An Analysis of the Commonwealth Accountability Testing System

Project Staff

Office of Education Accountability

Marcia Ford Seiler, Director
Gerald Lunney, Ph.D.
Sabrina Olds
Pam Young

Legislative Research Commission

Barry Boardman, Ph.D.
Mike Clark, Ph.D.
Kara Daniels
Greg Hager, Ph.D.
Dan Jacovitch
Erin McNees
John Perry
Jon Roenker
Ginny Wilson, Ph.D.

Research Report No. 328

Legislative Research Commission
Frankfort, Kentucky
lrc.ky.gov

Accepted July 6, 2005 by the
Education Assessment and Accountability Review Subcommittee

Paid for with state funds. Available in alternative form by request.

Foreword

The 2004 General Assembly directed the Office of Education Accountability to study the Commonwealth Accountability Testing System (CATS). Included in the study was a review of the validity and reliability of student-level scores, the state and local cost of the assessment system, the opinions of school personnel and others of CATS, and alignment of CATS with No Child Left Behind requirements. To carry out the study staff, worked with Legislative Research Commission staff, external contractors, and Kentucky Department of Education staff.

Office of Education Accountability staff would like to thank the many people who worked on and provided information required for this study. Petie Day, Vickie Terry, Bill Insko, Roger Ervin, Mary Ann Miller, Bonnie Brinly, Starr Lewis, Cherry Boyles, and other staff from the Kentucky Department of Education provided finance and assessment data as well as information and background on the assessment and accountability system.

The expertise and knowledge of LRC staff who worked on this study is appreciated. In addition, appreciation is extended to LRC Education Committee staff, including Sandy Deaton, Audrey Carr, and Jonathan Lowe.

Finally, valuable assistance was provided by the members of the National Technical Advisory Panel on Accountability and Assessment; AEL, Inc.; and all of the teachers, principals, superintendents, school board members, parents and guardians, and students who provided guidance through the focus groups and responded to our surveys.

Robert Sherman
Director

Legislative Research Commission
Frankfort, Kentucky
July 6, 2005

Contents

List of Tables

List of Figures

Summary

In 2004, the General Assembly adopted Senate Joint Resolution 156 directing the Office of Education Accountability to review the Commonwealth Accountability Testing System (CATS).

Chapter 1: An Overview of the Commonwealth Accountability Testing System

In 1990, the General Assembly enacted House Bill 940, known as the Kentucky Education Reform Act (KERA). To measure student achievement of the academic goals of the law, the Kentucky Board of Education created academic expectations, which were used to create the Program of Studies: "the minimum content required for all students before graduating from Kentucky high schools." The Core Content for Assessment is the essential content drawn from the Program of Studies that will be covered on the state assessment.

The General Assembly enacted House Bill 53 in 1998, giving guidelines for the Kentucky Board of Education to replace the previous assessment and accountability system with the Commonwealth Accountability Testing System. Under CATS, there are two types of assessments administered to students: the Kentucky Core Content Test (KCCT) and the Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5). CTBS/5 is the nationally norm-referenced test that assesses students in reading, math, and language arts in grades 3, 6, and 9. KCCT is the criterion-referenced assessment administered to students in grades 4, 5, 7, 8, 10, 11, and 12. KCCT assessments cover reading, math, science, social studies, arts and humanities, practical living and vocational studies, and writing.

Based on their performances on each of the assessments that cover the core content, students are divided into four categories: novice, apprentice, proficient, and distinguished. The performance levels of novice and apprentice are further divided for reading, mathematics, science, and social studies.

Kentucky uses multiple-choice and open-response questions in accordance with state law. For students in grades 4, 7, and 12, there is a section of KCCT in which they are assessed based on their writing. Students are given two writing prompts and must respond to one prompt of their choosing within a time limit. A contractor scores the CTBS/5 tests, KCCT subject tests, and on-demand writing prompts.

As part of CATS, students' writing portfolios are assessed in grades 4, 7, and 12. The writing portfolio is a compilation of each student's best writings, produced over time, with at least one piece from different categories of writing. Writing portfolios, scored by trained teachers, are scored using the performance categories of novice, apprentice, proficient, and distinguished. The rater reads the whole portfolio and assigns one score based on the overall performance of the student.

KCCT is administered using different test forms to provide greater coverage of the core content. As a result, students completing different forms receive different questions, which can result in differences in difficulty. Test forms are designed to minimize these differences. Some differences can persist, so scores are adjusted.

Procedures for including special student populations in the state assessment program are established in regulation. Special populations addressed include students with disabilities and students for whom English is not their primary language. The regulation establishes under what conditions students may receive an accommodation, modification, or both. An accommodation is a change in the testing environment or process. A modification is a change in the instrument used for assessment.

In spring 2004, less than 1 percent of students with limited English proficiency took KCCT tests with accommodations or modifications. The percentage of students using accommodations or modifications based on disability varied by grade, ranging from 6 percent to more than 11 percent.

Students with moderate to severe disabilities for whom traditional assessments are inappropriate are assessed via the Kentucky Alternate Portfolio, which is submitted and assessed during grades 4, 8, and 12.

Schools and school districts are held accountable based on students' performance on the various types of assessments and relevant nonacademic measures such as attendance or graduation rates. Schools that meet or exceed established goals are eligible for financial rewards. Those not meeting goals are subject to consequences.

The academic indices for the different KCCT content areas are combined with nonacademic factors and the results of CTBS/5 tests to construct the school's overall accountability index. The long-term goal for each school and district, and for the state as well, is to reach proficient by 2014. Proficient is defined as a score of 100 or greater on a 140-point scale. Intermediate goals are based on the school's progress toward scoring 100 or greater in 2014 from its own starting point.

By statute, parents of each student must receive a report card on the performance of the student's school. Each year, districts receive an individual report on each student's performance on CATS assessments. There is broad discretion at the district level as to the form and substance of individual student progress reports. This means that districts are not required to distribute the student-level CATS reports to parents. For districts that do distribute the reports, the means to do so vary.

Chapter 2: Alignment of the Commonwealth Accountability System With the Requirements of the No Child Left Behind Act

Twelve years after the General Assembly adopted KERA, the federal government adopted the No Child Left Behind Act (NCLB). Both KERA and NCLB required the establishment of education standards and assessments. The specific requirements under

these two pieces of legislation, however, differed. The Kentucky Board of Education, which is responsible implementing both CATS and NCLB, decided to retain CATS unchanged and incorporate the additional requirements of NCLB.

There are several key differences between CATS and NCLB. The first difference is the subjects evaluated under the two systems. CATS evaluates performance in seven subjects; math, reading, science, social studies, writing, practical living and vocational studies, and arts and humanities. NCLB only evaulates performance in reading and math. In 2007, science will be added for NCLB but will not be included in determining whether schools have met their standards. A second difference is that under CATS each school has its own standards to acheive, under NCLB the standards are the same for each level of school (elementary, middle, or high). Third, while CATS is designed to measure the progress of schools, NCLB is designed to measure the progress of schools and groups of students within the schools. Finally, performance is evaluated every two years under CATS but is evaluated annually under NCLB.

NCLB requires that states test students in reading and math in grades 3 through 8. CATS, however, does not provide reading and math tests in each of these grades. In order to comply with NCLB, the Kentucky Board of Education decided to add norm-referenced reading and math tests in the remaining grades. The norm-referenced tests that will be used in some grades will not provide complete coverage of the core content as required by NCLB. Therefore, these tests will be augemented to provide complete coverage.

Under NCLB, schools must achieve what is deemed to be adequate yearly progress toward their educational goals. The goal for each school is to reach a predefined level of proficiency by 2013-2014. Adequate yearly progress consists of three objectives. First, each school must have a certain percentage of its students scoring at the proficient level each year. NCLB also requires that at least 95 percent of the students within a school participate in the testing. Finally, NCLB requires that adequate yearly progress include academic indicators in addition to student assessments. In Kentucky, this will include graduation rates for high schools and each school's full CATS accountability index for the prior year for elementary and middle schools. The index includes attendence and retention rates for both elementary and middle schools and dropout rates for middle schools.

Schools face separate accountability requirements under NCLB and CATS, with separate consequences and sanctions for failing to meet their goals under either system. While both provide escalating consequences the longer a school fails to meet its goals, the type of consequences and the timeline for the consequences are different. Under NCLB, a school that fails to meet adequate yearly progress for two consecutive years is identified as needing improvement. School districts must provide students in these schools with the option to tranfer to another school that is not classified as being in need of improvement. Schools that continue to not meet their goals must provide supplemental services such as tutoring and may eventually be subject to alternative governance such as the replacement of staff or the direct management of the school by the Kentucky Department of Education.

NCLB also requires that teachers who teach "core academic subjects" meet certain standards defining "highly qualified" teachers. The term "core academic subjects" refers to English, reading, math, science, foreign languages, civics and government, economics, arts, history, and geography. Teachers must have a baccalaureate-level degree and full Kentucky certification. In addition, teachers must demonstrate subject knowledge and teaching skills, which can be demonstrated by passing various exams or completing an undergraduate major or graduate degree in the subject being taught.

In order to comply with the provisions of NCLB, the Kentucky Department of Education and the Kentucky Board of Education have augmented CATS. CATS will, however, continue to operate as it has in the past. The modifications for NCLB are detailed in Kentucky's state accountability plan, which was approved by the United States Department of Education. Kentucky recently submitted additional modifications detailing how Kentucky will set education standards and assess student and school performance. The U.S. Department of Education will review these changes at some point in the future.

Chapter 3: Reliability and Validity of the Commonwealth Accountability Testing System

One of the goals of education assessments is to make inferences about the performance of a student or a group of students. For these inferences to be useful as a basis for making decisions, the assessments must be sufficiently reliable and valid indicators of the student's performance or the group's performance. Reliability refers to whether an assessment would yield similar results if conducted multiple times. If students receive similar scores each time they take an assessment and their true understanding of the material did not change, the assessment would be considered reliable. Validity refers to whether the assessment actually measures the skills or knowledge it was intended to measure. CATS was designed to measure whether the students in a school as a group have mastered the core content.

AEL, a nonprofit educational research and development firm, reviewed the relevant research literature addressing the reliability and validity of education assessments. The research noted various factors that influence the reliability of an assessment. Reliability can be improved by increasing the number of test items students are required to complete. A larger number of test items reduces the probability that a student's score is due to chance. Reliability can also be improved by increasing the number of individuals who rate an individual student's performance. Even with explicit instructions for rating a student's work, raters may differ in their interpretations of these instructions. Increasing the number of raters improves the consistency of how work is evaluated.

The consistency with which various raters evaluate students' work is often measured by the percentage of times different raters agree on the score to assign to a student's work. A higher percentage of agreement suggests greater consistency across raters. A low percentage suggests that the raters are interpreting the work differently. The reliability of a score is reduced if raters score work inconsistently. In 2001, the percentage of times that raters agreed on the scores provided for open-response and on-demand writing

questions in CATS ranged from 68.5 percent to 90 percent, depending on the subject and grade. Agreement rates were generally higher for middle school students.

Because assessments are not perfectly reliable, there is some chance that a student's or a school's true performance is not actually at the level that it appears to be based on the results of an assessment. This possibility can result in misclassification of the student or school. Classification accuracy refers to the probability that an assessment correctly classifies a student's or school's performance. For CATS, the classification accuracy refers to the probability that a school is correctly classified as novice, apprentice, proficient, or distinguished. The Human Resources Research Organization, or HumRRO, concluded that the classification accuracy of CATS was 77 percent. That is, there is a 77 percent chance that a school is correctly assigned. This suggests that out of 100 schools, 77 percent would be correctly classified and 23 percent would be incorrectly classified.

While CATS is designed to measure performance at the school level, it does result in student-level scores, which provide some signal of a student's performance. Student-level scores, however, are generally less reliable than school-level scores, and there are concerns over whether they are valid indictors of student skills and knowledge. Therefore, any use of CATS scores at the student level would require an understanding of the risk of incorrectly classifying a student and an acceptance that taking this risk is approriate.

The literature notes two issues associated with using student-level scores. The first is the likelihood that the scores provide an accurate indication of a student's ability. The second is the consequences associated with a specific use of the scores. The research suggests that decisions that result in greater consequences should require higher levels of reliability to reduce the risk of incorrectly assigning these consequences. The National Technical Advisory Panel for Assessment and Accountability (NTAPPA) suggested that CATS scores at the student-level are not sufficiently reliable to be used alone or for decisions that would have "high stakes" consequences. The group also suggested, however, that student-level scores could be used in the context of additional measures of student performance for decisions such as determining Kentucky Educational Excellence Scholarship awards or determining whether a student needs additional assistance.

CATS was designed to provide complete coverage of the core content at the school level. Each student, however, is assessed on only a portion of the core content. By not completely covering the core content, student-level CATS results may be considered somewhat incomplete. NTAPPA suggested that the validity of student-level scores could be improved by increasing the number of tasks on the assessments, increasing the coverage of the core content, and allowing students to retake the tests. The group did note that there would be additional costs associated with these changes.

Chapter 4: A Summary of the Surveys of Kentucky Teachers, Principals, Superintendents, Students, Parents and Guardians, and School Board Members

In 2004, Legislative Research Commission contracted with AEL, which conducted focus groups and surveys of teachers, principals, superintendents, high school students, parents and guardians, and school board members in Kentucky. Focus groups were conducted in six regions across the state for the primary purpose of providing information that was used to create questions for the surveys. The number of respondents and response rates—percentage of surveys returned—were relatively low for most groups.

Almost all teachers, principals, and superintendents who responded indicated that multiple-choice questions were appropriate for grades 4 and 5, 7 and 8, and 9 through 12. Most members of each group rated open-response and on-demand writing as appropriate for grades 7 and 8, and 9 through 12. Nearly half of the teachers and principals did not agree that on-demand writing was appropriate for 4^{th} and 5^{th} graders.

Teachers, principals, and superintendents were asked whether six KCCT subject tests (reading, math, science, social studies, arts and humanities, and practical living and vocational studies) were valid measures of core content knowledge for special education students and for students with limited English proficiency. A majority of each group of educators responded that the six tests were invalid measures for either group. More than half of principals and superintendents disagreed that the alternate portfolio was a valid measure of knowledge of core content for special education students.

Teachers, principals, and superintendents were also asked whether the six KCCT tests were valid measures of knowledge of the core content for all other students. At least two-thirds of superintendents agreed that all six tests were valid for elementary, middle, and high school students. This was also true for approximately 60 percent or more of principals. The majority of teachers responded that most of the tests were valid, but support for this position was less widespread than for the other two groups.

Approximately 60 percent or more of surveyed high school students agreed each test was a good measure of their knowledge of the subject. More than 60 percent of surveyed parents and guardians indicated that CATS tests were fair measures of their children's knowledge of school subjects.

Less than half of teachers, principals, and superintendents agreed that the writing portfolio, on-demand writing, or the alternate portfolio was a valid measures of special education students' writing. Less than 40 percent of each of the three groups agreed that either the writing portfolio or on-demand writing was valid for students with limited English proficiency.

For all other students, a majority of superintendents rated portfolios and on-demand writing as valid at all levels. Teachers and principals distinguished between the types of writing assessment. More than half of teachers and principals agreed that on-demand

writing was a valid means of assessment for students at all levels. There was no majority of either group agreeing that portfolios were valid for any school level.

Two-thirds of students agreed that the writing portfolio was a "good measure" of their writing. The corresponding figure for on-demand writing was 59 percent.

Educators and students were asked about the integration of writing portfolios into a student's learning experience. There was broad agreement that work on portfolios was done throughout the school year rather than during a block of time set aside for them. Approximately 80 percent of students responded that work on portfolios was done in several classes. More than half of teachers, principals, and superintendents agreed. Approximately two-thirds each of teachers and principals reported that a student working on a portfolio worked with more than one teacher rather than just one. Students' and superintendents' responses were nearly evenly divided on the question.

Educators were asked whether they agreed or disagreed "that the amount of time it takes to prepare writing portfolios is appropriate to the benefit received by students." More than 80 percent each of teachers and principals and 77 percent of superintendents disagreed or strongly disagreed that portfolios were worth the time. More than one-half of the principals strongly disagreed, and nearly one-half of teachers did so.

The writing portfolio was the only item that a majority of each group of respondents would reduce or eliminate from the school academic index. More than 40 percent each of teachers and principals wanted on-demand writing to be emphasized less in the index or removed.

Generally, more than one-half—sometimes much more—of teachers, principals, and superintendents agreed that

- getting ready for or taking CATS tests took too much time,
- CATS testing was too stressful and reduced enjoyment of teaching and learning,
- teaching what was to be covered on the tests was too limiting, and
- teachers and students were forced to cover material too quickly.

Educators also agreed with positive statements about CATS testing. More than one-half of teachers, principals, superintendents, and school board members agreed that testing provided needed focus and organization. At least 75 percent of each group agreed that testing helped align the curriculum.

The average teacher's estimate was that 29 percent of work time during the school year was spent preparing for CATS testing. The average student's estimate was 32.5 percent. Approximately a third of each group responded that preparation took 10 percent or less of work time. Approximately one-half of teachers and students answered that preparation took 20 percent or less of work time. Fourteen percent of teachers and 18 percent of students answered that preparation for testing took up more than half of their work time.

Almost all the parents and guardians who responded to the surveys indicated that they received the CATS student and the school report card. More than three-fourths also answered that the information in each type of report was presented in a way that was easy to understand. More than 60 percent answered that the school report card helped them to understand how their children's schools were doing. Just under one-half indicated that the CATS tests reports helped them understand how their children were doing in school. Only 14 percent of the parents and guardians who responded to the survey indicated that someone from their child's school usually discussed the CATS tests reports with them.

Chapter 5: The Cost of CATS

The total cost of CATS for FY 2004 was estimated to be $37.2 million. This reflects both estimated state-level expenditures of $20.8 million and estimated local-level expenditures of $16.4 million related to CATS. This amounts to about $78 being spent per assessed student. For comparison purposes, the estimated per student assessment expense represents slightly less than 1 percent (0.97 percent) of the $8,029.84 spent per student in Kentucky during the 2004 school year. While the $8,029.84 does not contain state-level assessment expenditures, the comparison provides some context to the magnitude of estimated total CATS-related spending.

Obtaining state-level CATS-related expenditures was relatively direct: expenditures of state agencies charged with various activities related to the accountability system were added together. However, obtaining local-level expenditures was not so direct. Since there is no formal local-level accounting of CATS-related expenditures, local-level CATS-related expenditures were collected through two cost surveys. The first cost survey was administered to the central office financial officer in each school district. The second cost survey was given to similar administrators at individual schools.

