appropriate amount of time, the Department may want to consider whether their use is the most cost-effective way to obtain the desired information.

Concerns also were raised about whether local scoring of the tests would allow districts to inflate their scores to show improvement. The Department has built in some safeguards to detect inflated scores, and will need to continue monitoring this so that interested parties can have assurance that the test scores are reliable.

Recommendations

1. To ensure that all educators are informed about the appropriate and inappropriate uses of the State assessment tests, the Department of Education should continue to work with school district superintendents and local boards of education to ensure that this information is communicated to all educators in each school district.
2. To lessen the likelihood that test score information contained in building report cards it issues for school districts will be misinterpreted, the Department should do the following:
 - a. include clear cautionary language on the report card that will alert the reader to limitations in the use of test scores.
 - b. advise school districts about what additional information they should have assembled to provide to members of the press or other interested parties to help to explain differences between individual schools and districts.
3. Assessment test scores are only reliable within a wide range for comparing individual student scores to the "Student Performance Expectations" the Legislature has required the Department to establish. As a result, the Department should assess what changes would need to be made to those assessment tests to make such comparisons more reliable and valid, and what the associated cost would be. The Department should present that information to the education committees of the Legislature by the start of the 1997 session; so that the Legislature can make an informed decision about whether it wants to retain the student performance expectations, and what it wants them to be used for.
4. To ensure that the results of the assessment tests are as reliable and valid as possible, the Department of Education should continue to monitor the consistency of local scoring of exams, and evaluate whether changes in scores reflect changes in what is being taught and how it is being taught.
5. To ensure that the State assessments are done in a cost-effective manner, the Department should continue to monitor the performance components of the tests to ensure that the information they provide is worth the cost, and to determine whether there are other less costly means of obtaining the desired information about how well schools are teaching Kansas students.

How Do Kansas' State Assessment Tests—and Its Methods of Measuring Student Performance—Compare With Other States?

Because of federal initiatives, other states also are developing assessment tests. Although Kansas is further along than about half the 13 states we contacted, at least five other states have assessment test programs in place. Like Kansas, these states give students at the elementary, junior high, and high school levels criterion-referenced assessment tests designed to measure over time how well students in a particular school are learning the things called for in each states' curriculum standards, goals, or expectations. The tests in these other states include a variety of types of questions, although the emphasis is on open-ended and performance-based questions. A difference between Kansas and the other states was that two of these five require other standardized tests to be given. In the other three states, as in Kansas, giving standardized tests is a local school district option.

Kansas is Further Along Than More Than Half the States We Contacted in the Development of its Assessment Tests

As discussed in the Overview, in January 1994, Congress enacted the Goals 2000: Educate America Act. To receive federal funding under the Act, states have to develop a set of high-quality student assessment tests that test each state's "challenging" standards. A related federal law, Title I of the Improving America's Schools Act, provides funds for "at-risk" students. To be eligible for these funds, schools also have to develop state assessments at three grade levels.

We contacted 13 states including the surrounding states and Iowa, to find out about their assessment programs. Two of the states we contacted (Arizona and California) reportedly had assessment tests similar to Kansas and stopped using them. Both of these states apparently dropped their assessment tests because they were being used for inappropriate purposes. Information about Arizona and California is contained in the box below.

California and Arizona Had Assessment Tests, But Dropped Them

These states dropped their assessment testing because the results were being used for inappropriate purposes. The tests in Arizona originally had been designed to measure students' knowledge of essential skills, similar to the Kansas curriculum standards. They actually were being used as a graduation requirement, but weren't reliable for that use. The California assessment tests had been designed to measure school improvement, but were being

used to measure individual student achievement. Concerns were raised that the results weren't reliable for this use. Because of the public outcry that arose when the results were used at an individual student level, the assessment tests were dropped.

Both states are in the process of developing new assessment tests to measure students' achievement of their states' curriculum standards.

Six states we contacted either hadn't finalized their assessment tests or didn't plan to develop them. These couldn't be compared to Kansas' testing program. Information we were able to gather about those states can be found in the box below.

Several States Are Developing Assessment Tests, and One Has Decided Not To Require Them

We contacted officials in several states to find out whether they had state assessment programs. The following states did not have assessment tests in place yet, although all but Iowa were in the process of developing them.

Wisconsin, Virginia, Colorado, and Missouri are in the process of developing assessment tests.

The Virginia assessment tests will measure the state's new curriculum standards. In addition, the state will set performance expectations for individual student achievement, and the assessment tests will be used as a graduation requirement.

The Missouri assessment tests, which are in the pilot stages of development, will measure students' mastery of the state's newly legislated academic performance standards. These tests will be performance-based. Officials said the focus of these tests will be on building improvement. Until these assessment tests are completed, Missouri will continue to use its current achievement tests, which are based on core competencies and key skills, rather than the new state academic performance standards.

Colorado is scheduled to pilot test its new assessment tests in the 1996-97 school year. These assessment tests will be based on the state's newly developed content standards. Officials told us they would be contracting with a private test developer during the summer of 1996 to have assessment tests developed in reading, writing, mathematics, science, history, and geography.

Wisconsin also is developing assessment tests, although the Wisconsin program currently is on hold because of funding problems.

Nebraska is in the process of developing state curriculum standards, but the assessment tests will be developed by local districts instead of at the state level.

Nebraska school districts are required to give some norm-referenced test at three grade levels. In addition, local school districts give a criterion-referenced test in reading, writing, and mathematics beginning in the 5th grade to measure students' progress toward the state content standards. The state is in the process of developing state curriculum (performance) standards, and local districts will be responsible for making sure their local criterion-referenced assessment tests measure these new standards.