There was significant variation in reported local-level expenditures between schools and between central offices. The variation might be due to local officials classifying costs differently. The variation might also suggest that districts and schools respond to CATS differently. For example, schools might develop different programs to improve students' writing. In addition, some schools or districts might associate certain costs with CATS that other schools and districts consider normal costs of education.

CATS-related expenditures were classified into three categories: advisory and research; administration and implementation; and accountability. Total spending in FY 2004 was approximately $697,000 for advisory and research, all spent at the state level. Administration and implementation expenditures were estimated to be approximately $22 million, fairly evenly split between local- and state-level expenditures. Accountability costs were estimated to be about $14.5 million. State-level spending comprised approximately 63 percent, or about $9.2 million, of the total accountability portion of expenditures, while local-level spending was estimated to total $5.3 million.

Chapter 1

An Overview of the Commonwealth Accountability Testing System and This Report

Introduction

The Kentucky Board of Education instituted academic standards to implement the learning goals of the Kentucky Education Reform Act. Schools are accountable for their students' learning of the academic content based on these standards, which is assessed using the Kentucky Core Content Test.

In 1990, the General Assembly enacted House Bill 940, known as the Kentucky Education Reform Act (KERA). The requirements of KERA most directly relevant to the assessment and accountability system are that schools develop their students' ability to

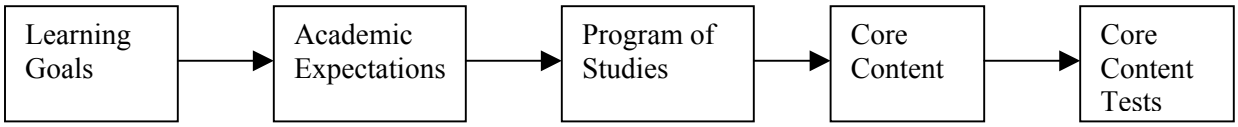
1. use basic communication and mathematics skills for purposes and situations they will encounter throughout their lives;
2. apply core concepts and principles from mathematics, the sciences, the arts, the humanities, social studies, and practical living studies to situations they will encounter throughout their lives;
3. become self-sufficient individuals of good character exhibiting the qualities of altruism, citizenship, courtesy, honesty, human worth, justice, knowledge, respect, responsibility, and self-discipline;
4. become responsible members of a family, work group, or community, including demonstrating effectiveness in community service;
5. think and solve problems in school situations and in a variety of situations they will encounter in life; and
6. connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned and build on past learning experiences to acquire new information through various media sources (KRS 158.6451).

The Kentucky Board of Education has the responsibility of establishing appropriate academic standards to implement the goals above. According to KRS 158.6453, the assessment program does not measure students' progress in meeting goals 3 and 4 from the list above.

As shown in Figure 1.A, the context for the Kentucky Core Content Test, which comprises the bulk of the Commonwealth Accountability Testing System (CATS) assessment, begins with the goals above—generally referred to as the learning goals. To measure student achievement of the goals, the Kentucky Board of Education has created Academic Expectations for each goal. Based

on the Academic Expectations, the Program of Studies outlines "the minimum content required for all students before graduating from Kentucky high schools" (Commonwealth. Kentucky Dept. of Ed. *Program of Studies*).

Figure 1.A
The Path From Learning Goals to Core Content Tests



Finally, the Core Content for Assessment is the essential content drawn from the Program of Studies that all students should know and that will be covered on the state assessment. All the content found in the Program of Studies is supposed to be taught, but schools are held accountable through CATS only for teaching the core content, which varies by elementary, middle, and high school level.

An element assessed on CATS can be traced through this process. Academic Expectation 2.1 is that students "understand scientific ways of thinking and working and use those methods to solve real-life problems." Note that this expectation applies across the curriculum, not just in science classes.

The Program of Studies is more specific. An example is Scientific Inquiry—Intermediate Science (S-4-PS-2): students "will understand that materials can exist in different states and some common materials (e.g., water) can change states."

Here is an example of what is included in the Core Content for Assessment for middle school science: "Cells carry on the many functions needed to sustain life. They grow and divide, thereby producing more cells. This requires that they take in nutrients, which they use to provide energy for the work that cells do and to make the materials that a cell or an organism needs" (SC-M-3.1.3).

The Kentucky Instructional Results System (KIRIS) was established as the state assessment and accountability system in 1992 to measure progress toward the learning goals established under KERA. In 1994, a panel of measurement specialists was appointed by the Office of Education Accountability and the General Assembly to investigate the technical quality of KIRIS. According to the panel's report, the test frameworks did not

communicate clearly what students were expected to know, the test scores reported for schools were not adequately reliable for accountability purposes, the student performance standards lacked standardization, and writing portfolio scores were not reliable (Hambleton et al.). The Task Force on Public Education, established in 1996 to investigate the effectiveness of KIRIS, recommended changes in Kentucky's system of assessment and accountability.

In 1998, the General Assembly enacted HB 53, which established guidelines for replacing the existing accountability and assessment system with the Commonwealth Accountability Testing System (CATS).

In 1998, the General Assembly enacted House Bill 53, giving guidelines for the Kentucky Board of Education to replace KIRIS with CATS. CATS is designed to ensure school accountability for student achievement, but among the components to be included is a "technically sound longitudinal comparison of the assessment results for the same students." The legislation also required that teachers play a significant role in the design of the new assessments. Appendix A summarizes relevant Kentucky statutes and regulations.

Description of This Study

How This Study Was Conducted

The 2004 General Assembly adopted Senate Joint Resolution 156 directing the Office of Education Accountability to complete a review of the Commonwealth Accountability Testing System. In conducting this study, staff of OEA and Legislative Research Commission (LRC) reviewed relevant state and federal documentation. In addition, the Office of Education Accountability contracted the services of AEL, a nonprofit educational research and development firm based in West Virginia, to review the research literature on the reliability and validity of education assessments systems. AEL was also contracted to conduct focus groups of parents, students, teachers, principals, superintendents, and school board members. The Office of Education Accountability and LRC staff developed surveys based on the information obtained from the focus groups. AEL administered the surveys. The Office of Education Accountability and LRC staff surveyed officials from school districts and schools to determine the costs that were attributed to CATS.

Organization of the Report

The remainder of Chapter 1 provides a brief summary of the conclusions and describes how CATS is administered and applied to schools.

Chapter 2 discusses the changes and additions that have been made in order to comply with the requirements of the federal No Child Left Behind Act. The chapter also discusses the requirements that have not been fully addressed.

Chapter 3 discusses the issues of reliability and validity for educational assessment systems, such as CATS. This chapter also summarizes issues with how CATS scores might be used at the school and student levels.

Chapter 4 summarizes the results of surveys sent to teachers, principals, superintendents, students, parents and guardians, and school board members regarding CATS.

Chapter 5 describes the costs associated with CATS. These costs include both state and local costs that schools and school districts attribute to CATS.

Major Conclusions

The study's major conclusions are as follows:

The Kentucky Board of Education and the Kentucky Department of Education have modified CATS to comply with the requirements of the No Child Left Behind Act. Kentucky's state accountability plan, which details many of the modifications, has been approved by the United States Department of Education. Kentucky recently submitted additional modifications to demonstrate how Kentucky will set educational standards and assess student and school performance. The U.S. Department of Education will be reviewing these changes, but this review has not yet been scheduled.

The research on education assessments does not provide a clear indication of what level of reliability and validity must be achieved on an assessment. Instead, the research suggests that the needed level of reliability and validity varies depending on how the scores will be used. The level of reliability and validity needed for a specific use will depend on the tolerance for making an incorrect decision.

The cost of CATS is borne at both the state and local levels. State-level costs for FY 2004 were estimated to be $20.8 million. Local-level expenditures, which were incurred by districts and schools, were estimated to be $16.4 million. There was considerable variation in the types and amounts of costs reported at the local level. This variation may reflect differences in the actual amounts spent on CATS, differences in how various costs were classified, and differences in whether certain types of costs were attributed to CATS. In total, the state and local costs for CATS was estimated to be $37.2 million: or $78 per assessed student.

Overview of the Commonwealth Accountability Testing System

CATS consists of various types of assessments that evaluate the performance of students within schools. The scores on these assessments are used to determine whether a school is making sufficient progress toward its educational goals. The following section describes the groups that provide advice on the design of CATS, the assessments that are administered to students, and the accountability provisions that are applied to schools.

Advisory Groups

House Bill 53 created roles for groups to advise policy makers and education officials on issues related to CATS.

House Bill 53 also created roles for four groups to advise policy makers and education officials on issues related to CATS. The General Assembly's Education Assessment and Accountability Review Subcommittee is empowered to review administrative regulations, advise the Kentucky Board of Education on the implementation of the state's system of assessment and accountability, and advise and monitor the Office of Education Accountability. The Office of Education Accountability is to advise the Legislative Research Commission and the Kentucky Board of Education on CATS.

The National Technical Advisory Panel on Assessment and Accountability (NTAPAA) meets quarterly to advise LRC and, with the approval of the LRC director, the Kentucky Board of Education and the Kentucky Department of Education. LRC appoints the members: professionals with expertise in education testing and measurement. Currently, the panel is composed of six professors from the University of California at Los Angeles, University of Colorado, University of Florida, University of Kansas, University of Pittsburgh, and Vanderbilt University. HB 53 stipulates that if NTAPAA specifies that any student assessment

scores are valid and reliable at the individual level, they are to be included on student transcripts.

The fourth formal advisory group is the School Curriculum, Assessment, and Accountability Council, which is charged with studying, reviewing, and making recommendations on academic standards, assessment of learning, school accountability, and ways to help schools improve. Pursuant to KRS 158.6452, the governor appoints the 17 members. The membership consists of teachers, principals, superintendents, school board members, school district assessment coordinators, parents, employers, and university professors with backgrounds in education assessment and measurement.

CATS Assessments

Under CATS, there are two types of tests administered to students, each of which covers multiple subjects. The Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5) is a nationally norm-referenced test. The Kentucky Core Content Test (KCCT) is criterion referenced. The performance of students on the core content tests makes up the majority of a school's academic index.

Under CATS, there are two types of assessments administered to students: the Kentucky Core Content Test (KCCT) and the Comprehensive Test of Basic Skills, Fifth Edition (CTBS/5). CTBS/5 is the nationally norm-referenced test that assesses students in reading, math, and language arts in grades 3, 6, and 9. KCCT is the criterion-referenced assessment administered to students in grades 4, 5, 7, 8, 10, 11, and 12. KCCT assessments cover reading, math, science, social studies, arts and humanities, practical living and vocational studies, and writing. Disabled students for whom KCCT is inapprorpriate may be assessed based on portfolios related to their work. Table 1.1 indicates which tests are given by grade level.

CTBS/5

The CTBS/5, also referred to as Terra Nova, is designed to assess student achievement compared to the performance of other students on a nationally norm-referenced assessment. Specifically, the CTBS/5 test measures how Kentucky's students are progressing compared to a group established as the norm in 1996. Comparisons of percentile rankings over time are also possible with other states that are using the same form of the test. Each school's students' performance on the CTBS/5 counts for 5 percent of the index through which the school is held accountable. The CTBS/5 is developed by CTB McGraw-Hill and is commercially available to all schools. The CTBS/5 tests, which consist entirely of multiple-choice questions, are not designed specifically to cover Kentucky's core content.

Table 1.1
CATS Assessments

	Kentucky Core Content Test								Writing Portfolio	CTBS/5
	(6 open-response, 24 multiple-choice questions each)				(2 open-response, 8 multiple-choice questions each)					(Multiple-choice questions in math, reading, language arts)
Grade	Reading	Math	Science	Social Studies	Arts and Humanities	Practical Living & Vocational Studies	On-demand Writing	Alternate Portfolio		
3										●
4	●		●				●	●	●	
5		●		●	●	●				
6										●
7	●		●				●		●	
8		●		●	●	●		●		
9										●
10	●					●				
11		●	●	●	●					
12							●	●	●	

Source: Commonwealth. Kentucky Dept. of Ed. *2004 CATS Interpretative Guide* 17.

Kentucky Core Content Test

Based on their performances on each of the assessments that cover the core content, students are divided into four categories: novice, apprentice, proficient, and distinguished.

A criterion-referenced, or standards-based, test such as KCCT assesses how well students are doing relative to a predetermined performance level on a specified set of educational goals or outcomes. The purpose of KCCT is to measure how well schools are educating students to master the core content. The performance of students on KCCT makes up the majority of a school's accountability index. Based on their performances on each of the assessments that cover the core content, students are divided into four categories:

1) Novice
   - Student demonstrates minimal, limited, underdeveloped, and at times inaccurate content knowledge and reasoning.
   - Student communication is ineffective and lacks detail with no evidence of connections within or between content areas.
   - Student uses strategies that are inappropriate.

2) Apprentice
   - Student demonstrates some basic content knowledge and reasoning ability.
   - Student communicates reasonably well but draws weak conclusions or only partially solves or describes.
   - Student attempts appropriate strategies with limited success.
3) Proficient
   - Student demonstrates broad content knowledge and is able to apply it.
   - Student communication is accurate, clear, and organized with relevant details and evidence.
   - Student uses appropriate strategies to solve problems and make decisions.
   - Student demonstrates effective use of critical thinking skills.
4) Distinguished
   - Student demonstrates an in-depth, extensive, or comprehensive knowledge of content.
   - Student communication is complex, concise, and sophisticated with thorough support, explicit examples, evaluations, and justifications.
   - Student uses and consistently implements a variety of appropriate strategies.
   - Student demonstrates insightful connections and reasoning (Commonwealth. Kentucky Dept. of Ed. *2004 CATS Interpretative Guide* 92-93).

There are more detailed descriptions for each grade level and content area. The performance levels of novice and apprentice are further divided for reading, mathematics, science, and social studies. Novice is divided into non-performance, medium, and high. Apprentice is divided into low, medium, and high.

There are seven KCCT subject tests: reading, math, science, social studies, arts and humanities, practical living and vocational skills, and on-demand writing.

Subject Tests. Kentucky uses multiple-choice and open-response questions in accordance with KRS 158.6453. Each item type seeks to identify a student's level of thinking and knowledge. The mix of question types for six KCCT subject tests varies by subject. For reading, math, science, and social studies tests, students are assessed based on 24 multiple-choice questions and 6 open-response questions on six forms of each test. For arts and humanities and practical living and vocational skills, students are assessed based on 8 multiple-choice and 2 open-response questions on 12 forms of each test. KCCT also includes one open-response and four multiple-choice questions per content area that do not count for accountability but are used for developing future tests.

Multiple-choice questions are those in which four answers are offered as alternative responses for each item. Multiple-choice questions ask for the best answer to a question or an incomplete statement. Answers are scored as either 0 (incorrect) or 1 (correct).

Example: Grade 7 science multiple-choice question
An earthquake occurs when the tectonic plates below Earth's surface suddenly shift. These shifts of the tectonic plates are caused by: O movements in Earth's core. O movements in Earth's mantle. O deposition of sediments. O eruption of volcanoes.

Source: Commonwealth. Kentucky Dept of Ed. *Grade 7 Sample Released Questions* 26.

An open-response question requires students to answer questions in writing. Answers are scored using a 0 to 4 scale. Open-response questions differ in how much freedom is permitted to the student in making a response. Restricted responses require brief and precise answers to specific questions. Extended responses reflect more comprehensive questions that allow greater freedom in structuring a response. The KCCT uses Extended Constructed Response.

Example: Grade 7 science open-response question
Scientists have evidence that the landforms we see on Earth, such as mountains, islands, and canyons, as well as the shapes of continents, are the result of constructive and destructive forces at work over a long period of time. Describe in detail **two** pieces of evidence that show that landforms on Earth are constantly changing. Provide a specific example for each piece of evidence.

Source: Commonwealth. Kentucky Dept. of Ed. *Grade 7 Sample Released Questions* 30.

For students in grades 4, 7, and 12, there is an on-demand writing section of KCCT. Students are given two writing prompts and must respond to one prompt of their choosing within a time limit of 90 minutes.

For students in grades 4, 7, and 12, there is a section of KCCT in which they are assessed based on their writing. Students are given two writing prompts and must respond to one prompt of their choosing within a time limit of 90 minutes (students may be allowed extra time as long as they are being productive). Students may be asked to respond in the form of a letter or article (grades 4, 7, 12), editorial (grades 7, 12), or speech (grade 12).

Example: Grade 12 on-demand writing prompt
Situation: From a very young age we are taught to follow the rules at home, then at school, and, later, on the job. Think of a time when you needed to know the rules. What happened? Why were the rules important? Could someone else learn from your experience? Writing Task: Write an editorial for your school newspaper about the importance of rules. Support your idea by writing about a time when knowing the rules proved to be important.

Source: Commonwealth. Kentucky Dept. of Ed. *Grade 12 Sample Released Writing Tasks* 4.

On-demand writing is scored based on "the degree to which the writer

- maintains a focused purpose to communicate with an audience…;
- develops and supports main ideas and deepens the audience's understanding…;
- creates unity and coherence to accomplish the focused purpose…;
- creates effective sentences…;
- demonstrates [appropriate] word choice…, concise use of language…, and correct usage/grammar…; and
- demonstrates correct spelling, correct punctuation, and correct capitalization" (Commonwealth. Kentucky Dept. of Ed. *Sample Released Questions* 67).

The writing portfolio is a compilation of students' best writings, produced over time, with at least one piece from each of the four specified categories of writing. As part of CATS, students' writing portfolios are assessed in grades 4, 7, and 12.

Writing Portfolios. As part of CATS, students' writing portfolios are assessed in grades 4, 7, and 12. The writing portfolio is a compilation of each student's best writings, produced over time, with at least one piece from each of the following broad categories of writing:

- Reflective writing is in the form of a letter to the reviewer of the portfolio that discusses the student's growth as a writer and reflects on pieces in the portfolio.
- Personal expressive writing can take the form of a personal narrative that focuses on a single event in the student's life (grades 4, 7, 12); a memoir focusing on the significance of the student's relationship with a particular person, place, animal, or thing, supported by memories of specific experiences (grades 4, 7, 12); or a personal essay focusing on a central idea about the student or the student's life, supported by a variety of incidents (grades 7, 12).