Iowa doesn't require state assessments.

Local school districts can determine what educational outcomes are important, and select appropriate means of measuring those outcomes. Iowa officials told us they were trying to get a waiver from the federal government so they could receive funding without having state assessment tests.

Five states—Illinois, Kentucky, Maine, Maryland, and Oklahoma—are using assessment tests, which generally are similar to those being used in Kansas. The table on the next two pages shows the numerous similarities in the tests administered in these states.

As the table shows, the main similarities were in the type of test given, the level at which the results were intended to be used, and the grade levels at which the tests were administered.

All five states used criterion-referenced tests which, as explained earlier, are designed to measure a specific criteria or yardstick, such as state curriculum standards or goals. As in Kansas, these states use a variety of formats to test their students: multiple-choice or open-ended questions, performance activities (such as group or individual projects), and portfolios (which are an on-going record of a student's work). Clearly, however, the emphasis is on open-ended and performance-based questions. Even in the two states (Illinois and Oklahoma) that otherwise use only multiple-choice questions, the writing assessment is open-ended.

The major focus in each of these states is on academic improvement at the individual school level, although a couple of the states expanded that focus to the district level as well. Finally, each of these states generally administered the tests annually at the elementary, middle, and high school levels.

The main differences we noted were in how the assessment tests are scored, and whether other standardized tests are required to be administered. Scoring the assessment tests generally is being done centrally in other states. In Kansas, local teachers score the open-ended and performance parts of the assessment tests. As noted in question one, this practice is encouraged as a good staff development activity. However, concerns also were expressed that scoring could be inconsistent because of this practice. In the other five states, scoring generally is done centrally, either by state scorers or by the contractor. One exception is Maine, which uses teachers to score the writing assessment test only.

Finally, two of the states—Maryland and Oklahoma—also require other standardized tests to be administered by local school districts. Illinois doesn't require any standardized tests, although the state tests a sample of students on a periodic basis to allow national comparisons. In Maine, Illinois, and Kentucky, as in Kansas, the decision about whether to give additional standardized tests is a local school district decision.

COMPARISON OF STATE ASSESSMENT TESTS

State	KANSAS	Illinois	Kentucky
Developed by	consultant in conjunction with local teachers	consultant in conjunction with local teachers	consultant in conjunction with local teachers
Purpose	to generate improvements in what's being taught at the individual school level	school and district accountability/improvement	measure a school's improvement
Measures Achievement Towards	Kansas Curriculum Standards	State goals for learning; are developing State standards	State academic expectations
Types of Questions	multiple-choice, multiple-mark, open-ended, and performancebased questions	all multiple-choice questions except writing is all open-ended questions	open-ended and performance-based questions, and portfolios
Scored by	scored by a contractor except open-ended & performance-based questions scored locally	scored by a contractor	scored by a contractor
Frequency	annually	annually	annually
Administered in Grades	4,7, and 10 mathematics 3,7, and 10 reading 5,8, and 10 writing 5,8, 11 science & social studies	3,6,8, and 10 reading, math, and writing 4,7, and 11 social science and science	3,8, and 12 in all subject areas
Started in	1990-91	1991-92	1991-92
Cost to the State	About \$850,000 in 1996 (Doesn't include local scoring costs)	About \$6 million annually	About \$8 million for 1996
Who tested reliability and validity	contractor with oversight by the Department of Education	consultant and Department of Education	contractor with oversight by the Department of Education
Other Standardized Tests Required	None, local decision if one is administered	None, local decision; State periodically uses Stanford test for a sample of students	None - looking into normreferenced tests for math and reading

FOR KANSAS AND FIVE OTHER STATES

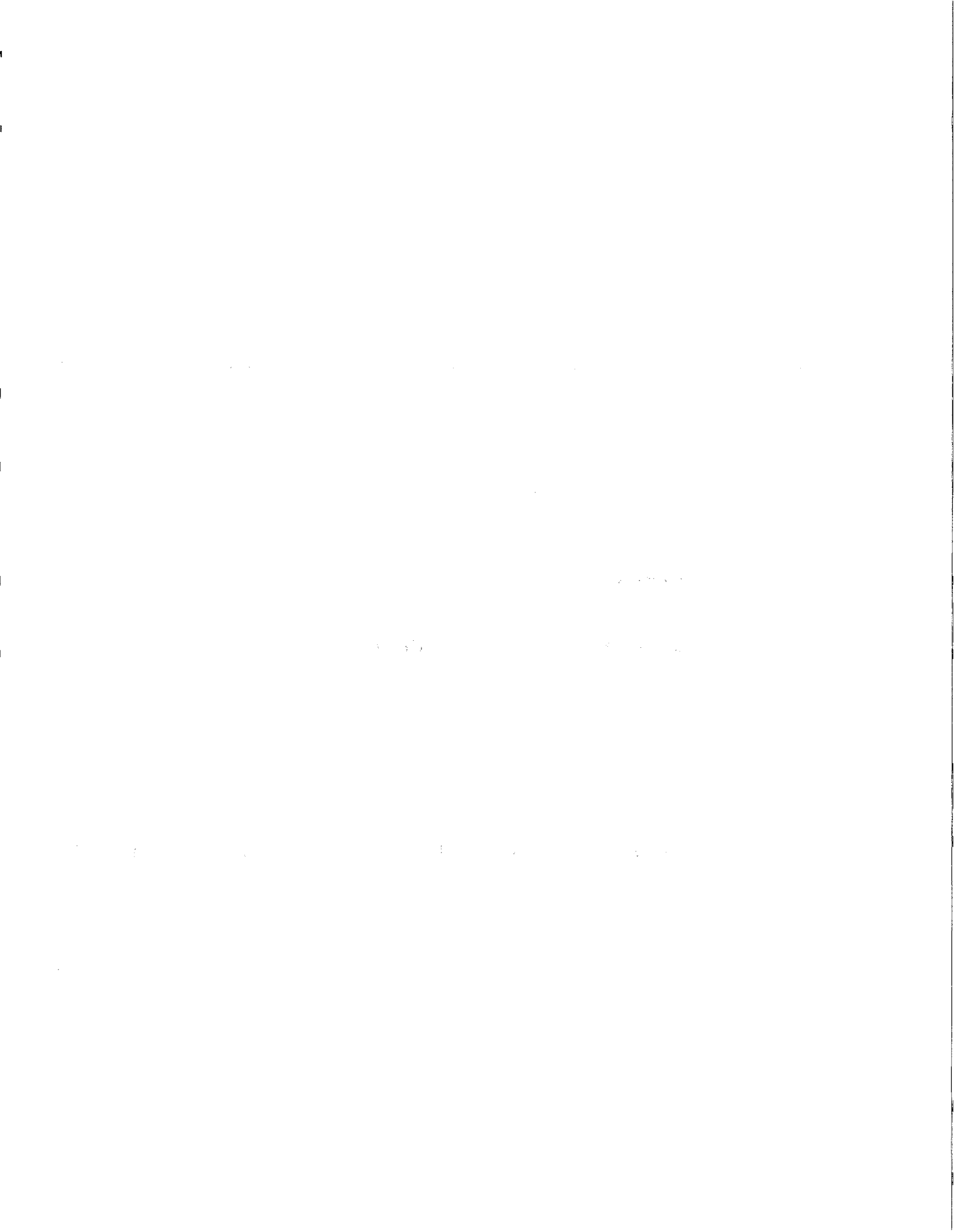
Maine	Maryland	Oklahoma
consultant in conjunction with local teachers	consultant in conjunction with local teachers	local teachers developed standards and consultant developed test questions
building/district program evaluation	school accountability	student mastery of core skills at building and district level
State content standards and performance indicators	State outcomes or learning goals	State specific levels of student performance on assessment tests
all open-ended questions	all open-ended questions	all multiple-choice questions, except writing is all open-ended questions
writing questions scored by State scorers, all others are scored by test contractor	scored by the State	scored by a contractor
annually	annually	annually
4, 8, and 11 in all subject areas	3, 5, and 8 in all subject areas—developing high school assessments	5, 8, and 11 in math, reading, writing, and science—US history and government in pilot phase
1985-86	1990-91	1994-95
About \$1.4 million in 1996	na	About \$2 million in 1996
contractor	Department of Education	contractor
None - local decision if one is administered	Comprehensive Test of Basic Skills-4 in grades 2, 4, 6, and 8; Minimum Competency in grades 7-12; Large districts have own Criterion-referenced tests	Iowa Tests of Basic Skills in grades 3-7



APPENDIX A

Comparing the State Assessment Tests With The Iowa Test of Basic Skills and The National Assessment of Educational Progress

The State assessment tests are criterion-referenced tests administered by the State Board of Education as part of the State's accreditation process known as Quality Performance Accreditation. Other types of well-known assessment tests administered to students include the Iowa Test of Basic Skills (ITBS) and the National Assessment of Educational Progress (NAEP) both of which are norm-referenced tests. This appendix shows the differences and similarities among the three tests in such areas as the purpose of the tests, what subjects are tested, how often the tests are administered, and the like.



Comparing the State Assessment Tests With the Iowa Test of Basic Skills and the National Assessment of Educational Progress

	Kansas State Assessment Tests	Iowa Test of Basic Skills	National Assessment of Educational Progress	Comments
Purpose	to measure students' progress towards meeting the Kansas Curriculum Standards	to measure a student's progress in the basic skills	to measure a student's achievement in specific subject areas	
Subjects Tested	1995-96: Mathematics Reading Science Social Studies 1994-95: Mathematics Reading Science Writing	Reading Language Vocabulary Mathematics Social Studies	1996: Mathematics Science 1994: Reading History Geography	NAEP is required by law to assess reading and math at least every 2 years, science and writing at least every 4 years, and history, geography and other subjects once every 6 years
Students Tested	Mathematics: Grades 4,7,10 Reading: Grades 3,7,10 Writing: Grades 5,8,10 Science: Grades 5,8,11 Social Studies: Grades 5,8,11	Tests are available for grades K-8 (a similar basic skills test for grades 9-12 --TAP--is available from the publisher). In 1994, about 37% of Kansas school districts administered the ITBS	only a sample of students are tested in grades 4-8, and 12 In 1994, 39 states administered the NAEP assessment tests. Kansas doesn't participate in NAEP	the sample size for Kansas for NAEP would be about 2,700 students of the State's 430,000 student population
How Often is Test Administered	Mathematics and reading tests are given annually; others are administered every other year	Varies by district; however, most districts administer the test annually	once every two years during the even-numbered years Kansas doesn't participate in NAEP	
Type of Test (a)	Criterion-referenced	Norm-referenced	Norm-referenced	
How Scores are Reported	The results are generally reported as percent correct. The performance-based questions are generally scored on a 5-point scale.	The results are reported as a "percentile" based on the average score of those students who took the test at a particular grade level in a particular subject (norm-referenced group) -- latest norm-reference (comparison) group for ITBS is 1995	The results are reported as a "percentile" based on the average score of those students who took the test at a particular grade level in a particular subject during the test year (norm-referenced group)	
Who Scores Tests	Multiple-choice questions are scored electronically by the contractor Performance-based questions are scored locally - considered staff development	All test questions may be scored by the publisher Suggest that performance-based questions be scored locally as part of staff development	All test questions are scored by the testing center	
How Tests Results Can Be Used	Building (school) improvement towards the Kansas Curriculum Standards over time Is not appropriate for student, building, or district comparisons	Compare student's achievement to a norming (comparison) group Can compare students, buildings, districts, and states	Compare national, regional, and state scores Is not appropriate for any comparisons at the building or student levels	NAEP does not generate results at the building or student levels

(a) Any test can give both criterion- and norm-referenced information. This table lists the most common way information is presented by each of these tests.