- Literary writing can be a short story, poem, script (grades 4, 7, 12), or play (grades 7, 12).
- Transactive writing is produced "to get something done" in the real world. Acceptable forms include a letter to a newspaper, a newspaper editorial, a magazine article, or a speech.

The number of portfolio entries varies by grade. For 4th-grade students, four writing pieces are required for the portfolio, one in each category. In grades 7 and 12, five writing pieces are required, one in each of the four categories plus an extra entry in either personal expressive, literary, or transactive.

Answers to the open-response questions and on-demand writing prompts are scored holistically. This means that there may be specific elements to be evaluated, but there is one overall score for each response. Writing portfolios are also scored holistically: there is one score for the entire portfolio.

Scoring of Assessments. Answers to the open-response questions and responses to the on-demand writing prompts are scored holistically. Holistic scoring means that although there may be specific elements to be evaluated, there is one overall score for each answer. The alternative is analytic scoring in which an overall score is the sum of weighted scores for specific tasks. The KCCT uses a unique scoring rubric for each open-response question. On-demand writing is scored using the same scoring guide as the writing portfolio.

Writing portfolios, scored by trained teachers, are scored using the performance categories of novice, apprentice, proficient, and distinguished. The rater reads the whole portfolio and assigns one score based on the overall performance of the student. The scores of student writing portfolios are included in calculation of the school academic index.

Students with moderate to severe disabilities for whom traditional assessments are inappropriate are assessed via the Kentucky Alternate Portfolio. The portfolio is submitted and assessed during grades 4, 8, and 12.

Alternate Portfolio. Students with moderate to severe disabilities for whom traditional assessments are inappropriate are assessed via the Kentucky Alternate Portfolio. Each student's individualized education plan must indicate this method of assessment and declare that the student is not on the regular diploma track. The alternate portfolio is submitted and assessed during grades 4, 8, and 12. The portfolio measures individual student progress toward the Academic Expectations. The portfolio is also scored based on five other dimensions, including progress based on objectives in the student's individualized education plan. In 2004, there were 1,258 students statewide assessed using alternate portfolios, an increase of approximately 100 students from 2003 (Commonwealth. Kentucky Dept. of Ed. *Alternate Portfolio 2004* and *Alternate Portfolio 2003*). In calculating the school index, the score obtained by a student on an alternative portfolio (novice, apprentice, proficient, or distinguished) is applied to each subject area measured by KCCT for that student.

Administration of the Assessments

KCCT is administered using six different test forms to provide greater coverage of the core content. As a result, students completing different forms receive different questions. If necessary, scores are adjusted to account for differences in difficulty of the forms.

KCCT is administered using different test forms to provide greater coverage of the core content. As a result, students completing different forms receive different questions, which can result in differences in difficulty. Test forms are designed to minimize these differences. Some differences can persist, so scores are adjusted. For example, two students who receive the same raw score on two different test forms could ultimately receive somewhat different scores for use in calculating their school's academic index. These adjusted scores are referred to as scale scores. Scale scores adjust for small differences in difficulty between test forms. Two students of the same ability who get the same scale score could get different raw scores on different KCCT forms.

A contractor scores the CTBS/5 tests, KCCT subject tests, and on-demand writing prompts. Methods to assure accuracy are integrated into the scoring processes. Local teachers score portfolios. Each year, KDE selects schools and audits the scoring of each school's student portfolios.

State regulations establish procedures for including special student populations, such as disabled students or students with limited English proficiency, in the state assessment program.

Accommodations and Modifications. Procedures for including special student populations in the state assessment program are established in 703 KAR 5:070. The regulation identifies several populations and outlines how the populations are included in state assessment and accountability programs. Special populations addressed include students with disabilities; students in vocational-technical, special education, preschool, alternative, and non-district-operated programs; students for whom English is not their primary language; homebound students; and students with temporary medical conditions requiring accommodations, modifications, or both.

The regulation establishes under what conditions students may receive an accommodation, modification, or both. An accommodation is a change in the testing environment or process. An example would be allowing extra time. A modification is a change in the instrument used for assessment, a large-print version, for example.

The accommodation or modification for each student must be included in the student's Individual Education Plan, 504 Plan (based on Section 504 of the U.S. Rehabilitation Act of 1973), or Program Services Plan for students with limited English proficiency. The accommodation or modification must have been used during the whole school year. For example, if a student's

Individual Education Plan allows a scribe (writer) as part of regular instruction, the student may have a scribe during assessment. Other permissible accommodations or modifications include "reading text in English…, use of technology, [and] use of extended time" (Commonwealth. Kentucky Dept. of Ed. *Kentucky Core Content Tests 2002* 1-3).

Table 1.2 shows the distribution of disabled students and students with limited English proficiency (LEP) who took KCCT tests with accommodations or modifications in spring 2004. Less than 1 percent of LEP students take KCCT tests with accommodations or modifications. However, more than 11 percent of 4th and 5th graders taking tests did so using accommodations or modifications based on disability; and more than 9 percent of 7th and 8th graders, approximately 7 percent of 10th and 11th graders, and 6 percent of 12th graders did so.

Table 1.2
Disabled and LEP Students Taking KCCT Tests
With Accommodations or Modifications
(Spring 2004)

	Disabled		Limited English Proficiency	
Grade	Number	As % of All Students Taking Tests	Number	As % of All Students Taking Tests
4	5335	11.1	344	0.7
5	5664	11.6	333	0.7
7	4944	9.6	203	0.4
8	4629	9.2	236	0.5
10	3281	7.2	224	0.5
11	2830	7.0	205	0.5
12	2155	5.7	137	0.4

Note: For grades 4, 5, 7, 8, 10, and 11, numbers are for the subject tests. The numbers of students with accommodations for on-demand writing and writing portfolios are slightly lower. There are no grade 12 subject tests; the numbers shown are the averages of numbers of students with accommodations for on-demand writing and writing portfolios.
Source: Commonwealth. Kentucky Dept. of Ed. *Spring 2004*.

Exemptions to assessment are allowed for specified reasons such as medical incapacity or limited English proficiency. Approximately 0.6 percent of students were exempted in 2004.

Exemptions. Foreign exchange students are exempted from CATS assessments. Other students may be exempted if they are determined to be medically unable to participate. Schools also have the option of exempting from accountability calculation students with limited English proficiency in their first year ever in a U.S. school. Based on the data in the Spring 2004 Kentucky Performance Report, approximately 0.6 percent of students were exempted that year. The numbers varied by grade level and type of

assessment. Generally, fewer students were exempted for portfolios than for other types of assessment. Among these other types of assessment, the percentage exempted ranged from 0.3 percent for 5th graders to 1 percent for 12th graders (Commonwealth. Kentucky Dept. of Ed. *Spring 2004*).

School Accountability

Schools and school districts are held accountable based on students' performance on the assessments and relevant nonacademic measures such as attendance or graduation rates.

Schools and school districts are held accountable based on students' performance on the various types of assessments and relevant nonacademic measures such as attendance or graduation rates. Schools that meet or exceed established goals are eligible for financial rewards. Those not meeting goals are subject to consequences.

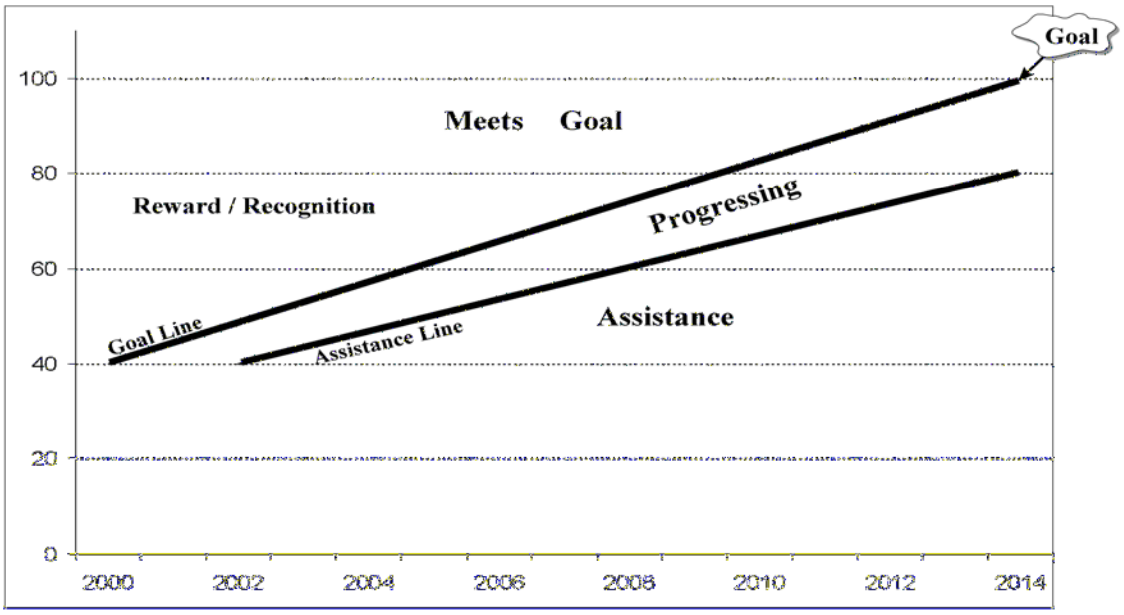
The long-term goal for each school and district is to reach an accountability index score of 100 or greater on a 140-point scale by 2014.

The long-term goal for each school and district, and for the state as well, is to reach proficient by 2014. Proficient is defined as a score of 100 or greater on a 140-point scale. Intermediate goals are based on the school's progress toward scoring 100 or greater in 2014 from its own starting point. Each school and district is evaluated every two years to determine whether it has achieved the appropriate level of progress.

Each school and district is evaluated every two years to determine whether it has achieved the appropriate level of progress.

Each school's growth chart is formulated as if the school would reach the long-term goal of 100 in equal steps, with each step taking two school years. Reaching or surpassing the goal of 100 earlier than 2014 would be better, obviously. A school with a starting point near 100 in 2000 has smaller biennial steps. Schools that began with baselines further below 100 must attain greater biennial improvements. Figure 1.B is an example of what a school's growth chart might look like.

Figure 1.B
Example of a School Growth Chart



Source: Constructed by LRC staff based on 703 KAR 5:020, Section 8.

The Kentucky Board of Education is responsible for the details of Kentucky's accountability system, including the process for calculating biennial goals for schools. The biennial school goals were calculated as follows:

1. Each school's 1999 and 2000 test results were averaged to determine its starting point or baseline.
2. The baseline was subtracted from 100 to determine how much the school needed to improve to reach the long-term goal of 100 in 2014.
3. There are seven two-year periods between 2000 and 2014, so the amount the school must improve was divided by 7 to determine how much improvement is needed every two years.
4. A standard error of measurement is used in conjunction with each two-year goal to reflect the test's margin of error. The margin of error accounts for fluctuations in scores due to chance occurrences that would affect scores. The margin of error varies by level of school and the number of students in a school. KDE, using 1999 and 2000 test data, calculated each school's margin of error, which ranges from approximately 0.5 to 3.0.

An assistance line was calculated using similar procedures, beginning at the baseline in 2002 and ending with 80 in 2014. Each school that is at or above its goal line is classified as meets goal. A school that is below its goal line but at or above the assistance line is classified as progressing. Schools with an index below the

assistance line are eligible for assistance from the state and the school district. For each classification, the determination of a school's status takes its margin of error into account. Table 1.3 shows the numbers and percentages of schools for each classification for the 2002 and 2004 accountability cycles.

Table 1.3
Schools Classified as In Need of Assistance, Progressing, and Meets Goals
(2000-2002 and 2002-2004 Accountability Cycles)

School Classification	Accountability Cycle			
	2000-2002		2002-2004	
In Need of Assistance	88	7.4%	48	4.1%
Progressing (Decline)	69	5.8%	48	4.1%
Progressing	439	37.1%	419	35.6%
Meets Goal	586	49.6%	661	56.2%
Total	1182	100.0%	1176	100.0%

Source: Commonwealth. Kentucky Dept. of Ed. *Kentucky Core Content Test Scores Released.*

Based on its accountability index, a school is classified as meets goal (on track to the target of proficiency), progressing (not on track but above the school's designated "assistance line"), or in need of assistance (not on track and below the assistance line).

It should be noted that a school can be classified as progressing even though its accountability index score declined from the previous cycle. This distinction is made in the table. Compared to the 2002 cycle, there were fewer progressing schools with declining scores in 2004. From 2002 to 2004, the number of schools classified as in need of assistance declined from 88 to 48. The number of schools meeting goal increased from just under half to approximately 56 percent (Commonwealth. Kentucky Dept. of Ed. *Kentucky Core Content Test Scores Released*).

Rewards and Assistance

Schools meeting goals or progressing qualify for rewards if they also meet the criteria for reducing novice classification and dropout rates. Assistance schools qualify for assistance and face consequences in some cases.

Schools meeting goals or progressing may be eligible for rewards, which also depend on other factors. Schools performing below the assistance line qualify for assistance.

Rewards. In order to qualify for rewards for growth, a school's biennial accountability index must be at or above its goal line. Schools also qualify for rewards for meeting the designated index points 55, 66, 77, 88, and 100. To receive any award, a school must also reduce its dropout rate and reduce the number of students in the school classified as novice. The dropout rate must be reduced to 5.3 percent or be reduced by at least 0.5 percent from the dropout rate of the previous biennium and be less than 6 percent. The number of students classified as novice must be

reduced steadily so that by 2014 no more than 5 percent of the school's students are classified as novice.

No funding was authorized for rewards beginning with the 2002-2004 biennium.

No funding was authorized for rewards beginning with the 2002-2004 biennium. However, if rewards had been available, some schools that otherwise could have qualified would have been ineligible due to the novice and dropout requirements. Of the 467 schools classified as progressing, 56 percent did not meet the dropout/novice criteria. Of the 661 schools that met goals, only 11 did not meet the dropout and novice criteria. For schools that were progressing or meeting goals, the share of schools disqualified for dropout or novice status was lower in 2002-2004 than in the previous accountability cycle (Commonwealth. Kentucky Dept. of Ed. *Kentucky Core Content Test Scores Released*).

Schools with accountability indices below their assistance lines are ranked according to their indices and grouped into three categories.

Assistance and Review. Schools with accountability indices below their assistance lines are ranked according to their indices and grouped into three categories. Schools in all three categories are eligible for financial assistance from the state. Level 1 schools (the top third) are required to perform a self-review facilitated by the district's professional development coordinator. Level 2 schools (middle third) are reviewed by a team established by the Kentucky Department of Education (KDE) that must include members from the local school district. An external team coordinated by KDE reviews Level 3 schools (lower third). The Level 3 review can include recommendations that certain staff be dismissed or transferred. Level 3 schools will also receive guidance from a highly skilled educator—a teacher and administrator identified by the commissioner of education and an advisory committee as exhibiting strong organizational, curricular, and interpersonal skills. If a school is classified as Level 3 two biennia in a row, the school district must allow students enrolled at the school to transfer to another school with an accountability index above its assistance line, within the district or outside it if the district has an agreement with another district.

Calculation of Accountability Index

A school's accountability index is based on the assessment scores of its students and on nonacademic measures such as attendance and dropout rates. Each student's performance in a particular content area is scored and assigned to a performance level based on the scale score. For example, a scale score of 560 in elementary school science is proficient (Commonwealth. Kentucky Dept. of Ed. *Kentucky Core Content Tests 2002* 8-5).

The school's academic index for a particular content area is determined by the percentage of students who score at the various performance levels. Each performance level is associated with a weight. The weight for proficient is 100. The weight for each level is multiplied by the percentage of students scoring within the level to develop a weighted score. These weighted scores are totaled to develop the academic index for the content area. An example of how individual student scores within a content area are compiled to determine the school's academic index for the subject is shown in Table 1.4.

Table 1.4
Sample Calculation of a School's Academic Index for 4th-grade Reading

Performance Level	Weight	Distribution of Student Scores	Calculation	Weighted Score (Weight x Percent)
Novice Non-performance	0	5%	0 x .05	0
Novice Medium	13	10%	13 x .10	1.3
Novice High	26	15%	26 x .15	3.9
Apprentice Low	40	20%	40 x .20	8.0
Apprentice Middle	60	25%	60 x .25	15.0
Apprentice High	80	15%	80 x .15	12.0
Proficient	100	8%	100 x .08	8.0
Distinguished	140	2%	140 x .02	2.8
Academic Index				51.0

Source: Commonwealth. Kentucky Dept. of Ed. *2004 CATS Interpretive Guide* 23.

The academic indices for the different KCCT content areas are combined with nonacademic factors and the results of CTBS/5 tests to construct the school's overall accountability index. As shown in Table 1.5, the weights applied for the components vary by level of school. Dropout rates are only included in the accountability indices of middle and high schools. Measures of transition to adult life are only included for high schools.[1]

[1] Indicators of a student's successful transition to adult life include full-time college, vocational or technical school, military, work, and a combination of work/school. The measures are obtained each fall by surveys done by school personnel of the previous year's graduates (Commonwealth. Kentucky Dept. of Ed. *2004-2005 Nonacademic Data* 11). According to KDE, the percentage of students making a transition to adult life increased from 92.6 percent in 1993 to 96.0 percent in 2003 (Commonwealth. Kentucky Dept. of Ed. *Nonacademic Data* 13).

A school's performance is evaluated by biennium, so two years of the accountability index are averaged to produce one value for each accountability cycle. For example, schools will be evaluated in 2006 based on the average of their accountability indices for the school years 2004-2005 and 2005-2006.

Table 1.5
Academic and Nonacademic Weights in the Accountability Index

Component	Grade Level		
	Elementary	Middle	High
Reading	19.00%	14.25%	14.25%
Mathematics	19.00	14.25	14.25
Science	14.25	14.25	14.25
Social Studies	14.25	14.25	14.25
On-demand Writing	2.85	2.85	2.85
Writing Portfolio	11.40	11.40	11.40
Arts and Humanities	4.75	7.125	7.125
Practical Living and Vocational Studies	4.75	7.125	7.125
Attendance Rate	3.80	3.80	1.90
Retention Rate	0.95	3.80	0.48
Dropout Rate	—	1.90	3.56
Transition to Adult Life	—	—	3.56
CTBS/5	5.00	5.00	5.00
Total	100%	100%	100%

Source: 703 KAR 5:020.

By statute, parents or guardians of each student must receive a report card on the performance of the student's school. The minimum content of the report is specified. Individual reports on students are also sent to each school district, but districts have discretion as to whether and how to send the reports to parents or guardians.