APPENDIX B

Evaluating Reliability and Validity

This appendix describes the work done by the University of Kansas' Center for Educational Testing and Evaluation to ensure that the State assessment tests were reliable and valid for measuring students' average performance in a school.

Steps Taken To Ensure the Reliability and Validity of State Assessment Tests

Our review showed that the Center for Educational Testing and Evaluation followed the steps required by professional standards and good practices to ensure that assessment tests were reliable and valid for measuring students' progress toward meeting the State curriculum standards. Concerns were raised by legislators and others about the reliability and validity of the assessment tests as measures of students average performance in a school, and about whether the Center received any input from any other groups in developing the assessments. Generally, we found that Center officials took the appropriate steps to ensure the reliability and validity of the assessments as measures of students average performance in a school. The following sections of the audit describe these findings in more detail.

Professional standards guide the development of educational tests. These standards require that tests be reliable and valid. Reliable means that in general, if a student took an assessment test a number of different times, he or she would score within the same general range each time he or she took the test. A test is considered valid if it measures what it is supposed to measure--in this case, whether schools have improved at teaching students what they are supposed to learn as established by the Kansas curriculum standards.

To determine whether the Center took the steps it should have, we reviewed the professional standards in this area, spoke with experts from other universities about what steps should be taken, and reviewed pertinent literature. The table below displays the general consensus of the experts about the kinds of things that should be done to evaluate the reliability and validity of a test. The left hand column identifies the primary steps that need to be taken to make sure a test can be used for its stated purpose. The right-hand column lists the steps the University of Kansas' Center for Educational Testing and Evaluation took to ensure the reliability and validity of the State assessment tests for measuring students' performance toward the *Kansas Curriculum Standards*. Our reviews showed that the Center did those things it was supposed to do.

To evaluate reliability:

What steps should be taken

- gather evidence to show that scores are consistent over time
- gather evidence to show that scores are consistent between raters

What the Center did

- performed statistical tests to ensure the student scores were consistent
- performed statistical tests to judge whether there were enough questions on each sub-category of an assessment test to determine at what level the scores were reliable
- performed statistical tests to ensure that the scores given by different raters would be approximately the same
- developed scoring rules and guidelines and provided training to raters on how to score the test

To evaluate validity:

What steps should be taken

- show the test questions cover the material in the standards
- review the questions for bias

What the Center did

- did a series of pilot tests for each assessment
- obtained input from educators in the field about whether the questions measured the curriculum standards
- reviewed alternative test question formats, such as multiple-choice or open-ended
- examined whether expected correlations existed between test results and students' attitudes toward the subject area (for example, it was expected that students who liked reading would do better on the reading assessment)
- had assessment test questions reviewed by many educators and members of different minority groups for bias
- performed statistical tests for bias

Several on-going monitoring efforts are planned to ensure the assessment tests continue to be reliable and valid. For example, the Center plans to examine whether there is any systematic inflation of scores for assessment tests that are scored by teachers in local school districts as compared with those scored by State scorers.

Center officials also told us they want to conduct external validation efforts to examine schools with improved scores to ensure that the improvements were the results of changes in the school's curriculum or instruction, rather than because of some other factor such as students' improved test-taking ability. A related study would examine whether changes in the assessment test results really measure student achievement by comparing the overall assessment results with changes in other measures of student progress, such as results from standardized tests. To help ensure the continued reliability of the results, the Center is planning to include blocks of questions to help them determine whether teachers are teaching specific questions that are repeated on the tests over time.



APPENDIX C

Summary of Interviews with School District Superintendents

We contacted a sample of 63 of the 304 public school superintendents and four private school superintendents in Kansas. One of the private school superintendents said his school didn't use the assessment tests. As a result, we ended up with 66 usable interviews.

The purpose of the interview was to determine if schools were using standardized tests, what the superintendents had been told about the assessment tests, how the assessment test results are used in their districts, and other related questions. This appendix is a summary of the responses from the interviews. For open-ended questions, we only reported responses given by five or more superintendents.

Superintendent Interview

Public School **63**
Private School **3**

1. Have you changed which standardized test(s) your district uses since starting the assessments?
Yes - 24 (10 were not due to the Assessments)
2. Have you changed how frequently standardized tests are given in your district since starting the assessments?
Yes - 23 (2 were not due to the Assessments)
3. Do you plan to make any changes to the frequency of the standardized tests your district gives next year?
Yes - 21
4. What other indicators does your district use to determine student progress?
 - **Criterion-referenced Tests - 32**
 - **Grades - 22**
 - **Teacher -Developed Tests - 18**
 - **Student portfolios - 19**
5. Based on what the KSBE has told you about the assessments, how can you use the results of the Assessments? Does your district use the Assessments for this purpose?

	KSBE			District use			
	<u>yes</u>	<u>no</u>	<u>no response</u>	<u>yes</u>	<u>no</u>	<u>somewhat</u>	<u>unsure</u>
• evaluate or assist in evaluating individual students	8	49	9	13	40	12	1
• change instructional techniques	45	14	7	53	10	2	1
• change the curriculum in a particular area	43	16	7	48	17	1	0
• used to evaluate teacher performance	5	52	9	3	57	4	2
• part of graduation requirements	0	56	10	1	63	0	2
• compare buildings within the district	8	40	18	11	33	1	21
• compare the district to others in the State	8	42	16	11	54	1	0
• measure progress toward the statewide curriculum standards	52	8	6	60	5	0	1
• part of QPA process	60	2	4	64	2	0	0
• others							

6. Has the information provided by KSBE been useful in helping you interpret the assessment results?

Useful	usable, but left questions	unclear, left many questions	worthless, provided no guidance
22	25	14	2

7. Do you follow the State's schedule for offering the Statewide Assessments or some other schedule? Describe.

Yes - 56 No - 10 (offer writing more frequently)

8. Who do you release the results of the assessments to?

- parents (individual student results) 20
- principals 66
- teachers 62
- school board members 63
- the media 13
- state legislators 4
- students 19

9. Has your district made changes in the following due to assessment results? Briefly describe.

- | | <u>yes</u> | <u>no</u> | <u>somewhat</u> |
|-----------------------|------------|-----------|-----------------|
| • curriculum | 37 | 20 | 9 |
| • teaching | 46 | 16 | 4 |
| • in-service training | 48 | 13 | 5 |
| • other | | | |

10. As a superintendent, what is your overall opinion of the Statewide Assessment exams?

like very much	like	dislike	dislike very much
4	35	18	6

11. Do you have any information on the costs of the State Assessments to your district?

- Several Thousand Dollars (\$5 to \$6 Thousand) 5
- \$2 Thousand or less 6
- Costs include taking instructor's time from teaching class 11

12. Comments

see attached comments

Comments from Superintendent Interviews

Changes in Curriculum Due to Assessments

- 7 Added reading or other English classes
- 7 Added science class
- 6 Added additional Math classes
- 7 Increased emphasis on problem solving.
- 7 Focus on strengths, weaknesses, necessary skills, and what is tested.

Changes in Teaching Due to Assessments

- 7 Emphasize teaching problem solving.
- 8 Emphasize teaching the Curriculum Standards, thinking skills.
- 10 Provide instruction with the 6-trait writing model.
- 11 Include more cooperative learning and teaching and group work.

Changes in In-Service Training Due to Assessments

- 10 Have held meetings to incorporate teaching and curriculum changes.
- 10 Have held meetings to discuss needed improvements.

Concerns about Validity and Reliability of Assessments

- 13 Tests not yet stable. Some tests are good, some are bad.
- 14 Tests may not measure achievement of skills, too few questions to judge performance, tests cover material not yet taught.

Administering the Assessments - Expense, Timing,

- 12 Testing students too much, too time consuming.
- 6 Assessment format difficult, goals are set too high.

Using Assessment Results

- 8 Results are worthless, no new information, not needed.
- 5 Assessments don't track students over time or measure student progress.
- 5 Difficult to understand purpose and interpret results.

Positive Comments about the Assessments

- 11 Assessments are driving curriculum and instruction in a positive way, good for evaluating progress.

Other Concerns

- 5 Want to maintain local control, have district developed tests.
- 7 Not good tests for small districts.

APPENDIX D

Summary of Survey Responses from Principals

On March 22, 1996, we mailed surveys to a sample of principals from public and private schools in Kansas. The sample consisted of every third elementary, middle, and high school listed in the *1995-1996 Kansas Educational Directory*. In all, 492 surveys were mailed to public school principals and 51 surveys were mailed to private school principals. We received 336 surveys from principals (307 public school and 29 private school principals), giving us an overall response rate of 62%. This appendix is a summary of the responses from the surveys returned by principals. For open-ended questions, we reported only those responses given by five or more principals.

Principal Survey

The Legislative Post Audit Committee of the Kansas Legislature has directed the Legislative Division of Post Audit to conduct a performance audit of the State Assessment tests used as one measure of performance under Quality Performance Accreditation. One of the objectives of the audit is to determine how schools are using the Assessment tests and how satisfied they are with the tests. To help meet this objective, we are surveying Kansas Principals who have used the Assessment tests.

We would appreciate it if you would answer the following questions. The returned surveys will be included in the audit working papers which become part of a public document once the audit is completed. Please return the survey in the accompanying, self-addressed, postage-paid envelope (or fax to the attention of Ellyn Sipp at 913-296-4482) by **Friday, April 5, 1996**. If you have questions about the audit or the survey, please contact Ellyn Sipp or Scott Brunner at Legislative Post Audit, 800 S.W. Jackson, Suite 1200, Topeka, KS 66612 or phone 913-296-3792.

Please provide the following information about your school.

Private School 29 City _____

Public School 307 USD# (if applicable) _____

What is your building level? elementary middle/junior high high school

197 48 91

What grades are taught in your building? _____

How long have you been a Principal in Kansas schools? _____ years

On the following questions, circle the response that most closely reflects your opinion about the State Assessment tests.

1. The State Assessment tests provide information that:

- assists in evaluating student performance toward meeting the Kansas Curriculum Standards

strongly agree	agree	disagree	strongly disagree	no response
23	249	48	14	2

- provides feedback facilitating change in curriculum and instructional practice

strongly agree	agree	disagree	strongly disagree	no response
24	235	58	16	3

- helps evaluate teacher performance

strongly agree	agree	disagree	strongly disagree	no response
3	71	192	64	6

- allows building comparisons within the district on the State Curriculum Standards

strongly agree	agree	disagree	strongly disagree	no response
9	168	102	42	15

- allows district comparisons on the State Curriculum Standards

strongly agree	agree	disagree	strongly disagree	no response
9	191	92	39	5

For the following questions, check the boxes next to ALL items that apply to your building or district.