Reporting of Results. By statute, parents or guardians of each student must receive a report card on the performance of the student's school.[2] At a minimum, the reports are to include student academic achievement, including the results of assessments; nonacademic achievement, including attendance, retention, and transition to adult life; and learning environment, including involvement of parents and guardians. For subgroups of 10 or more students, the results are to be reported by race, gender, and disability.

Each year, districts receive an individual report on each student's performance on CATS assessments. KRS 160.345(3)(b) requires a local school board policy on assessment of individual student

[2] By statute, a district report card is to be published in the newspaper with the largest circulation in the county. Language in the enacted budget for FY 2006 allows districts the alternative of publishing the information on their Web sites or having a printed copy available at a public library within the school district. If either of these methods is chosen, notification must be given in the newspaper with the largest circulation in the county.

progress, including testing and reporting student progress to students and parents or guardians. However, there is broad discretion at the district level as to the form and substance of individual student progress reports. This means that districts are not required to distribute the student-level CATS reports to parents or guardians. For districts that do distribute the reports, the means to do so vary.

Chapter 2

Alignment of the Commonwealth Accountability Testing System With the Requirements of the No Child Left Behind Act

The No Child Left Behind Act (NCLB) requires states to establish education standards and assessments.

In 2002, the federal government adopted the No Child Left Behind Act (NCLB), which mandates that states establish education standards and assessments. As a result of this legislation, each state must set annual goals for schools and students and assess their progress. The ultimate goal of the legislation is to have schools and students in each state reach a predefined level of proficiency by the 2013-2014 school year.

Twelve years earlier, the Kentucky General Assembly passed the Kentucky Educational Reform Act. This Act also established performance goals and an assessment mechanism for Kentucky's schools. Over time, various changes were made to Kentucky's accountability system. One of these changes established the Commonwealth Accountability Testing System (CATS).

Kentucky's current accountability system, CATS, does not meet all of the NCLB requirements.

The educational standards and assessments that evolved in Kentucky are similar in many ways to the provisions that were later mandated in NCLB. Although NCLB provides each state with some flexibility in designing its standards and assessment programs, many of the provisions required in NCLB are somewhat specific. As a result, CATS does not always satisfy the requirements of NCLB.

The Kentucky Board of Education is currently augmenting CATS in order to comply with the requirements of NCLB.

The Kentucky Board of Education is responsible for implementing both CATS and NCLB within the structure of state and federal laws. At its meeting on August 7, 2003, the board decided to retain CATS unchanged and incorporate the additional requirements of NCLB. CATS and NCLB will operate concurrently, and the board is currently augmenting CATS in order to satisfy the requirements of both systems.

This chapter summarizes the changes and additions that have been made to CATS in order to comply with NCLB and discusses the requirements that have not yet been fully addressed. The information presented below comes primarily from various reports provided by the Kentucky Department of Education (KDE), the Kentucky Board of Education, and the United States Department

of Education (USDOE). The changes that Kentucky Board of Education has made or plans to make to comply with NCLB are detailed in the "Kentucky's Consolidated State Application Revised Accountability Workbook."

NCLB Approval Process

NCLB requires every state that applies for federal Title I education grants to submit a plan that satisfies all of NCLB's requirements.[1] The USDOE is responsible for approving the state plans. The approval process consists of two phases: approval of the states' accountability plans and approval of the states' standards and assessment plans.

The United States Department of Education (USDOE) has approved the initial NCLB provisions in Kentucky's and 27 other states' accountability plans.

State accountability plans lay out the basic provisions of how each state intends to comply with the requirements of NCLB. These requirements include methods for determining adequate progress toward goals for schools, levels of student participation, annual determinations of schools' progress, and various other provisions. The Government Accountability Office reports that the USDOE approved Kentucky's and 27 other states' accountability plans as of July 31, 2004. The remaining 23 states and the District of Columbia have accountability plans that have been approved with conditions.

The USDOE will be reviewing each state's educational standards and assessments plans. These plans detail the knowledge and skills that students should possess and assessments designed to measure students' knowledge and skills.

The second phase of the approval process requires states to submit a standards and assessment plan to the USDOE. States must develop standards that students and schools will be held accountable for achieving. These standards detail the knowledge and skills that students are expected to possess. States must also develop assessments that measure students' mastery of the state's academic standards. States must show that they have produced valid assessments to measure student performance. The standards and assessment plan also lays out consequences and rewards based on students' performance.

A group of national education experts, under the guidance of USDOE, will determine whether states are in compliance with NCLB requirements based on evidence submitted by the states. States must administer an assessment system that meets the NCLB requirements by the 2005-2006 school year (Simon). Kentucky has partially, but not fully, met all of the standards and assessment provisions. KDE is currently working on these provisions.

[1] The Title I program was developed to assist schools with high percentage of low-income students (United States. Dept. of Ed. Title I).

Implementation of NCLB

Kentucky will use the same definition of proficiency for NCLB as it does for CATS. There will, however, be several differences between the two systems.

Both NCLB and CATS share the long-term goal of student proficiency by 2014. NCLB allows each state to define what it considers "proficient" for use within the state. Kentucky uses the same definition of proficient for NCLB as it does for CATS. Proficiency denotes a certain level of performance, or score, on the various assessments. The level required for proficiency varies across the different types of assessments. While both testing systems have the same long-term goal of proficiency, there are some important distinctions in how to obtain that goal. Some key differences between the two testing systems are noted in Table 2.1.

Table 2.1
Key Differences Between NCLB and CATS

NCLB	CATS
Only reading and mathematics results are used in accountability decisions. Science will be added in 2007 but will not be included in the determination of adequate yearly progress.	Results for seven core content subjects are used in accountability decisions: (math, reading, science, social studies, writing, practical living and vocational studies, and arts and humanities).
All schools in the state have the same baseline based on whether they are elementary, middle, or high school.	Each school has its own baseline.
Baseline is set using 2002 CATS scores.	Baseline is set using the average of the 1999-2000 CATS scores.
The assessment is intended to measure the progress of groups of students and schools.	The assessment is intended to measure the progress of schools.
Annual accountability performance judgments are produced.	*Biennial* accountability performance judgments are produced.

Source: Compiled by staff.

Augmented Assessments

NCLB requires tests in different grades than CATS requires.

Both CATS and NCLB require various types of tests in various grade levels. While there is some overlap of the testing requirements, both NCLB and CATS include some tests that are not required in the other. The CATS tests cover more subjects than required by NCLB, such as arts and humanities. CATS does not test students, however, at all of the grade levels required by NCLB.

NCLB requires states to test students in reading and math in grades 3 through 8 and once in high school. States must administer these reading and math tests by the 2005-2006 school year. Currently, Kentucky administers two types of tests. The CTBS/5 is a test used to assess students in reading, math, and language arts in grades 3, 6, and 9. The CTBS/5 consists of multiple-choice questions and is norm referenced to allow for comparisons of Kentucky students to students in the rest of the nation. The second test is the Kentucky Core Content Test (KCCT) that is used to assess students in several subjects including reading and math. The KCCT tests cover reading in grades 4, 7, and 10 and math in grades 5, 8, and 11.

Additional math and reading tests will be added to CATS in order to comply with NCLB.

These tests left Kentucky without reading and math tests in some grades that are required by NCLB. At its February 2004 meeting, the Kentucky Board of Education decided to change the assessments in order to comply with NCLB requirements (Commonwealth. Kentucky Board of Ed. *Issues Brief — Improving CATS for Optimum Results*). These changes are shown in Table 2.2. According to KDE, CTBS/5 will be replaced with a new norm-referenced test called Terra Nova Form D. Terra Nova Form D does not provide adequate coverage of the reading and math components of the core content as required by NCLB for grades 3 through 8. Therefore, the reading and math portions of the Terra Nova Form D will be augmented in these grades. The KCCT assessments will continue. The Kentucky Board of Education adopted a proposal at its February meeting, however, to administer the on-demand writing components in grades 5, 8, and 12, rather than in grades 4, 7, and 12. This change was not to comply with NCLB but to spread the writing assessments over more grades.

Table 2.2
Changes in Assessments

● = Continued ✓ = Added x = Removed

Assessments	Grade 3	4	5	6	7	8	9	10	11	12
KCCT										
Reading	–	●	–	–	●	–	–	●	–	–
Math	–	–	●	–	–	●	–	–	●	–
Science	–	●	–	–	●	–	–	–	●	–
Social Studies	–	–	●	–	–	●	–	–	●	–
Writing	–	x	✓	–	x	✓	–	–	–	●
Arts & Humanities	–	–	●	–	–	●	–	–	●	–
Practical Living & Vocational Skills	–	–	●	–	–	●	–	●	–	–
Writing Portfolio	–	●	–	–	●	–	–	–	–	●
Alternative Portfolio	–	●	–	–	–	●	–	–	–	●
Terra Nova Form D (Norm-Referenced Test)*										
Reading / Language Arts	✓	–	✓	✓	–	✓	✓	–	–	–
Math	✓	✓	–	✓	✓	–	✓	–	–	–
CTBS/5 (Norm-References Test)	x	–	–	x	–	–	x	–	–	–

* Terra Nova Form D will be augmented for grades 3 through 8 in order to provide adequate coverage of the core content.

Source: Commonwealth. Kentucky Dept. of Ed. *2004 CATS Interpretive Guide* 17; Commonwealth. Kentucky Board of Ed. *Issues Brief – Improving CATS for Optimal Results.*

Adequate Yearly Progress

Under NCLB, schools must meet three objectives in order to achieve adequate yearly progress.

To achieve the goal of having all students being proficient by 2014, all public schools and districts are required under NCLB to demonstrate satisfactory improvement each year toward that goal. Based on criteria included in NCLB, the Kentucky Board of Education has established specific targets for adequate yearly progress in reading and math. Under NCLB, schools must meet the following three objectives to achieve adequate yearly progress:

1. an annual measurable objective,
2. student participation goals, and
3. other academic indicator goals.

Each of these objectives and Kentucky's approach to meeting them are discussed below.

Three separate sets of annual objectives were developed for elementary, middle, and high schools.

Objective 1: Annual Measurable Objective. Under NCLB, three separate baselines are calculated: one for elementary schools, one for middle schools, and one for high schools. All schools of a

particular level must make the same adequate yearly progress. The baselines are to be calculated using test scores from the 2001-2002 school year and can be based on either the "lowest-achieving group of students in the state" or "the school at the state's 20th percentile in terms of the proportion of students at proficient levels" (Education Commission of the States 5).

The baselines for the annual objectives were based on "the percentage of students at or above the proficient level who are in the school at the 20th percentile in the state" (703 KAR 5:020 (10)).

Under CATS, each school has its own baseline, which is the average of its 1999 and 2000 test scores. Because the CATS baselines used older data, they do not meet the NCLB requirements. Therefore, the Board of Education set the baselines for NCLB using CATS data on reading and math from the spring of 2002 (Commonwealth. Kentucky Dept. of Ed. "Kentucky's Consolidated"). The Kentucky Board of Education decided to use "the percentage of students at or above the proficient level who are in the school at the 20th percentile in the state" for setting the baseline (703 KAR 5:020 (10)). Table 2.3 shows the NCLB baselines for elementary, middle, and high schools along with the adequate yearly progress goals. The figures in the table represent the percentage of students who must score at the proficient level within each school. The goals for adequate yearly progress generally provide constant increments through the 2013-2014 school year. The exception is from 2004-2005 through 2006-2007 during which the goals do not increase. According to the "Issue Brief" for the board's October 2003 meeting, this is to allow "schools time to understand and to adjust to the requirements of NCLB" (5).

Table 2.3
Kentucky's NCLB Baselines and Annual Measurable Objectives
(Percentage of Students Within Each School
Who Should Score at the Proficient Level)

	Elementary		Middle		High	
School Year	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**
2003-04	47.27%	22.45%	45.60%	16.49%	19.26%	19.76%
2004-05	53.86	32.14	52.40	26.93	29.35	29.79
2005-06	53.86	32.14	52.40	26.93	29.35	29.79
2006-07	53.86	32.14	52.40	26.93	29.35	29.79
2007-08	60.45	41.84	59.20	37.37	39.45	39.82
2008-09	67.04	51.53	66.00	47.81	49.54	49.85
2009-10	73.64	61.23	72.80	58.25	59.63	59.88
2010-11	80.23	70.92	79.60	68.68	69.72	69.91
2011-12	86.82	80.61	86.40	79.12	79.82	79.94
2012-13	93.41	90.31	93.20	89.56	89.91	89.97
2013-14	100.00	100.00	100.00	100.00	100.00	100.00

Source: Commonwealth. Kentucky Dept. of Ed. "Kentucky's Consolidated State Application Revised Accountability Workbook."

Under NCLB, improvements in student scores only affect whether a school makes its adequate yearly progress if it results in more students scoring at or above the proficient level.

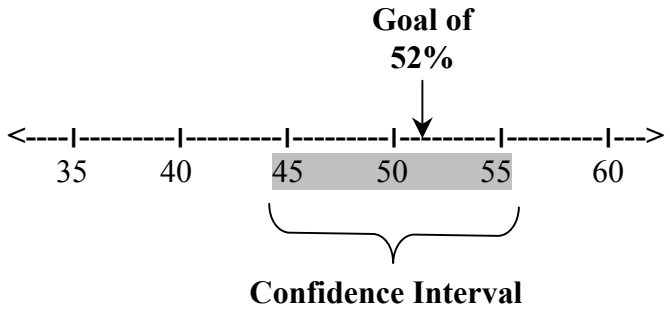
Schools and districts are held accountable for progress every year under NCLB. With CATS, assessments are administered each year, but the schools are held accountable on a biennial cycle. Additionally, NCLB and CATS focus on slightly different aspects of a school's performance in determining success or failure. NCLB identifies the primary indicator of success as the percentage of students who have attained proficiency. Since students are either proficient or not proficient, improvements in student scores that do not increase the number of students who are proficient will not directly help a school meet its NCLB goals. For example, a student who improves her score by 10 percentage points but does not reach the proficient level would not improve the school's chances of meeting adequate yearly progress. Under CATS, a school's accountability index reflects the school's scores in every performance category; therefore, an increase that resulted in more students scoring in higher categories, such as when students move from novice to apprentice, would increase the accountability index, even though there may have been little or no increase in the number of students who are proficient.

States may allow schools to use a confidence interval to determine whether schools meet their goals. This can result in a school achieving its adequate yearly progress if its score is statistically close to its goal.

States also have been given the flexibility to apply a statistical test, such as a confidence interval, to their adequate yearly progress calculations. Confidence intervals take into account the fact that there can be sampling error in test scores such as lower scores due to a student illness. Therefore, scores can be higher or lower simply due to circumstances. Sampling error is a more significant problem for small schools or for schools with a small number of students in subgroups. One absent student might not significantly change the overall results for a large school but might have a greater effect on the results for a small school. Confidence intervals compensate for this sampling error by "passing" schools or subgroups that come very close to achieving their annual goals. If the confidence interval associated with a school's score encompasses the goal, the school is considered as having met its adequate yearly progress.

For example, if the score on the reading test for a group of students was 50 percent, and the 99 percent confidence interval is applied, then it might be estimated that these students could have scored between 45 and 55 percent on this same assessment, were it given on another day under similar circumstances. Therefore, if the required score for adequate yearly progress was 52 percent, a score of 50 percent with a confidence interval of 45 to 55 percent would be considered meeting this goal because the confidence interval encompasses the required score of 52 percent. This is illustrated in Figure 2.A.

Figure 2.A
Example of a 99% Confidence Interval



Kentucky and 28 other states use confidence intervals to determine whether schools meet their goals.

The U.S. Government Accountability Office indicates that 29 states are using confidence intervals when calculating adequate yearly progress. In August 2004, the Board of Education adopted a proposal to consider schools with subgroup scores that fall within a 99 percent confidence interval of the school's goal as achieving adequate yearly progress (Commonwealth. Kentucky Board of Ed. "Issues Brief — Assessment").

Adequate Yearly Progress of Student Subgroups

Schools must also demonstrate adequate yearly progress for different groups of students, such as minority and disabled students.

Reducing achievement gaps between different groups of students is one of the primary goals of No Child Left Behind. NCLB requires schools to make adequate yearly progress toward proficiency, not just among the student population as a whole, but also for racial and ethnic minority students, economically disadvantaged students, students with disabilities, and students with limited English proficiency. Subgroups of students are held to the same proficiency goals as the entire school. Therefore, if 52.4 percent of students in the school must score at the proficient level for the school to make adequate yearly progress, then 52.4 percent of the students in each of the subgroups must also score at the proficient level for the school to demonstrate adequate yearly progress. If a school meets adequate yearly progress for the student population as a whole but not for one of these subgroups, the school would be deemed as not meeting its goal under NCLB. The Kentucky General Assembly passed Senate Bill 168 in April 2002 that requires Kentucky school councils to address gaps in achievement among subgroups but also notes that a school can still meet its CATS performance goal even if a particular subgroup fails to meet the goal.

Schools will not be held accountable for these groups of students if there are too few of such students enrolled at the school.

NCLB does give states some flexibility in testing subgroups. Under the Kentucky NCLB plan, a school will not be held accountable for a subgroup if there are fewer than 10 students of the subgroup in a grade and 60 students in the subgroup within the school or if the subgroup accounts for less than 15 percent of the school population (703 KAR 5:001 (52-53)).

Safe Harbor

NCLB also contains a "safe harbor" provision that allows a school or district to achieve adequate yearly progress without meeting the standard performance thresholds. If a school or subgroup does not meet the performance threshold but does reduce the percentage of students who scored below proficient in the previous year by 10 percent or more, it will be considered as meeting its adequate yearly progress goal. These schools are recognized as having met adequate yearly progress goals because such improvement is considered significant under NCLB. KDE adopted administrative regulations to allow schools to utilize the "safe harbor" provisions of NCLB (703 KAR 5:001 (39)).

NCLB allows up to 5 percent of students to miss the tests. A zero score is assigned for these students. This provides an incentive for the school to stress high participation.

Objective 2: Participation Goals. NCLB and CATS differ on whether certain students' scores are included in accountability decisions. NCLB requires a 95 percent participation rate on the tests. Therefore, up to 5 percent of students in a school could miss the testing, and those missing scores would not count against the school's score. When evaluating the accountability index under CATS, a school receives a score of zero for every student that does not take the test; therefore, schools have more of an incentive under both CATS and NCLB to include every student in testing. Although administrative regulations have been adopted by KDE to comply with NCLB's required participation rate, CATS will continue to assign a score of zero for students not taking the tests.