2. I/we use the results of the State Assessment tests...

- to compare individual student achievement **159**
- to recommend changes in teaching methods **181**
- to identify needed changes in the curriculum **253**
- to measure progress toward the State Curriculum Standards **260**
- to evaluate individual teachers **6**
- to compare my building with others in the district **84**
- to assist teachers' evaluation of student performance **187**
- as a requirement for graduation **6**
- other (please list)

3. Individual student results of the State Assessment tests are provided to...

- students' parents **143**
- teachers in my building **306**
- other (please list)
- QPA teams* **2**
- School Improvement Team/Site Council* **6**
- Administrators/District Office* **13**

4. Summary level results of the State Assessment tests are provided to...

- students' parents **138**
- teachers in my building **312**
- the local School Board **296**
- the local media **92**
- other (please list)
- QPA teams or QPA reporting* **7**
- Site Councils* **34**

5. How satisfied are you with each of the following features of the Kansas Assessment program?

- as measures of the Kansas Curriculum Standards

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
4	187	98	24	12	11

- packaging and dissemination of materials

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
12	244	51	16	8	5

- test administration instructions

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
16	266	32	12	4	6

- instructions on how to score the performance parts of the tests

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
6	233	63	15	13	6

- information on district and building results (hard copy and data disks)

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
8	241	54	8	14	11

- Center for Educational Testing and Evaluation (CETE) assistance

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
12	162	57	9	84	12

- performance parts of the tests (includes the writing assessment)

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
14	201	72	27	3	19

- time to administer the tests

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
6	239	54	28	1	8

- the Kansas Assessment tests overall

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no reponse
1	172	98	35	2	28

6. Please explain why you either like or dislike the State Assessment tests?

See attached comments.

Please indicate the number of surveys you distributed to teachers in your building. **250**

The staff of Legislative Post Audit wishes to thank you in advance for your cooperation by completing this survey. Please return the completed survey in the enclosed, self-addressed, postage-paid envelope or fax the completed survey to the attention of Ellyn Sipp at 913-296-4482.

Principal survey comments

Concerns about Validity and Reliability of Assessments

- 28 Shouldn't compare scores between districts, between urban and rural schools, or between social and economic groups.
- 27 Scoring is inconsistent and subjective among districts, between teachers.
- 27 The Assessments do not test what students know or what we teach in class. They are above the level of knowledge of students.
- 22 Tests always changing. Can't compare results, and schools can't show improvement.
- 13 No evidence the tests are reliable and valid.
- 9 Open ended questions are too hard, have too many answers, or are not developmentally appropriate
- 8 Teachers will teach to the test or inflate scores on the Assessments
- 7 Assessment results shouldn't be used to compare buildings over time. Changes in classes effect results over time.

Administering the Assessments - Expense, Timing,

- 48 Take too much time to score, away from class
- 10 Instructions are not clear for administering and grading the assessments, hard to administer
- 9 Give too many tests in grade 5 and grade 10.
- 7 Need more flexibility in timing of exam. Spring semester is not a good time to give the tests.
- 6 Schools need additional money to grade Assessments.
- 6 Assessments are too expensive for districts.

Positive Comments about the Assessments

- 31 The Assesments drive and are an incentive for curricular change and school accountability.
- 15 The Assessments are a good measure of what we do, better than norm referenced tests.
- 9 Statewide test is a good idea. The Assessments are improving.
- 5 Open ended questions and the performance items are good

Using Assessment Results

- 17 Results are difficult to interpret.
- 12 Not better than norm referenced tests, can't compare Assessment results to norm referenced tests.
- 10 Assessments don't have individual student reports, and can't provide results to parents.
- 10 No longitudinal data or base line data from Assessments.

Other Concerns

- 14 There is little followup from CETE on questions and results are not provided in a timely manner. CETE doesn't take suggestions from schools.
- 9 Assessments enforce a state curriculum, reduce local control of curriculum and schools.
- 8 Fear misuse of results. Assessments are politically motivated. Assessments are Legislative involvement and interference in education.
- 7 Students don't take the Assessments seriously



APPENDIX E

Summary of Survey Responses from Teachers

On March 22, 1996, we sent a packet of teacher surveys to every eighth private and public school principal in our sample. We asked the principals to distribute the surveys to teachers in their schools who had experience with the assessment tests. Elementary school principals received ten surveys to distribute and middle or high school principals received fifteen surveys. We received 216 surveys from teachers for analysis (208 public school and 8 private school teachers). This appendix is a summary of the responses from the surveys returned by teachers. For open-ended questions, we reported only those responses given by 5 or more teachers.

Teacher Survey

The Legislative Post Audit Committee of the Kansas Legislature has directed the Legislative Division of Post Audit to conduct a performance audit of the State Assessment tests used as one measure of performance under Quality Performance Accreditation. One of the objectives of the audit is to determine how schools are using the Assessment tests and how satisfied they are with the tests. To help meet this objective, we are surveying Kansas teachers who have used the Assessment tests.

We would appreciate it if you would answer the following questions. The returned surveys will be included in the audit working papers which become part of a public document once the audit is completed. Please return the survey in the accompanying, self-addressed, postage-paid envelope (or fax to the attention of Ellyn Sipp at 913-296-4482) by **Friday, April 5, 1996**. If you have questions about the audit or the survey, please contact Ellyn Sipp or Scott Brunner at Legislative Post Audit, 800 S.W. Jackson, Suite 1200, Topeka, KS 66612 or phone 913-296-3792.