Prior to NCLB, CATS counted students if they were enrolled in the school the day before the testing window, resulting in almost every student being included in accountability decisions. Following NCLB, the assessment scores of students enrolled in a school for less than a full academic year (defined as any 100 days of attendance during the school year from the first day of school until the first day of the testing window) are not included in school-level accountability calculations for either CATS or NCLB (703 KAR 5:001 (21) and (22)).

NCLB requires that students with limited English proficiency be included in the testing.

Including students with limited English proficiency is also different under NCLB than under CATS. Students who have limited English proficiency had been given two years to master English before being tested under CATS. Now these students are tested in their first year and are included in accountability decisions earlier due to NCLB.

NCLB requires consideration of other academic indicators, such as graduation rates. In addition to graduation rates, Kentucky also considers attendance and retention rates when determining whether a school has met its adequate yearly progress.

Objective 3: Other Academic Indicators. NCLB also requires that adequate yearly progress include academic indicators in addition to student assessments. For high schools, NCLB states that the additional indicator be graduation rates. For elementary and middle schools, the Kentucky Board of Education decided to use each school's full accountability index for the prior year under CATS (Commonwealth. Kentucky Dept. of Ed. “Kentucky's Consolidated”). The full index includes attendance and retention rates for both elementary and middle schools and dropout rates for middle schools. The NCLB graduation rate target set by KDE starts at 71 percent in 2002, increasing by 2.25 percentage points each year to an ultimate goal of a 98 percent graduation rate in 2014.

Accountability Under NCLB

A school faces separate consequences for not meeting its goals under NCLB and CATS.

Schools face separate accountability requirements under NCLB and CATS. As mentioned earlier, the two accountability systems operate independently of one another. If a school fails to meet its goals under either testing system, there are separate consequences and sanctions.

NCLB specifies escalating consequences the longer a school fails to reach its goals. While CATS also provides escalating consequences, the consequences and the timeline for implementing them differ from those required in NCLB. KDE has implemented administrative regulations that detail the consequences required by NCLB (703 KAR 5:020 (11-12)).

Students enrolled in a school that failed to meet its NCLB goals for two consecutive years must be allowed to transfer to another school that did meet its goals.

The consequences under both NCLB and CATS are described in Table 2.4. Under NCLB, if a school fails to make adequate yearly progress in any content area for two consecutive years, the school is identified as a school in need of improvement. The identification of the school must take place before the beginning of the school year following the second notice of failure to make adequate yearly progress. School districts must also provide students in Title I schools not meeting adequate yearly progress for two consecutive years with the option to transfer to another public school that has not been identified as needing improvement, within the same school district. The school district must give priority to the lowest-achieving children from low-income families. Schools that continue to not meet their goals must provide supplemental services such as tutoring, and may eventually be subject to alternative governance, such as the replacement of staff or the direct management of the school by the state Department of Education (United States. Dept. of Ed. “Accountability and AYP”).

Table 2.4
Consequences for a School That Does Not Meet Its Goals Under NCLB and CATS

NCLB intervention grows with each year a school or any subgroup at the school falls short of its goal:	**CATS** intervention grows the farther a school is from its two-year goal (levels are based on rankings relative to other schools not meeting their CATS goals):
Tier 1 (two years of not making adequate yearly progress) ❑ Notify Parents. ❑ Allow school choice. ❑ Write or revise school plan. Tier 2 (three years of not making adequate yearly progress) ❑ Continue school choice. ❑ Revise school plan. ❑ Offer supplemental services. Tier 3 (four years of not making adequate yearly progress) ❑ Continue school choice. ❑ Revise school plan. ❑ Continue offering supplemental services. ❑ Take corrective action. Tier 4 (five years of not making adequate yearly progress) ❑ Continue school choice. ❑ Revise school plan. ❑ Continue offering supplemental services. ❑ Continue corrective action. ❑ Write a plan for alternative governance.	Assistance Level 1 ❑ Scholastic review. The leader of the review team is designated by the commissioner of education in consultation with the superintendent. ❑ Eligible for improvement funds. Assistance Level 2 ❑ Scholastic review. The leader of the review team is designated by the commissioner of education. ❑ Eligible for improvement funds. Assistance Level 3 ❑ Scholastic audit, including analysis of staff evaluation needs. ❑ Possible demotions, transfers, or dismissals of teachers or principals. ❑ Assistance from a highly skilled educator. ❑ Eligible for improvement funds. Assistance Level 3 for two biennial cycles ❑ Continue Level 3 interventions. ❑ Allow students to transfer. ❑ Audit team may recommend removal of school council members.

Source: 703 KAR 5:020 and 703 KAR 5:120

There have been some interpretation and implementation issues associated with the provisions of NCLB. Interpretation questions result when schools meet their goals under one accountability program but not the other. This has resulted in delays in determining whether schools have achieved their goals and whether students should be allowed to transfer to other schools.

It is possible for schools to meet their goals under one system and not under the other.

While both NCLB and CATS use results from some of the same state tests, there are differences in the particular scores used and how the scores are evaluated. NCLB holds schools accountable specifically for the reading and math scores of students, while CATS holds schools accountable for student scores in all subject areas within the accountability index. NCLB requires certain levels of progress among subgroups of students that are not required in CATS. These differences can result in a school meeting its goal under CATS but not under NCLB or vice versa.

In 2004, 67 schools met their goals under CATS but not under NCLB.

In 2004, 67 schools met their goals under CATS but failed to meet their goals under NCLB. Wilt Elementary in Jefferson County, for example, obtained a CATS index score of 72.1, which exceeded its 2004 biennium goal of 63.8. The school, however, did not meet its overall NCLB goals. While the school met nearly all of its goals under NCLB, the percentage of disabled students scoring at the proficient level was below the school's goal.

Conversely, 301 schools met their goals under NCLB but not under CATS.

In 2004, 301 schools met their NCLB goals but did not meet their CATS goals. The reasons these schools did not meet their CATS goals vary. For example, 10 schools did not sufficiently reduce the number of students scoring at the novice level as required by CATS. Breckinridge Middle School in Breckinridge County was another school that met its NCLB goals but did not meet its 2004 CATS biennium goal. While a school's accountability index is used to determine if the school met its goals under both NCLB and CATS, the index from the previous year is used to determine if is met its NCLB goals. Therefore, Breckinridge Middle School's decline in its accountability index score for 2004 did not affect its NCLB score because the 2003 index was used for NCLB rather than the 2004 index.

Due to the slow grading of open-ended questions, in 2004, transfer decisions were based on preliminary, rather than final, results.

NCLB requires that assessment results for determining adequate yearly progress be provided to schools prior to the beginning of a school year. Preliminary test data for NCLB was given to schools and districts in early August 2004. Due to the slow grading of open-ended questions, only the NCLB multiple-choice data was available in August. Final NCLB data were released to schools in

early October along with the CATS results (Commonwealth. Kentucky Dept. of Ed. "Kentucky's Consolidated").

In some instances, the decision to permit students to transfer based on these preliminary scores was incorrect.

Because final scores were not available in August, transfer decisions were based on preliminary scores. In some instances, the decision to allow students to transfer was incorrect, as the final score differed from the preliminary score. Initial adequate yearly progress determinations were incorrect for 78 schools (Commonwealth. Kentucky Dept. of Ed. "2004 No"). These schools accounted for approximately 6 percent of Kentucky's schools. Of these schools, 38 appeared to make adequate yearly progress and 40 appeared to not make adequate yearly progress based on preliminary results. Initial determinations were later revised based on the final results.

In one instance, according to a *Lexington Herald-Leader* article, before the start of the 2004-2005 school year, about 50 Fayette County students transferred from Tates Creek Middle School to Southern Middle School because Tates Creek Middle was deemed as not meeting its NCLB goals in 2003 (Deffendall). Southern Middle was a transfer option because it appeared to meet its NCLB goals based on preliminary results. After reviewing final scores, however, it was determined that Southern Middle had not actually met its goals. Therefore, students should not have been allowed to transfer there.

The testing window will be moved back one week in the future to allow sufficient time for determining school scores.

NCLB requires that adequate yearly progress decisions be provided with sufficient time for parents to make decisions as to whether to transfer their children. To address the timeliness of the decisions, the testing window will be moved back one week for spring 2005 and beyond. Testing will start no earlier than April 1 and will last only two weeks (Commonwealth. Kentucky Dept. of Ed. "Kentucky's Consolidated"). Schools should be notified of the final decisions on adequate yearly progress by August, assuming that KDE is able to negotiate the earlier timeline with its assessment contractor.

Safe Schools

NCLB requires that students be allowed to transfer to another school if the school is considered "persistently dangerous" or if the student is a victim of a violent crime while on school grounds.

NCLB requires each state to implement a statewide policy to ensure that students may attend safe schools. The policy requires that any student who attends a persistently dangerous public school or who becomes a victim of a violent crime while on the grounds of a public school will be allowed to transfer to a safe public school within the local school district. Each state, in consultation

with local school districts, will determine the definition of "persistently dangerous public school."

At its June 2003 meeting, the Kentucky Board of Education adopted unsafe school choice policies that comply with NCLB. In addition, KDE's definition of a persistently dangerous school is a school in which "conditions exist over a period of time that expose students to injury due to violent criminal acts" (Commonwealth. Kentucky Dept. of Ed. "No Child Left Behind — Unsafe" 1). Violent criminal acts or offenses include acts such as forcible rape, robbery, and assault. School safety data for three years will be used to determine if these conditions exist in a school.

Teacher Quality

NCLB requires that teachers who teach "core academic subjects" be "highly qualified."

NCLB requires each school district to ensure that all teachers who teach "core academic subjects" meet certain requirements defining "highly qualified" teachers. The Education Professional Standards Board has released a brief report explaining the requirements and how it will address these requirements. NCLB makes a distinction between teachers who are new to the profession and those who have teaching experience. Teachers who are new to the profession must meet the requirements upon employment. Experienced teachers have until the 2005-2006 school year to meet the requirements. The term "core academic subjects" refers to English, reading, math, science, foreign languages, civics and government, economics, arts, history, and geography. The state's plan must include an annual increase in the percentage of highly qualified teachers employed in a school.

Teachers must have a baccalaureate-level degree and demonstrate knowledge in the subjects they teach.

All teachers, regardless of whether they are new to the profession and regardless of the level they teach, must have a baccalaureate-level degree and full Kentucky certification. All teachers must also demonstrate subject knowledge and teaching skills. How knowledge and skill are demonstrated vary. In Kentucky, all teachers may demonstrate their knowledge and skill by passing the PRAXIS II exam that measures prospective teachers' knowledge on various subjects to be taught (ETS). Elementary school teachers who are not new to the profession may also demonstrate their knowledge and skill by meeting the HOUSSE — "highly objective uniform state standard of evaluation" — requirements (Commonwealth. Education Professional Standards Board 8). HOUSSE provides credit for college courses, experience, and other professional development activities. Middle and secondary school teachers, regardless of whether they are new to the profession, have similar options to demonstrate their subject knowledge and

teaching skills, including completing an undergraduate major or graduate degree in the subject.

Under NCLB, paraprofessionals, such as teachers' aids, must have at least an associate's degree or two years of college, or be able to demonstrate knowledge in instructing reading, writing, and math.

NCLB also creates new requirements regarding the training and qualifications of paraprofessionals, such as teachers' aides, who work in public schools. All new paraprofessionals hired after January 8, 2002, must have obtained an associate's degree or higher; completed at least two years of college; or demonstrated that they have "knowledge of, and the ability to assist in instructing reading, writing, and mathematics" through a formal state or local assessment (20 U.S. Code Sec. 6319). All paraprofessionals must also have a high school diploma. All existing paraprofessionals, regardless of hiring date, must meet the same requirements no later than January 8, 2006. Kentucky statutes currently require teachers' aides to have a high school diploma or general equivalency diploma. The statutes do not require teachers' aides to take any formal assessment or obtain additional education.

Reporting

Both NCLB and CATS require annual school report cards that detail each school's performance.

Both NCLB and CATS require that an annual school report card be prepared for each public school in the Commonwealth. CATS requires schools to report information on academic achievement, attendance, retention, learning environment including parental involvement, enrollment, teacher qualifications, school safety information, spending per pupil, pupil/teacher ratio, and availability of technology. In addition, CATS requires high schools to report school dropout rates and student transition to adult life, which shows the percentage of graduating students that received jobs or enrolled in college. NCLB requires the following information to be on its annual report card:

- Aggregated NCLB scores in reading and math;
- Disaggregated NCLB scores by subgroup: race/ethnicity, disability, socioeconomic status, gender, migrant status, and limited English proficiency;
- Most recent two-year trend data reported by subject area and grade level in areas where assessments are required;
- Comparison data between actual achievement levels of each group of students to annual student achievement goals;
- Percentage of students not tested, disaggregated by student subgroups listed above;
- High school graduation rates; and
- Data on teacher qualifications, including the number of teachers with emergency certification and the percentage of classes not taught by highly qualified teachers.

A copy of a school report card for 2004 is shown in Appendix A. Kentucky's CATS school report cards include much of the information required by NCLB. Information on the percentage of students not tested and teacher qualifications are not detailed in Kentucky's report cards. In "Kentucky's Consolidated State Application Revised Accountability Workbook," however, KDE states that nearly all students are assessed, as required by state regulations. KDE also points out that data on teacher quality is available from the Kentucky Education Professional Standards Board. KDE has also noted that while the NCLB reports do not include data on teacher qualifications, the school report cards do include data on qualifications.

NCLB is designed with the intent to measure the progress of groups of students and schools, whereas CATS measures school progress. Specifically, NCLB requires an assessment that produces "individual student interpretive, descriptive, and diagnostic reports... that allow parents, teachers, and principals to understand and address the specific academic needs of students, and includes information regarding achievement on academic assessments aligned with State academic achievement standards" (Public Law 107-110 20 USC 6311 SEC. 111).

The addition of augmented assessments in grades 3 through 8 is an attempt to follow classes of children from year to year in reading and math. It is yet to be determined if the new assessments will provide the data required by NCLB. The new assessment plan has not yet been approved by USDOE.

Districts must also notify parents of students enrolled in a school that has been identified for improvement.

In addition to sending the report cards, districts must notify parents of students enrolled in a school that has been identified as needing improvement. According to KDE staff, districts must inform parents of the school's status, of the reason for the school's status, of what will be done to improve the school's status, and that students may transfer to another school. According to KDE, 132 Title I schools were identified as needing improvement based on scores from 2003-2004.

Conclusions

Because the specific education accountability provisions required in the NCLB Act differed from the provisions that existed in CATS, the Kentucky Board of Education and KDE have augmented CATS to comply with the NCLB provisions. Kentucky's state accountability plan, which details many of the modifications, has been approved by USDOE. Kentucky recently

submitted additional modifications to demonstrate how Kentucky will set educational standards and assess student and school performance. USDOE will be reviewing these changes at some point in the future, but this review has not yet been scheduled.

Chapter 3

Reliability and Validity of the Commonwealth Accountability Testing System

Educational assessments attempt to measure student performance and provide information that can be used to make educational decisions.

Educational assessment systems attempt to determine students' true educational achievements by evaluating their performance on several questions or tasks. The performance on these tasks provides a measure, or indication, of the student's achievements, which potentially can help educators and parents make decisions about the student's education.

Various factors can cause the assessment to misrepresent students' achievements. The difference between the measure of the students' achievements and their true achievements is referred to as error.

The goal in designing an assessment is typically to yield a score that accurately reflects the students' achievements. Scores, however, can be affected by a number of factors that cause them to misrepresent a student's true achievements. For example, if a student was ill on the testing day, performance might suffer. As a result, this student's score would likely understate true achievements. A score might also misrepresent a student's achievements if the individuals grading the test cannot properly score the student's work or if the tests required skills that were not intended to be evaluated. Some of these types of factors introduce error into the measure of students' performance by causing the measurements of the students' achievements to differ from their true achievements. Other factors cause the scores to mean something other than what was intended.

Some assessments, including CATS, attempt to measure the achievements of a group of students or of a school. These measures may also be subject to error.

Some educational assessment systems attempt to measure the achievements of a group of students rather than an individual student. CATS, for instance, was specifically designed to yield an accountability index that measures the overall performance of students within each school. These school-level measures are also subject to error, as a school's index may differ from the true ability of the students enrolled in the school.

To be useful, assessments should by sufficiently reliable and valid. An assessment is reliable if it yields similar results when repeated under similar circumstances. Error tends to reduce the reliability of an assessment.

For an assessment score to be useful, it must be both sufficiently reliable and valid. Reliability speaks to whether the assessment would yield similar results if it were performed several times. If students receive a similar score each time they take the assessment and their true understanding of the material did not change, an assessment would be considered reliable. If there is too much error associated with the assessment, a low score might be interpreted as signaling that a student does not know the material, or it may be dismissed as simply a chance occurrence. The error discussed

above reduces the reliability of the assessment. It should be noted that some degree of error exists within all educational assessments; therefore, scores might not perfectly represent a student's performance even on an assessment that is considered reliable.

An assessment is considered valid if it measures the achievements that were intended. CATS was designed to measure the degree to which students who are enrolled in a school as a group have mastered the core content.

Validity speaks to whether the scores indicate what they were intended to indicate. CATS was designed to measure knowledge of the core content. If questions are asked that require skills not included in the core content, then the CATS scores would not necessarily indicate knowledge of the core content but some other set of skills.

This chapter discusses the various issues relating to the reliability and validity of CATS and the use of CATS scores as indicators of school-level and student-level performance. The discussion that follows draws on the review of the relevant literature conducted by AEL, a nonprofit educational research and development firm based in West Virginia. The discussion also draws on statements about the reliability and validity of CATS from the National Technical Advisory Panel for Accountability and Assessments and staff analysis (National).

Reliability

Reliability is affected by several factors such as the number of questions or tasks required and the consistency of the raters who score students' work.

The research reviewed by AEL discussed several factors that affect the reliability of assessment scores. These factors include the number and types of questions or tasks required and the consistency of the raters who score students' work. The CATS accountability index attempts to measure the performance of the students enrolled at each school in Kentucky. As these factors affect the reliability of a school's accountability index, they affect the ability to make inferences about the school's performance.

Error can be associated with the student, the school, and the design of the assessment.