Please provide the following information about your school.

Private School City _____
 Public School USD# (if applicable) _____

What grade do you teach? _____ How many years have you been teaching? Average 16 years

Which Assessment tests have you given?

Math	Reading	Science	Writing	Social Studies
96	84	57	90	52

For the following questions circle the response that most closely reflects your opinion about the State Assessment Tests.

1. The State Assessment tests provide useful information that:

- assists in evaluating student performance toward meeting the Kansas Curriculum Standards

strongly agree	agree	disagree	strongly disagree	no response
8	124	57	18	9

- gives me feedback on areas where my teaching methods need to change

strongly agree	agree	disagree	strongly disagree	no response
6	95	82	28	5

- gives me feedback on areas where the curriculum needs to change

strongly agree	agree	disagree	strongly disagree	no response
12	115	66	17	6

2. How have you been told you can use results from the Assessments?	Yes	No	No Response
• to change your teaching methods	111	93	12
• to change the content of your classroom curriculum or lesson plans	122	84	10
• to evaluate individual students	110	94	12

3. How familiar are you with the State Assessment tests? Check all that apply.			
• I have read the State Assessment test.	208	4	4
• I have taken the State Assessment test in my subject.	111	94	11
• I have scored the written or performance components of the Assessments.	188	23	4

4. How satisfied are you with each of the following features of the Kansas Assessment program?

• as measures of the Kansas Curriculum Standards	very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
	1	82	76	36	11	10

• packaging and dissemination of materials	very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
	18	153	20	9	13	3

• test administration instructions	very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
	19	150	34	9	1	3

• instructions on how to score the performance parts of the tests	very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
	9	124	39	21	17	6

• information on district and building results (hard copy and data disks)	very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
	9	81	49	27	43	7

- Center for Educational Testing and Evaluation (CETE) assistance

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
3	39	26	7	132	9

- performance parts of the tests (includes the writing assessment)

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
7	81	63	36	13	16

- time to administer the tests

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
17	145	34	9	2	9

- the Kansas Assessment tests overall

very satisfied	satisfied	dissatisfied	very dissatisfied	no basis for opinion	no response
4	71	73	43	3	22

5. Please explain why you either like or dislike the State Assessment tests?

The staff of Legislative Post Audit wishes to thank you in advance for your cooperation by completing this survey. Please return the completed survey in the enclosed, self-addressed, postage-paid envelope or fax the completed survey to the attention of Ellyn Sipp at 913-296-4482.

Teacher survey comments

Number Who Said They Liked or Disliked Specific Assessment Tests, and Why

Reading Assessment

- 5 like
- 8 don't like: too vague, boring reading selections

Writing Assessment

- 18 like: six-trait model, authentic assessment
- 9 don't like : not collect a good sample of work, instructions not clear, scoring is subjective.

Math Assessment

- 2 like: open ended questions
- 33 don't like : timed exercises, instructions not clear, open ended exercises are too hard, require too much reading, don't match curriculum or what students know.

Social Studies Assessment

- 0 like
- 17 don't like : performance parts too hard, can't compare, take too long, not what we teach, elementary students not prepared for the work

Science Assessment

- 3 like:
- 12 don't like : hard to administer, test items are not what is taught, should be at 11th grade, performance influenced by writing ability

Concerns about Validity and Reliability of Assessments

- 18 Scoring is inconsistent/subjective among districts, between teachers
- 17 Not test what students know, what we teach in class, above the level of knowledge of students
- 7 Students can't do the multiple correct, multiple choice questions.
- 5 Not clear the tests are reliable and valid

Administering the Assessments - Expense, Timing,

- 29 Take too much time to score, away from class
- 12 Instructions are not clear for administering/grading the assessments, hard to administer

Using Assessment Results

- 6 No individual student reports, can't provide results to parents
- 6 Shouldn't compare scores between districts. Comparisons between urban and rural schools. Comparisons between SES groups.

Positive Comments about the Assessments

- 6 Assessments have helped be reevaluate my teaching.

Other Concerns

- 14 Lack of followup from CETE, not provided results in a timely manner, CETE doesn't take suggestions from schools.
- 7 Students don't know what is expected of them on the Assessments.
- 5 Students don't take the Assessments seriously



APPENDIX F

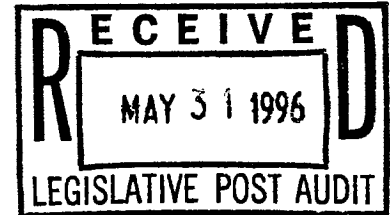
Agency Response

On May 21, we provided copies of the draft audit report to the State Board of Education. Its response is included as this appendix.

Kansas State Board of Education

120 S.E. 10th Avenue, Topeka, Kansas 66612-1182

May 30, 1996



Ms. Barbara Hinton
Legislative Post Auditor
Mercantile Bank Tower
800 Southwest Jackson Street, Suite 1200
Topeka, Kansas 66612-2212

Dear Ms. Hinton:

Thank you for your invitation to provide "comments, corrections, or clarifications" to your draft copy of the audit report on the Kansas Assessments. Agency staff have reviewed the document, and our response to your recommendations is attached. It is understood that our response will be included in an appendix to your report.

Let me express my appreciation for the professional manner in which your staff conducted this audit and their willingness to undertake the evaluation of complex subject matter. As always, it is a pleasure to work with you and your staff.

If I may be of further assistance with this report, please let me know.

Sincerely,



Dale M. Dennis
Interim Commissioner of Education

DMD:ng
Enclosure

58.