Some of the error that occurs with an assessment is associated with the individual students rather than with the assessment. Even if the exact same test was provided repeatedly to an individual student, the student's score might differ each time even if the student's skills and knowledge remained the same. Error can also be associated with the schools, such as illnesses in the school lowering attendance prior to exam time or a breakdown in the air conditioning unit making test conditions uncomfortable. As with the student, if an entire school was assessed multiple times using the same assessment, there still might be differences in the scores each time the assessment was taken. Error can also occur due to the manner in which an assessment is designed or scored.

Excessive error from these various sources tends to limit the ability to make inferences about students' educational performance.

Assessments represent a sample of a group of students' knowledge rather than their total knowledge and therefore limit the inferences that can be made about the students.

The questions that are asked on an assessment represent a sample of a group of students' knowledge rather than their total knowledge. Educators use this sample of student's knowledge to make inferences regarding their overall knowledge. For example, if a group of students does well on a test consisting of 10 questions, one might infer that the students are very knowledgeable even though the test might not cover all concepts they are expected to know. However, there are limitations to the inferences that can be made because a sample evaluates only a portion of the students' knowledge and therefore might provide an inaccurate measure of their true knowledge.

Assessments tend to be more reliable when they include a larger number of questions or tasks.

Assessments tend to be relatively less reliable if only a few questions are asked. Consider, for example, a situation where students are expected to know 100 concepts but are assessed based on their answer for only one question. The answer to this one question would not provide a very reliable assessment. The question might happen to be the one concept that an otherwise low-performing student knew. Alternatively, the question might happen to be the one concept that an excellent student did not know. If the question were a true/false or a multiple-choice question, a student could provide the correct answer simply by guessing. In each case, the assessment would not accurately represent the student's performance.

The research reviewed by AEL points out that reliability can be improved by increasing the number of items on which students are evaluated. Increasing the number of items reduces the probability that a chance occurrence might affect the overall score. For example, it would be more difficult for students who were guessing to achieve a high score if there were more questions. Similarly, the results of a test with 10 questions would be less likely than the results of a test with one question to be affected by just happening to include a question on the one area some students did not know.

The reliability of open-response questions, on-demand writing questions, and writing portfolios is affected by the consistency of the raters who score them.

Reliability may also be affected by the raters who evaluate students' answers. In CATS, raters evaluate the work students provide for open-response questions, on-demand writing questions, alternate portfolios, and writing portfolios. For these types of assessments, raters must determine whether the students have adequately demonstrated the required skills based on a predefined standard called a rubric. The rubric details the required elements

that an answer should include to earn a certain score. For example, the writing assessment rubric requires that writers create "effective sentences that are varied in structure and length and are complete and correct" (Commonwealth. Kentucky Dept. of Ed. "Writing").

Raters might score students' work inconsistently due to unclear rating instructions or personal differences in how the raters apply the instructions.

In spite of efforts to have raters consistently score student work, raters do not always agree on the score a piece of work should be given. There are several reasons raters might disagree. Some raters might be more difficult than others. Rubrics might lack clarity as to what constitutes certain scores, which might introduce subjectivity into the rating process. As a result of these inconsistencies, students' scores can be partially affected by who scores the work rather than by the quality of the work. This introduces additional error into the measurement and reduces reliability.

Interrater agreement rates are one measure of scoring consistency. The rates indicate the percentage of times two or more raters agree on what score to assign a work.

AEL discussed some of the issues related to raters and reliability and some common measures used to evaluate the extent to which raters consistently score student work. Many studies have used interrater agreement rates to evaluate the extent to which raters agree or disagree. Interrater agreement rates are developed by having two or more raters score several pieces of work and calculating the percentage of times they agree.

Some researchers have provided suggestions for acceptable interrater agreement rates.

AEL cited interrater agreement rates from several studies of educational assessments. Several researchers have made recommendations as to what interrater agreement rates should be. For example, Reckase suggested that rates be 70 percent or higher. This suggestion appears to be based on the rates that are commonly found in other studies. In another study, Herman et al (1992) indicated that agreement rates should be 75 percent to 80 percent.

Comparisons of agreement rates across different types of assessments can be difficult, as agreement rates will tend to vary based on the number of categories in which a work might be scored. Assessments with fewer categories tend to have higher agreement rates simply because there is less opportunity to disagree.

While the agreement rates in these studies do provide some context against which to compare CATS, caution is warranted. Agreement rates will be affected by the number of categories in which a work might be scored. For example, under CATS, there are four scoring categories: novice, apprentice, proficient, and distinguished. When there are fewer categories available, agreement rates would be expected to be higher. Based on staff interviews with teachers who score writing portfolios, it appears that differences between raters tend to occur for portfolios that are near the borderline. That is, a portfolio might seem to meet the criteria for two adjacent categories. If raters were required to assign a student's work to one of two categories, agreement might be relatively high. If raters could assign a written work to one of six categories, however, there would likely be more differences as the difficulty of determining whether the work should be assigned to any specific

category is multiplied. Because the number of categories affects agreement rates, it is not clear whether the rates suggested above are relevant for evaluating CATS. An agreement rate of 70 percent might be considered high if there are six categories but might be considered low if there are only four categories.

In their studies, Reckase and Herman et al (1992) did not provide the information needed to determine whether the agreement rates discussed provide meaningful context for comparing the agreement rates in CATS. Therefore, the agreement rates in CATS are most informative when compared over time and within CATS.

Interrater agreement rates for the open-response and on-demand writing questions on CATS range from 68.5 to 86 percent.

Table 3.1 shows interrater agreement rates for the open-response and on-demand writing questions in CATS. For 4th graders, agreement rates ranged from 68.5 percent in practical living to 86 percent in reading and on-demand writing. That is, raters agreed on the scores in practical living for 68.5 percent of the responses evaluated. Raters initially disagreed for the remaining 31.5 percent. Agreement rates appeared to be highest for middle school students, ranging from 80 to 90 percent.

Table 3.1
Interrater Agreement Rates for CATS Open-response Questions (Percent)

	4th & 5th Grades		7th & 8th Grades		10th,11th, & 12th Grades	
	2000	2001	2000	2001	2000	2001
Reading	82.0	86.0	87.5	90.0	84.0	86.0
Math	85.0	77.8	84.0	84.0	88.0	81.8
Science	80.0	80.0	81.5	80.0	86.5	78.0
Social Studies	81.0	81.6	88.5	85.5	90.0	84.0
Arts & Humanities	81.0	84.0	84.0	89.0	88.0	72.0
Practical Living	83.0	68.5	80.0	81.0	79.0	86.0
On-demand Writing	80.0	86.0	80.0	82.0	90.0	72.0
Average	81.7	80.5	83.6	84.5	86.5	80.0

Source: Commonwealth. Kentucky Dept. of Ed. "Kentucky Core Content Tests 2002."

When discussing agreement rates, adjacent agreement rates are often mentioned. Adjacent agreement rates considered two raters to be in agreement if they only differed by one category. While raters typically would be considered as not agreeing if one assigns a score of proficient and the other assigns a score of apprentice, with adjacent agreements, these raters would be considered as if they did agree because there was only a difference of one category between the scores they assigned. Adjacent agreements will always

be higher than agreement rates based on an exact agreement of the score categories and might not provide much information that can be used to evaluate rater inconsistencies when there are few score categories.

Classification Accuracy

Classification accuracy indicates the probability that a student's or a school's performance is correctly classified.

As the goal of an assessment system is typically to draw inferences regarding a student's or school's performance, it is useful to understand how accurately a score represents the true level of performance. Classification accuracy refers to the probability that an assessment correctly classifies the student's or school's performance. For instance, a school might be classified as proficient based on the performance of its students, which might be the school's correct classification. Due to the error associated with the assessment, however, the school's true level of performance might really be apprentice or distinguished.

An analysis by HumRRO estimated the probability that schools were correctly assigned to one of three categories: meets goal, progressing, or in need of assistance.

The classification accuracy of CATS scores for schools was calculated for 2002 by the Human Resources Research Organization (HumRRO), a non-profit organization that conducts validity and reliability research for KDE. The analysis evaluates the probability that a school was correctly classified into one of three categories: meets goal, progressing, or in need of assistance.[1] The analysis was limited to just the accountability index and did not include an analysis of novice and dropout reduction, which can also affect the classification of a school ("The Accuracy of School").

HumRRO estimated that there is a 77 percent probability that a school is correctly classified.

HumRRO concluded that the classification accuracy associated with CATS was 77 percent. That is, there is a 77 percent probability that a school is correctly assigned and a 23 percent probability that the school is incorrectly assigned. This suggests that out of 100 schools, one could expect that 23 schools would be incorrectly classified. Some schools' true performance would likely be lower than indicated by the accountability index, and some schools' performance would be higher than indicated by the index.

Under CATS, schools are considered as meeting their goals if their accountability index is sufficiently close to their goal. This practice reduces the probability that a school is correctly classified.

The HumRRO report noted that one factor that reduces the classification accuracy was the use of a margin of error in determining whether a school meets its goals. There can be several schools with accountability indices that are very close to but just below their goals. The index does not perfectly reflect a school's

[1] Although schools classified as in need of assistance are also assigned to one of three categories within assistance, the analysis did not go to this level of detail.

true level of performance. The error associated with the index can make a school's index be higher or lower than the school's true level of performance. Schools that were disadvantaged by this error might have met their goals if the accountability index could perfectly measure their performance. One cannot determine, however, whether an individual school was advantaged or disadvantaged by the error. Recognizing that this error can cause some schools to be disadvantaged, all schools that are sufficiently close to their goals are classified as meeting their goals. HumRRO notes that this practice reduces "the risk of erroneously under-classifying schools" but also increases the risk of over-classifying schools (4). If a school's index falls in the progressing range, the school's true level of performance is more likely to be progressing than it is to be meeting its goal. Therefore, considering schools that are close to their goals as meeting their goals reduces the classification accuracy. In the absence of this practice, HumRRO concluded that schools would have an 82 percent probability of being correctly assigned rather than the 77 percent with this practice.

Writing Portfolios

Writing portfolios are also subject to error, which reduces their reliability.

While writing portfolios are different from traditional assessments, the same issues regarding reliability apply. Writing portfolios provide an indication of a student's writing skills but are subject to error. As with other assessments, this error can be due to the number of items a student must complete, how raters score portfolios, and other factors. The amount of error that exists in writing portfolio scores determines their reliability.

Each year, the writing portfolios for a random sample of schools are rescored by a team of auditors. Comparing the schools' original scores to the audited scores provides an opportunity to evaluate the reliability of portfolios.

Each year, a number of the writing portfolios submitted throughout the state are independently audited as a check of the scores that schools assign. Schools are selected for an audit because their writing portfolio index is significantly different from what was expected based on its other CATS scores or are selected randomly. In 2004, there were 26 schools and 1,538 portfolios that were sampled because the schools' scores appeared to be high relative to their other CATS scores; and there were 75 schools and 7,754 portfolios that were sampled randomly. Comparison of the scores assigned by schools and the auditors provides an opportunity to evaluate how differences among raters affect the reliability of portfolios. Several measures of reliability are presented below. All measures are based only on those portfolios that were randomly sampled.

In 2004, the scores provided by the schools and the auditors matched for approximately 75 percent of the portfolios. Agreement rates were highest for the 4th-grade portfolios.

Table 3.2 shows the interrater agreement rates from 1999 through 2004. These rates indicate the percentage of audited portfolios for which the scores provided by the schools and the auditors matched. The overall agreement rate between schools and auditors ranged from 72.7 percent to 80.6 percent and was 75 percent in 2004. A 75 percent agreement rate indicates that the scores provided by the schools and the auditors were the same for 75 percent of the portfolios. The schools and auditors disagreed on the remaining 25 percent of the portfolios. Fourth-grade agreement rates were consistently higher than either the 7th- or the 12th-grade agreement rates, indicating that schools and auditors were more likely to agree on the scores of portfolios written by 4th graders than those written by 7th and 12th graders. Except for the 12th grade in 2000 and 2004, and 7th grade in 2003, agreement rates have been above 70 percent.

As discussed earlier, AEL cited two studies that recommended certain levels of agreement. Reckase suggested agreement rates of 70 percent or higher and Herman et al (1992) suggested agreement rates of 75 to 80 percent. As mentioned, it is not clear how relevant these are for evaluating CATS.

Some agreement should be expected even if scores were randomly assigned by both the schools and the auditors.

It is also useful to note that some agreement should be expected. Even if the schools and auditors randomly assigned scores, the scores for some portfolios could agree simply due to chance. Staff analysis shows that if scores were randomly assigned, one would still expect agreement rates of 34 to 47 percent. In addition, most schools have two raters score portfolios and have procedures to address any differences between the raters. For example, in some schools, if two raters disagree about what score to assign to a portfolio, a third rater might be used to make a final decision on the score. These procedures should reduce the likelihood that a rater who might introduce error into the scoring process for some reason would misclassify a portfolio. This process could also result in fewer differences between the scores provided by schools and auditors.

Table 3.2
Agreement Rate and Correlation
Between School and Audited Scores

Year	Grade	Agreement Rate (Percent)	Correlation
1999	4	77.9	0.72
	7	78.2	0.57
	12	71.9	0.65
	Total	76.4	0.69
2000	4	80.3	0.77
	7	76.4	0.50
	12	60.2	0.48
	Total	73.4	0.59
2001	4	82.6	0.79
	7	83.1	0.80
	12	75.5	0.75
	Total	80.6	0.79
2002	4	82.7	0.79
	7	73.1	0.70
	12	74.0	0.72
	Total	76.7	0.79
2003	4	79.0	0.69
	7	68.5	0.67
	12	70.0	0.72
	Total	72.7	0.71
2004	4	81.8	0.68
	7	74.3	0.72
	12	66.5	0.61
	Total	75.1	0.70

Source: Staff analysis of data provided by the Commonwealth of Kentucky Dept. of Ed.

Reliability can also be evaluated by examining the correlation between the scores provided by schools and by the auditors. A high level of correlation indicates that the schools and the auditors consistently ranked the same portfolios high and the same portfolios low.

Reliability can also be evaluated by considering how consistently schools and auditors rank portfolios, which can be measured by the degree of correlation that exists between the two sets of scores. If schools and auditors tend to score the same portfolios high and the same portfolios low, correlation will be high. If schools tend to score certain portfolios high and auditors tend to score these same portfolios low, correlation will be low. Low levels of correlation signal inconsistencies in how schools and auditors rank students and indicate that reliability may be low.

The level of correlation between the scores provided by the school and those provided by the auditors has decreased since 2001, suggesting that there has been less consistency between the rating by schools and the rating by auditors in recent years.

Table 3.2 also shows the degree of correlation between the scores assigned by schools and the scores assigned by the auditors. Correlations are stated on a scale of 0 to 1. Zero indicates no correlation and 1 indicates exact correlation. The correlation for all grades was 0.70 in 2004. While the correlation for all grades

improved in 2001, it decreased in 2003 and 2004, suggesting that there has been less consistency in how schools and auditors score portfolios. Within grades, the levels of correlation were usually highest for 4th graders but lower for 7th and 12th graders. As with the agreement rates, this suggests that schools and auditors have an easier time agreeing on the scores for 4th graders.

When the schools and auditors disagreed about the score to assign, the auditors usually provided a lower score than the schools.

In most instances, the schools and auditors agreed on the scores to assign portfolios. When differences did exist, however, the auditors usually provided lower scores than the schools. The auditors provided lower scores for 86 percent of the portfolios for which there was a disagreement. This number has grown from 79 percent in 1999 to 92 percent in 2004. In 2004, the average writing portfolio index for schools was approximately 10 points lower when based on the scores provided by the auditors than they were when based on the scores provided by the schools. Table 3.3 shows the difference in the indices over time and by grade. The differences were smallest for 4th graders; and in two years, 1999 and 2001, the differences for 4th graders were not statistically significant.

The differences in scores suggest that there may be a problem with the validity of portfolio scores. The difference might be attributable to teachers having an incentive to be more lenient or it might be due to differences in the achievements of schools and auditors to score portfolios.

The fact that the auditors consistently score lower than these schools suggests that there may be a problem with the validity of the portfolio scores. It is possible that one set of scores or both sets of scores are not providing the intended measure of writing skills. It is not known which sets of scores, those from the schools or those from the auditors, are most "correct." One might argue that schools have an incentive to be more lenient when scoring portfolios or that the teachers are influenced by their knowledge of a student. On the other hand, one might argue that the teachers in schools have more experience evaluating student work. The differences above suggest that the schools and auditors apply the standards for rating portfolios somewhat differently. These rater differences are more pronounced for 7th- and 12th-grade portfolios.

A g-study measures the amount of error that various factors contribute to an assessment.

As mentioned, the error associated with writing portfolios comes from several factors such as the design of the portfolio assessment, the raters, the students, or the schools. In evaluating the reliability of assessments, researchers often attempt to determine the amount of error that is attributable to each of these factors. A generalizability-study, or g-study, measures the error from certain factors. The goal in designing an assessment is to minimize the amount of error attributable to the design and the raters.[2]

[2] Error can also be associated with the interaction of factors. For example, error might be relatively higher when certain raters score the portfolios of certain students.

Table 3.3
Average Scores of the Randomly Audited Writing Portfolios

Year	Grade	Average School Rating	Average Auditor Rating	Difference	Was Difference Statistically Significant?
	4	58.4	57.5	0.9	**No**
1999	7	35.1	28.1	7.0	Yes
	12	60.7	50.1	10.6	Yes
	Total	49.8	43.9	5.9	Yes
	4	57.7	55.0	2.7	Yes
2000	7	42.9	34.8	8.1	Yes
	12	65.4	48.5	16.9	Yes
	Total	53.9	45.2	8.7	Yes
	4	61.4	60.1	1.3	**No**
2001	7	40.9	35.0	5.9	Yes
	12	59.1	50.4	8.7	Yes
	Total	53.0	47.6	5.4	Yes
	4	66.6	63.8	2.8	Yes
2002	7	45.0	32.9	12.1	Yes
	12	65.3	57.2	8.1	Yes
	Total	58.2	50.4	7.8	Yes
	4	74.5	69.6	4.9	Yes
2003	7	54.5	41.0	13.5	Yes
	12	64.0	53.1	10.9	Yes
	Total	64.1	54.4	9.7	Yes
	4	75.5	70.4	5.1	Yes
2004	7	58.4	46.9	11.5	Yes
	12	69.9	56.3	13.6	Yes
	Total	67.0	57.2	9.8	Yes

Note: Averages were calculated by assigning the following values to the scores provided: Novice = 13, Apprentice = 60, Proficient = 100, Distinquished = 140, blank or incomplete = 0. Original portfolios assigned an incomplete or blank score were omitted. The statistical significance of the difference was evaluated using a t-test at the 95 percent level.
Source: Staff analysis of data provided by the Commonwealth of Kentucky. Dept. of Ed.

This analysis evaluated the error attributable to schools and to raters.

Table 3.4 shows the margin of error associated with raters from the school-level g-study on the writing portfolio indices from 1999 through 2004.[3] Only schools that were randomly selected to be audited were included in the analysis. In this case, the analysis evaluated the error attributable to schools and to raters. The figures in the table represent the margin of error that results from

[3] There are two types of g-studies that are typically performed: relative and absolute. Shavelson and Webb describe a relative g-study as measuring the reliability when making decisions about the relative rankings or ordering of students or schools. They describe an absolute g-study as measuring reliability when making decisions about assigning or categorizing students or schools based on a certain level of competency. As CATS scores are used to determine whether students within schools have mastered the core content, the figures in Table 3.4 reflect the statistics derived from an absolute g-study.

inconsistencies in how raters score portfolios.[4] In 2004, the margin of error associated with rater inconsistencies for a 4th-grade writing portfolio index was +/- 12.6 points.

The results of the g-study suggest that the schools and auditors have less difficulty determining the scores for 4th-grade portfolios than for 7th- and 12th-grade portfolios.

A larger margin of error indicates less reliability. As with the other measures of reliability, the g-study suggests that 4th-grade scores have been more reliable than 7th- or 12th-grade scores. In each year, rater error was lowest for the 4th grade. These statistics are consistent with the other measures discussed and suggest that raters have less difficulty determining the scores for 4th-grade portfolios than they do for 7th- and 12th-grade portfolios.

Table 3.4
Writing Portfolio Index
Margins of Error Associated With Inconsistencies in Rating

Year	4th Grade	7th Grade	12th Grade
1999	+/- 10.4	+/- 15.0	+/- 17.2
2000	+/- 9.7	+/- 14.0	+/- 24.6
2001	+/- 8.6	+/- 22.2	+/- 15.0
2002	+/- 10.0	+/- 18.0	+/- 15.1
2003	+/- 10.7	+/- 22.8	+/- 17.3
2004	+/- 12.6	+/- 16.7	+/- 19.8
Average	+/- 10.3	+/- 18.1	+/- 18.2

Note: The margins of error are based on an absolute generalizability-study at the school level using a single rater. Scores from the schools and the auditors were used to calculate the margins of error.
Source: Staff Analysis of data provided by the Commonwealth of Kentucky Dept. of Ed.

In spite of similar training for those rating portfolios at the schools and those rating portfolios for the audits, differences in the scores assigned continue to exist. Although the reason is not known, several factors could contribute to the difference.

The comparison of the scores provided by schools and the scores provided by the auditors show that there are differences in how portfolios are scored. These differences occur in spite of similar training for those rating portfolios at the schools and those rating portfolios for the audits. The cause of the difference is not clear. In its review of the research literature on portfolios, AEL noted that several factors could contribute to rater errors, including differences in severity or leniency, a teacher's perception of the student, a tendency to score portfolios in the middle of the scoring range, and a tendency to restrict the scores provided (46). When a difference does exist, the scores assigned by schools are usually higher than the scores assigned by the auditors. This difference might occur as a result of schools having an incentive to assign higher scores. Alternatively, the differences in scores might be caused by differences in the ability of schools and the ability of the

[4] The margins of error are based on a 95 percent confidence interval.

auditors to evaluate writing portfolios. While there have been efforts to address these types of problems, they may still exist to some degree and contribute to differences in how schools and auditors score portfolios.

Integrating Writing Portfolios

Some individuals argue that portfolios encourage good instructional practices. In its review of the literature, AEL found little empirical evidence to support this claim.

According to AEL's review of the research, some individuals argue that portfolios encourage teachers to use good instructional practices and that portfolios are better for assessing writing skills than multiple-choice questions (Reckase). AEL and Herman et al (1993) have found little empirical evidence to support this claim. AEL noted that there is some evidence that writing portfolios affect instruction. For example, HumRRO conducted several interviews with school administrators and faculty and determined that some schools developed additional student conferences to improve the students' portfolios ("Variations"). The research on the effects that portfolios have on teaching practices, however, does not demonstrate whether there are positive effects on learning.

The lack of standardization, which is cited as an advantage of portfolios, could reduce their reliability.

AEL pointed out that many of the advantages of portfolios can create measurement issues. For example, the lack of standardization might introduce subjectivity into the evaluation on the portfolios, which could result in less reliability or could affect whether portfolios actually measure what was intended. Concerns have been raised regarding the amount of assistance some students receive from others, such as teachers or parents. Gearhart et al studied the amount of support nine teachers provided to students who were preparing writing assignments. The authors stated that "the quality of work appeared to be a function of substantial and uncontrolled support as well as student competence. Thus the validity of inferences we can draw about student competence based solely on portfolio work becomes suspect" (7). If students receive a great deal of assistance, the portfolio may not truly represent the work of the student. AEL suggested that increased standardization could help mitigate these measurement issues, but this would provide teachers with "less flexibility" and might reduce many of the advantages of portfolios (58).

Much of the research regarding the effect of portfolios on teaching practices involved surveys of educators. For example, Koretz et al surveyed teachers and principals in Vermont regarding their views on mathematics portfolios and writing portfolios. The study found that there was confusion among teachers as to the intended purpose of the portfolios, that the number of revisions and amount of

assistance varied across teachers, that portfolios required considerable time, and that there were negative attitudes regarding portfolios. The negative attitudes focused on the amount of time and resources required by portfolios. Teachers surveyed indicated that they did change their teaching practices as a result of the portfolios. The authors noted that teachers devoted "more attention to problem solving and communication" and that the educators surveyed generally felt the additional burden was worthwhile (xvii). While the study evaluated the opinions of educators, it did not evaluate whether portfolios improved learning.

Validity

A valid assessment measures what was intended.

For an assessment to be valid, it must measure what was intended. CATS was intended to measure how well students within a school have mastered the core content. To the extent that the assessment tends to measure skills or knowledge not covered by the core content, validity would suffer. Likewise, if the assessment does not include certain portions of the core content, validity would suffer.

CATS consists of various types of assessment including a norm-referenced test, the Kentucky Core Content Test, the alternative portfolio, and the writing portfolios. Each type of assessment serves a different purpose within the Kentucky accountability system and affects the validity of CATS as a measure of the core content.

Norm-referenced Tests

The norm-referenced test that is included in CATS was not designed to cover the core content and would likely reduce the validity of CATS as an assessment of the core content. Its purpose, however, was to provide a comparison of the performance of Kentucky students to that of students from other states.

The norm-referenced test was added in 1997 to address concerns that the performance of Kentucky students on the KCCT could not be compared to the performance of students from other states.[5] The addition of this test was not to address validity but to provide a method for comparing the performance of Kentucky's students to that of students of other states. The norm-referenced test was not designed to cover the core content. As a result, it may include questions on areas that are not part of the core content or may not include questions on areas that are part of the core content. As such, the norm-referenced test would likely reduce the validity of CATS as an assessment of the core content.

[5] While the norm-referenced test was first administered in 1997, it was not included in the accountability index until 1999.

Types of Tasks Required

The KCCT consists of several types of tasks and questions, including multiple-choice questions, open-response questions, writing prompts, and a writing portfolio.

KCCT requires students to perform several types of tasks as part of the assessment. The majority of the tasks are multiple-choice questions. Students must also complete several open-response questions, respond to an on-demand writing prompt, and prepare a writing portfolio. AEL refers to these types of assessment tasks as performance assessments.

Some researchers argue that tasks such as open-response questions or writing portfolios are better than multiple-choice questions for assessing complicated skills. Other researchers disagree.

AEL reviewed several studies that discussed the advantages and disadvantages of these types of assessments. The studies generally concluded that using multiple-choice questions was a low-cost method of covering a large number of subject areas. Multiple-choice questions are also easy to score. There was some debate as to whether multiple-choice questions could be used to assess complicated skills. Performance assessments such as open-response questions were offered as an alternative that would allow students to demonstrate these more complicated skills. Performance assessments require more time to administer and to score. Although there appears to be some disagreement regarding whether performance assessments evaluate different skills than do multiple-choice questions, AEL concluded: "Using combinations of multiple-choice and open-response items … strengthens the content coverage and alignment of the assessments with the Kentucky Core Content" (5).

Accommodations and Modifications

Some students face barriers, such as learning or physical disabilities, that can reduce the validity of a traditional assessment.

Some students face certain barriers with traditional assessments that make it difficult to properly evaluate their achievements. For example, it would be difficult to assess a visually impaired student using a traditional written test form. In instances such as this, the assessment might not provide a valid indication of the student's achievements. These types of barriers also reduce the validity of school-level measures, particularly for schools that have a large number of students who face these barriers.

These barriers might cause the assessment to measure skills that were not intended. Accommodations and modifications attempt to improve the validity of assessments by removing these barriers.

In order to address these barriers, various accommodations and modifications have been used. The intent of accommodations and modifications is to ensure that the specific knowledge or skill of interest is tested rather than some other knowledge or skill. For example, a question intended to measure math skills that is written in English might prove very difficult for a student with limited English proficiency. In this case, the question would test both English and math skills rather than just math skills. By ensuring that the skill of interest is being tested, accommodations and

modifications improve the validity of the assessment. In a paper discussing validity issues with accommodations, Sireci noted that while accommodations and modifications should remove these barriers, they should not provide the accommodated students with an unfair advantage over other students.

Accommodations and modifications might reduce the validity of an assessment if they change the skills being evaluated.

According to Sireci, while accommodations and modifications are intended to improve the validity of an educational assessment, they can also reduce the validity. This author's research concluded that validity improves if the changes remove some aspect of the test administration that prevented certain students from demonstrating their mastery of the knowledge or skills being measured. Validity is reduced, however, if the accommodations and modifications change the skill or knowledge that is being evaluated. For example, administering a reading assessment with a computer might cause the assessment to evaluate computer skills in addition to reading skills.

Researchers evaluate accommodations and modifications by comparing their effects on students with and without the barriers.

In determining the effect that changes have on the validity of an assessment, researchers often compare the effect of the changes on both the students for which the changes are intended and on students in general. Ideally, the changes should improve the scores of the student for which the changes are targeted but not those of other students (Sireci).

A study by Kosciolek is an example of this type of analysis. In this study, both learning disabled students and general education students took a test with and without an accommodation. The accommodation was an audio cassette player that allowed students to hear the questions. The scores of both groups on each type of test were compared. The author concluded that the scores of learning disabled students improved with the accommodation, but the scores of the general education students did not improve with the accommodations. This result suggests that the accommodation improved the validity of the assessment by removing a barrier for some students but did not provide them with an unfair advantage over other students.

Several types of accommodations are available for students with disabilities, such as providing extended time or reading the test questions aloud.

Students With Disabilities. Standard assessments are often more challenging for students with disabilities than for other students. Several types of accommodations are available for these students, including providing multiple days to take a test, extending the amount of time allowed for completing the test, and reading aloud the test questions and instructions.

AEL found that there were inconsistencies in the research evaluating the effects of accommodations and modifications. These inconsistencies might be due to differences in how these changes were administered.

After reviewing the research, AEL concluded that "the effects of these accommodations on the performance of students with disabilities are inconclusive" (55). For all of the accommodations studied, there were studies suggesting that the accommodations were helpful to students with disabilities, but there were also studies suggesting that the accommodations did not improve the scores of these students. These inconsistencies might be due to differences in the research techniques used to measure the effects of the accommodations. Alternatively, it could be that the effects of the accommodations were sensitive to how the accommodations were administered or differed for students with different types of disabilities.

Students with limited English proficiency might experience greater difficulty in demonstrating certain skills, such as math, if the assessment requires knowledge of the English language in order to understand the question.

Students With Limited English Proficiency. Assessing a student with limited English proficiency (LEP) creates the additional problem of measuring the student's knowledge of a particular concept without making the test item more difficult due to a language barrier. A math question that is asked in English will require both the math skills necessary to solve the problem and the knowledge of English to understand the question being asked. LEP students will likely have more difficulty than will other students in answering these types of questions. If the goal of the question is partially to evaluate English language skills in addition to math skills, this may be a valid question. If the goal is simply to evaluate the student's math skills, however, this type of question would not provide a valid assessment of an LEP student's math skills. An LEP student with excellent math skills might answer incorrectly simply because he or she could not understand the question. If English proficiency is required to answer a question but is not part of the skills or knowledge being evaluated, then the question would not provide a valid measure of the skills or knowledge for LEP students.

Some studies have concluded that providing students with glossaries and allowing them additional time can improve the scores of students with limited English proficiency.

As with other accommodations, the accommodations for LEP students are intended to remove the barriers without providing them with some advantage over other students. Common accommodations for LEP students include extended time, glossaries, and modified English. AEL concluded that the research on accommodations indicates that modified English, extended time, and glossaries with extended time were useful for improving the scores of LEP students (Abedi et al. NAEP and Abedi et al. "Assessment"). One study concluded that glossaries without extended time were not effective (Abedi et al. NAEP). Although AEL was able to identify some studies that have evaluated the effectiveness of accommodations and modifications for LEP

students, it noted that conclusions from the research were "difficult to draw" (57).

While providing assessments to students in their native language can be useful, AEL noted that this type of accommodation can be cost prohibitive given the number of languages that might be required and the small number of students who speak each language.

The effect of accommodations and modifications on CATS is unclear. AEL concluded that the research on the types of changes permitted in CATS was not conclusive.

Accommodations Under CATS. CATS provides the following accommodations or modifications: scribes, extended time, use of technology, and interpreters. According to the Kentucky Department of Education's procedures for including special populations, students are only allowed to use accommodations or modifications that are part of their normal curriculum ("Inclusion"). The effects of these accommodations on the validity of CATS are unclear. AEL frequently noted that the research found "mixed" results (54). While the research on these accommodations did not provide clear support for their use, the research also did not clearly discredit their use. As the effects may differ due to factors such as which students receive the accommodations and how the accommodations are administered, an evaluation of the accommodations used within CATS may be required, as noted by AEL (55).

Accommodations and modifications for LEP students may become a greater concern with the implementation of the No Child Left Behind Act. Prior to NCLB, CATS did not assess LEP students for the first two years. NCLB, however, requires that LEP students be assessed during their first year, allowing them less time to learn English. One study of accommodations found that "the best predictor of math scores was the length of time the student has lived in the United States" (Abedi et al. NAEP ix). Given that LEP students will have less time to learn English, language might be a more significant barrier to assessing their knowledge of the core content.

CATS Assessments as Indicators of Student-level Performance

A frequent question is whether student-level CATS scores accurately measure a student's achievements.

Although CATS was designed to evaluate schools rather than students, CATS does yield student-level scores. A frequently asked question has been whether or not these student-level scores accurately measure a student's achievements. The reliability and validity issues that affect the use of CATS scores at the individual level are the same, except more dramatic, as the issues that affect

the use of CATS scores at the school level, such as coverage of the core content and consistency of scoring.

Coverage of the Core Content

CATS was designed to cover the entire core content at the school level. Students, however, are only tested on a portion of the core content.

CATS was designed specifically to provide reliable and valid assessments of school-level performance rather than student-level performance. Due to the time and cost associated with testing each student on the entire core content, multiple test forms were developed within each subject area. Each form covers a portion of the core content. For a school, these forms completely cover the core content. Each student, however, is only assessed on a portion of the core content.

By only being assessed on a portion of the core content, a student may happen to get a test form with questions that the student does not know or a test form with questions the student does know.

By only being assessed on a portion of the core content, an individual student may happen to get a test form with questions that the student does not know. If the student had extensive knowledge of all other areas of the core content, being tested on a form that just happens to have a question he or she does not know would make it appear that the student's mastery of the core content is lower than it actually is. Similarly, an otherwise low-performing student may happen to get a test form with questions that he or she can answer correctly. In both of these cases, the scores would provide a poor indication of the students' true knowledge of the core content.

The probability that students happen to get a form that includes questions they know increases when there is less coverage of the core content.

The possibility that students will just happen to get questions that they know or questions that they do not know will always exist. The probability of these occurrences increases when there is less coverage of the core content. Therefore, by only covering a portion of the core content, each test score may be considered somewhat incomplete for individual students. The degree to which the test is incomplete depends on the level of coverage.

Probability of a Student Being Misclassified

HumRRO estimated the probability that students are correctly classified based on their CATS scores.

HumRRO estimated the student-level classification accuracy for each school level and each area of the core content ("The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2001 and 2002" and "The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2003"). The student-level classification accuracy measures the probability that a student is misclassified as a result of the manner in which the assessment is designed and administered. As with school-level assessments, various types of error may affect a student's score. These probabilities are

estimated by examining the amount of error and the sources of error. Generally, these sources of error might consist of factors that are related to the design of the assessment or are related to the students.

The probability of a student being correctly classified was highest for high school students' reading scores.

Table 3.5 shows the student classification accuracy for KCCT scores by school level and subject. In 2003, the classification accuracy was highest for high school students' reading scores. This figure indicates that the probability that students were accurately classified based on their reading scores was 81.1 percent. Alternatively, there is an 18.9 percent chance that the students were misclassified. That is, a student who is classified as proficient in reading has a 18.9 percent chance that his or her true reading level falls within one of the other categories: novice, apprentice, or distinguished.

Table 3.5
Student-level Classification Accuracy
(Probability That a Student is Correctly Classified as Novice, Apprentice, Proficient, or Distinguished Within a Subject)

School Level & Subject	Grade	2001	2002	2003
Elementary Schools				
Arts & Humanities	5	74.1	72.9	62.2
Mathematics	5	76.2	75.3	74.2
Practical Living	5	59.6	62.5	58.8
Reading	4	81.0	80.7	78.7
Science	4	76.8	76.3	76.6
Social Studies	5	72.5	73.0	71.4
Middle Schools				
Arts & Humanities	8	70.6	70.1	60.3
Mathematics	8	80.4	81.2	80.2
Practical Living	8	65.4	68.3	61.8
Reading	7	81.9	81.7	79.9
Science	7	73.8	74.1	73.1
Social Studies	8	80.3	80.8	79.5
High Schools				
Arts & Humanities	11	70.5	69.6	65.3
Mathematics	11	79.1	79.5	79.6
Practical Living	10	63.8	64.7	61.4
Reading	10	82.5	82.3	81.1
Science	11	79.0	77.9	77.2
Social Studies	11	82.1	80.9	80.1

Sources: HumRRO. "The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2001 and 2002 Kentucky Core Content Tests"; and HumRRO. "The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2003 Kentucky Core Content Tests."