Dale M. Dennis
Deputy/Assistant Commissioner
Division of Fiscal Services and Quality Control
(913) 296-3871
Fax No. (913) 296-7933

KSBE RESPONSE TO RECOMMENDATIONS

***Recommendation #1:** To ensure that all educators are informed about the appropriate and inappropriate uses of the State assessment tests, the Department of Education should continue to work with school district superintendents and local boards of education to ensure that this information is communicated to all educators in each school district.*

Response #1: As was confirmed in this report by both superintendents and principals, both the Kansas State Board of Education staff and the Center for Educational Testing and Evaluation spend a great deal of time working with school districts to ensure that information is communicated about appropriate and inappropriate uses of the assessment. Eleven workshops were held last year alone for teachers, principals, and superintendents. Thirteen are scheduled for this year. Printed material, such as interpretation manuals, state reports, and periodic communication from the Kansas State Board of Education which deal with uses of the assessments have been distributed. These efforts will continue.

***Recommendation #2:** To lessen the likelihood that test score information contained in building report cards it issues for school districts will be misinterpreted, the Department should do the following:*

- a. include clear cautionary language on the report card that will alert the reader to limitations in the use of test scores.*
- b. advise school districts about what additional information they should have assembled to provide to members of the press or other interested parties to help to explain differences between individual schools and districts.*

Response #2: Any time that test score information is released, the Kansas State Board of Education staff and the Center for Educational Testing and Evaluation include cautionary language about the limitations of the uses of those scores. Quality Performance Accreditation materials and instructions are clear about the need for multiple sources of evidence in all instances. Therefore, this recommendation is in line with current Kansas State Board of Education policy and practice, especially in light of the potential for media misrepresentation and for judgments which could be made about school districts. The Kansas State Board of Education concurs with this recommendation.

***Recommendation #3:** Assessment test scores are only reliable within a wide range for comparing individual student scores to the "Student Performance Expectations" the Legislature has required the Department to establish. As a result, the Department should assess what changes would need to be made to those assessment tests to make such comparisons more reliable and more valid and what the associated cost would be. The Department should present that information to the education committees of the Legislature by the start of the 1997 session, so that the Legislature can make an informed decision about whether it wants to retain the student performance expectations, and what it wants them to be used for.*

Response #3: The Kansas State Board of Education staff and Center for Educational Testing and Evaluation have already begun to move toward the setting of individual standards of performance. Those standards will be set on the total scores for each subject area. Those total scores have reliabilities that are well within acceptable limits for use at the individual level. Individual score reliabilities are comparable to individual score reliabilities in other states. As is ALWAYS the case, the Kansas State Board of Education will strongly recommend that NO one test be used to make a high-stakes decision about either a building or an individual student. **Use of three sources of data is always recommended when a decision is being made**, whether we are talking about the Kansas Assessments or the ITBS, and whether we are talking about an individual student or a building. The Kansas State Board of Education staff will be pleased to report to the 1997

Legislature on the issue of student performance levels, the development of which was required by the 1995 Legislature.

***Recommendation #4:** To ensure that the results of the assessment tests are as reliable and valid as possible, the Department of Education should continue to monitor the consistency of local scoring of exams, and evaluate whether changes in scores reflect changes in what is being taught and how it is being taught.*

Response #4: The Kansas State Board of Education completely concurs with this recommendation. As you know, we are in the process of doing these studies at the present time. The studies will be available by the fall of 1997 and will be ongoing after that time, provided that we have continued funding for this purpose.

***Recommendation #5:** To ensure that the State assessments are done in a cost-effective manner, the Department should continue to monitor the performance components of the tests to ensure that the information they provide is worth the cost, and to determine whether there are other less costly means of obtaining the desired information about how well schools are teaching Kansas students.*

Response #5: Yes, we concur that Kansas State Board of Education staff should continue to monitor both the performance and multiple-choice/multiple-mark portions of the state assessment. The Kansas State Board of Education and Center for Educational Testing and Evaluation will also certainly continue to be a part of the professional dialogue about performance assessment.

It is the contention of the Kansas State Board of Education that very few experts will agree with the claim that performance assessment doesn't "yield that much more additional information than other types of tests." Kansas State Board of Education staff are active in such organizations as American Psychological Association, Kansas Council of Teachers of Mathematics, National Council of Teachers of Mathematics, Kansas Association of Teachers of Science, National Council of Teachers of English, International Reading Association, Kansas Reading Association, Kansas Council for Social Studies, National Council for Social Studies, American Educational Research Association, and National Council of Measurement in Education. Without exception, these organizations not only endorse, but strongly encourage the use of performance assessment. As you have noted, the trend nationally is toward, not away from, performance assessment as part of the large-scale assessment system.

Evidence supporting the use of performance assessment comes from authorities in every subject area specialization. Lauren Resnick from the New Standards Project says, "Many of tests we do use are unable to measure what should be the hallmark of a 'thinking' curriculum; the cultivation of students' ability to apply skills and knowledge to real-world problems. Testing practices may in fact interfere with the kind of higher order skills that are desired."¹ The business community has been using performance assessment in hiring and performance appraisals for years. Current intelligence theories, like those of cognitive psychologist Robert Sternberg, that stress multiple intelligences and different learning styles, endorse the need for performance assessment. Psychometric evidence of a variance between two assessment methods includes data from the Kentucky Assessment System, which reports low correlations between open-ended and multiple-choice questions.² This suggests that they are, indeed, providing different information.

¹ Resnick, L. B. (1987). *Education and Learning to Think*. Washington, D.C.: National Academy Press.

² Kentucky State Department of Education (1995). *KIRIS Accountability Cycle 1 Technical Manual*. Frankfort, Kentucky: Kentucky State Department of Education.