The probability of a student being correctly classified was lowest for practical living and arts and humanities.

In 2003, classification accuracies were lowest in the practical living and arts and humanities areas. This was true for all levels of schools. For elementary students, the probability of being correctly classified in practical living was approximately 59 percent. Alternatively, this represents a 41 percent chance of being misclassified. In its discussion of classification accuracy, AEL noted that the assessments for practical living and arts and humanities both have fewer test items, which likely contributes to the lower level of accuracy.

Classification accuracy improved from 2002 to 2003 for elementary school students in science and for high school students in mathematics. The accuracy decreased, however, for all other subjects and levels.

AEL noted that there was no standard for determining the appropriate level for these probabilities.

AEL noted that there is no standard for determining the appropriate level of classification accuracy. To provide some context, AEL was able to compare the classification accuracy of KCCT scores to scores from two other states: Massachusetts and Maine. The classification accuracy for Maine's assessment ranged from 56 to 82 percent. In Massachusetts, the classification accuracy ranged from 67 to 84 percent in English and from 68 to 71 percent in mathematics.

Uses of Student-level Scores

Student-level scores suffer from lower levels of reliability than do school-level scores. Therefore, the use of student scores is limited.

Limitations are associated with the use of CATS scores at the student level. Scores at this level often suffer from less reliability than do school-level scores and might not accurately signal a student's true level of skills and knowledge. In addition, several concerns regarding the validity of CATS scores have been raised, particularly at the student level.

While there are limitations to their use, student-level scores do provide some signal of a student's achievements.

While the scores do suffer from several limitations, they can provide some signal of the student's skill and knowledge of the core content. This signal is far from perfect and can result in incorrect determinations of a student's achievements. Therefore, any use of CATS scores at the student level would require an understanding of the risk of incorrectly classifying a student and an acceptance that taking this risk is appropriate.

Both AEL and the National Technical Advisory Panel for Assessment Accountability (NTAPAA) suggested that decisions that result in greater consequences should require higher levels of reliability.

Two specific concerns were identified both in the literature reviewed by AEL and in the comments provided by the National Technical Advisory Panel on Assessment and Accountability (NTAPAA). The first is the level of consequences associated with a specific use of the scores. The second is the likelihood that the

scores provide an accurate indication of a student's ability. The research suggested that decisions that result in greater consequences should require higher levels of reliability to reduce the risk of incorrectly assigning these consequences. For example, in its review of the research, AEL cited the standards that were established by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education as stating that "the need for precision increases as the consequences of decisions and interpretations grow in importance" (29-30). Lower levels of reliability might be acceptable if the consequences are not deemed significant.

Whether an assessment is sufficiently reliable depends on the tolerance for risking incorrect decisions given the consequences that will be incurred.

Consequences might include modifying teaching plans, requiring additional evaluation of a student, or holding a student back a grade. In determining whether student-level scores should be used for a particular function, it is important to consider the level of consequences associated with the use and the risk of making an incorrect decision regarding the student. Consider, for example, whether it would be appropriate to determine whether a student would graduate from high school based on an assessment that correctly determines a student's true ability 50 percent of the time. If the consequence of a low score is further evaluation, then making an incorrect decision might not be too costly. Ultimately, whether an assessment is sufficiently reliable depends on the tolerance for risking incorrect decisions given the consequences that will be incurred. If an incorrect decision has serious consequences for the student, higher levels of reliability might be required.

HumRRO concluded that the probability of incorrectly classifying a student was highest for high school students in reading.

Based on the analysis provided to the Kentucky Department of Education by HumRRO, the probability of an incorrect classification of a student was lowest for high school students in reading, which was approximately 19 percent. The student's true level of performance might be higher or lower than indicated by the CATS scores ("The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2003").

NTAPAA suggested that student-level CATS scores can be used for certain purposes when considered in the context of additional measures of student performance.

NTAPAA suggested that CATS scores at the individual level might serve certain purposes. The group cautioned that the scores are not sufficiently reliable to stand on their own and must be used in the context of additional measures of student performance. In addition, when determining whether a use was appropriate, NTAPAA made a distinction based on the level of consequences. The group suggested that there is too much error in individual student scores to make decisions that would have "high-stakes" consequences (3). AEL provided a similar caution, stating that a

single test is not adequate for important decisions about an individual student (64).

NTAPAA indicated that some uses of student-level scores would be appropriate. It did not indicate that including student-level CATS scores on transcripts would be sufficiently valid.

NTAPAA was asked whether certain uses of student-level scores were appropriate. Uses that NTAPAA suggested are appropriate include helping determine the Kentucky Educational Excellence Scholarship awards, determining whether a student needs additional assistance, and informing educators and parents of a student's academic performance. The group did not conclude that reporting students' CATS scores on transcripts would be a valid use. The panel stated that: "Because we could not anticipate all uses of KCCT test information — namely individual student subject scores — posted on transcripts, we could not judge the practice to be valid" (National 3).

It is important to note that with each of these uses, NTAPAA indicated that the CATS scores should be used along with other measures of student performance. Relying on several measures should reduce the risk of inaccurately determining a student's level of performance. If a score on one area of the core content appeared to be inconsistent with other measures of performance such as class grades, the CATS score might be given less weight.

Possible Changes To Improve Validity

NTAPAA provided suggestions for increasing the reliability and validity of student-level scores.

The primary limitations to using CATS scores at the student level are coverage of the core content, the number of items students are required to answer, and the consistency with which raters score certain items. NTAPAA made some suggestions for increasing the reliability and validity of student-level CATS scores. The research reviewed by AEL also included ways to improve the reliability and validity.

Increasing the number of tasks a student must perform would likely improve both reliability and validity.

Increasing the number of tasks students must perform or questions students must answer would serve to improve both validity and reliability. Validity would be improved because students would be tested on a greater portion of the core content, reducing the probability that students just happen to get questions in areas in which they are particularly weak or strong. The research reviewed by AEL frequently indicated that increasing the number of tasks also improved the reliability of assessments by reducing the probability that random occurrences affect scores. Both methods would make the scores better indications of a student's true achievements.

NTAPAA suggested expanding the language and math tests to cover the entire core content and allowing students to retake the tests to improve their reliability and validity.

NTAPAA suggested that the language and math tests might be expanded to cover the entire core content for these areas and that students be allowed to retake the tests. Retaking the tests would provide students who performed poorly an additional opportunity. The panel suggested that these changes could result in an assessment comparable to the ACT or SAT with regard to reliability and validity.

NTAPAA did caution that there would be additional cost associated with these changes. CATS was specifically designed as a school-level assessment in part due to concerns over the amount of time it would take to test students and the cost to administer the assessments.

AEL's review of the research suggested that increasing the number of raters could improve the consistency of scores and improve reliability.

Other changes might address the effect that inconsistencies among raters have on student-level reliability. The research reviewed by AEL suggested that increasing the number of raters for the on-demand questions and the writing portfolios might reduce inconsistencies across raters. The studies often found that improvements in reliability from increasing the number of raters were lower than from increasing the number of tasks.

Even with extensive changes, there will always be some chance that a student is misclassified as a result of an assessment.

While there are several options available for improving the reliability and validity of CATS, it should be noted that there will always be some probability that a student will be misclassified. The issue then becomes whether the probability can be reduced to a level deemed acceptable for the intended uses of the scores. And as the reliability of student-level scores is improved, the reliability of school-level scores would likely improve as well.

Conclusions

One of the goals of assessments is to make inferences about the performance of a student or a group of students. For these inferences to be useful as a basis for making decisions, the assessments must be sufficiently reliable and valid indicators of the student's performance or the group's performance.

Both AEL and NTAPAA suggested that higher levels of reliability and validity are needed if scores will be used to make decisions that are deemed significant. Ultimately, the level of reliability and validity needed for a specific use depends on the tolerance for making an incorrect decision.

The research did not provide a clear indication of what level of reliability and validity must be achieved. Instead, the research suggested that the needed level of reliability and validity varies depending on how the scores will be used. Both AEL and NTAPAA suggested that higher levels of reliability are needed if the consequences are deemed to be significant. Ultimately, the level of reliability and validity needed for a specific use depends on the tolerance for making an incorrect decision.

HumRRO estimated that the probability of correctly classifying a school was 77 percent. It also concluded that the probability of correctly classifying a student ranged from 58.8 percent to 81.1 percent depending on the school level and content area.

HumRRO estimated that these is a 77 percent probability of correctly assigning a school to a particular classification, such as meets goal or progressing ("The Accuracy of School"). This indicates that out of 100 schools, approximately 77 would be correctly classified and that 23 would be incorrectly classified. HumRRO estimated similar statistics for individual student-level classifications ("The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2001 and 2002"). The probability of correctly classifying a student varied by grade and content area. The probabilities ranged from 58.8 percent in practical living studies for elementary school students to 81.1 percent in reading for high school students.

NTAPAA concluded that CATS scores were sufficiently reliable and valid at the school level but that student-level scores were only sufficiently reliable and valid for certain uses and only in the context of other measures of student performance.

NTAPAA concluded that CATS scores were sufficiently reliable and valid for use at the school level. The panel concluded that the following uses of student-level CATS scores would be sufficiently reliable and valid: helping determine Kentucky Educational Excellence Scholarship awards, determining whether a student needs additional assistance, and informing educators and parents of a student's academic performance. However, NTAPAA did not conclude that student-level scores were sufficiently reliable and valid to be included on students' transcripts, as they could not determine how the scores might then be used.

Chapter 4

A Summary of the Surveys of Teachers, Principals, Superintendents, Students, Parents and Guardians, and School Board Members

Introduction

In 2004, LRC contracted with AEL, which conducted focus groups and surveys of teachers, principals, superintendents, high school students, parents and guardians, and school board members in Kentucky. Focus groups were conducted in six regions across the state for the primary purpose of providing information that was used to create questions for the surveys.

In 2004, LRC contracted with AEL, which conducted focus groups and surveys of K-12 teachers, principals, superintendents, students in grades 10 through 12, parents and guardians of students in grades 4 through 12, and school board members. Focus groups were conducted in six regions across the state for the primary purpose of providing information that was used to create questions for the surveys. In consultation with AEL staff, LRC personnel constructed a questionnaire for each group of respondents. AEL then mailed survey packets to randomly selected teachers, students, parents and guardians, and school board members in November 2004. For teachers, LRC staff asked that the sample be stratified to assure that the returned surveys would include adequate numbers of special education teachers and teachers of students with limited English proficiency. The questionnaires for principals and superintendents were made available via Web pages. In November 2004, AEL sent emails to randomly selected principals and to all superintendents in Kentucky describing the survey and providing a link to the relevant Web page so that the questionnaire could be completed online.

The number of respondents and response rates—percentage of surveys returned—were relatively low for most groups.[1] Based on the number of returned questionnaires, if the surveys are representative, the margins of error range from plus or minus 5.3 percentage points (teachers) to plus or minus 9.5 (parents and guardians). A margin of error of 5.3 would mean that if the sample indicates that 60 percent gave a particular answer to a question,

[1] The number of returned questionnaires, response rates, and margins of error are as follows: teachers, 337 returned surveys, 29 percent response rate, +/- 5.3 margin of error; principals, 230, 38 percent, +/- 5.9; superintendents, 107, 61 percent, +/- 6.0; students, 159, 26 percent, +/- 7.8; parents and guardians, 106, 17 percent, +/- 9.5; and school board members, 149, 28 percent, +/- 7.3. The margins of error are calculated based on a confidence level of 95 percent and the assumption that the true breakdown of responses to a question is 50%/50%.

there is a 95 percent probability that the true answer lies between 54.7 and 65.3. Within each group of respondents, the margin of error for any given question will be higher the fewer the number of respondents answering that question.

The remainder of this document consists of summaries of the results of the surveys. The format is for each section to cover the most relevant questions from the surveys of all groups asked about a particular subject. The subjects covered are

- question formats;
- validity of Kentucky Core Content Tests for special education students, for students with limited English proficiency, and for all other students;
- utilization of CATS test results;
- integration of writing portfolios into a student's learning experience;
- validity and value of writing assessments for different types of students;
- the effect of CATS testing on curriculum, instruction, and learning; and
- student CATS reports and school report cards.

When the wording of a question is not clear from the description in the text, a footnote provides the exact wording from the questionnaire. Many of the survey results are covered using tables. When the numbers of respondents to the questions under discussion are not given in the table, they can be found in the note below the table. Appendix B contains the results of the closed-ended questions from all six surveys.

Question Formats

Almost all teachers, principals, and superintendents who responded indicated that multiple-choice questions were appropriate for grades 4 and 5, 7 and 8, and 9 through 12. More than 75 percent of each group rated open-response and on-demand writing as appropriate for grades 7 and 8, and 9 through 12. Nearly half of the teachers and principals did not agree that on-demand writing was appropriate for 4th and 5th graders.

Teachers, principals, and superintendents were asked whether multiple-choice and open-response questions and on-demand writing were appropriate for three grade levels: 4 and 5, 7 and 8, and 9 to 12.[2] More than 90 percent of teachers replied that multiple-choice questions were appropriate for all three grade levels, and 100 percent of principals and nearly 100 percent of superintendents agreed. The three groups of respondents were also in agreement that the other two question formats were appropriate for the middle and higher grades. More than 80 percent of each group responded that the open-response and on-demand formats

[2] The question was "Are the following test formats appropriate for the grade levels indicated?" The response categories were "Yes" and "No."

were appropriate for grades 9 through 12. More than 75 percent reported that both formats were appropriate for grades 7 and 8.

There was no consensus, however, about open-response questions and on-demand writing for those in grades 4 and 5. More than three-fourths of principals and superintendents reported that open-response was appropriate at this level; only 64 percent of teachers agreed. Nearly half of the teachers and principals responding to the surveys did not agree that on-demand writing was appropriate for 4th and 5th graders. Three-fourths of superintendents said that on-demand writing was appropriate for grades 4 and 5, but this was the lowest level of support from the group for any of the three question formats.

Validity of Kentucky Core Content Tests

Validity for Special Education Students

Teachers, principals, and superintendents were asked whether CATS testing disadvantaged particular groups of students.[3] Assuming they were taught the same content as other students, more than half of teachers (65 percent), principals (58 percent) and superintendents (55 percent) responded that many special education students were disadvantaged. Less than a fifth of each group said that none or almost none of special education students were disadvantaged.

Teachers, principals, and superintendents were asked whether six KCCT subject tests were valid measures of core content knowledge for special education students. A majority of each group of educators responded that the six subject tests were invalid measures for these students.

Other questions addressed the validity of six Kentucky Core Content subject tests for special education students.[4] Table 4.1 below shows the percentages of each group of educators that agreed that a subject test was valid for special education students.[5]

[3] The question was "Assuming they are taught the same content of the tested subjects as other students, are any of the following types of students [includes 'Special Education Students' and 'Students With Limited English Proficiency'] disadvantaged by CATS testing?" The response categories were "None or almost none," "A few," and "Many."

[4] The question was "For each CATS test and type of student [includes 'Special Education Students'] with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' knowledge of the core content." The response categories were "Strongly disagree," "Disagree," "Agree," "Strongly agree," and "Not sure."

[5] In discussing the results of such questions, "Strongly disagree" will typically be grouped with "Disagree"; and "Strongly agree" will be grouped with "Agree."

Table 4.1
Percentages of Educators Agreeing That KCCT Subject Tests Were Valid for Special Education Students

Subject Test	Teachers	Principals	Super-intendents
Reading	25.3	32.7	42.7
Math	29.0	39.4	47.1
Science	26.9	38.9	44.8
Social Studies	27.8	39.3	43.2
Arts & Humanities	24.1	33.8	38.4
Practical Living/Vocational Studies	29.5	37.9	43.0

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by subject test. The number of teachers ranged from 220 to 241, the number of principals from 206 to 211, and the number of superintendents from 86 to 89.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

The most notable finding is that there was not a majority of any of the three groups of educators that agreed that any subject test was valid for special education students. Within each group, there was little distinction made among the tests. Less than 30 percent of teachers, less than 40 percent of principals, and less than half of the superintendents who responded agreed that the tests were valid. Put another way, a majority of educators that answered the questionnaires said that all six subject tests were invalid measures of special education students' knowledge of the core content.

In one sense, these results may even understate the lack of confidence of educators in the validity of KCCT subject tests for special education students. For most questions on the survey, there were relatively low percentages of respondents answering "Strongly disagree" or "Strongly agree." That was only partly the case for this set of questions. Few strongly agreed, but relatively large percentages strongly disagreed. The results varied somewhat by subject, but approximately 29 percent or more of teachers, 19 to 24 percent of principals, and 16 to 19 percent of superintendents strongly disagreed that any given subject test was valid for special education students.

More than half of principals and superintendents disagreed that the alternate portfolio was a valid measure of knowledge of core content for special education students.

Validity of the Alternate Portfolio. Students with moderate to severe disabilities for whom traditional assessments are inappropriate are assessed via the Kentucky Alternate Portfolio. Principals and superintendents were asked about their level of disagreement or agreement that the alternate portfolio was a valid

measure of knowledge of the core content.[6] More than 10 percent of principals and more than 20 percent of superintendents replied that they were not sure. Excluding these respondents, just more than half of each group disagreed or strongly disagreed that the alternate portfolio was a valid measure of knowledge of the core content. Compared to responses to most of the surveys' questions, the shares of respondents that strongly disagreed were high: more than 18 percent of each group.

Validity for Students With Limited English Proficiency

There were corresponding survey questions regarding students with limited English proficiency. Overall, educators were even less positive about CATS testing for these students than for special education students. Assuming these students were taught the same content as other students, more than three-fourths of teachers and principals answered that many students with limited English proficiency were disadvantaged. More than two-thirds of superintendents agreed. Less than a tenth of each group said that none or almost none of these students were disadvantaged.

Teachers, principals, and superintendents were also asked whether six KCCT subject tests were valid measures of core content knowledge for students with limited English proficiency. A majority of each group of educators responded that the six subject tests were invalid measures for these students.

Table 4.2 indicates the percentages of each group of educators who agreed or strongly agreed that a subject test was valid for students with limited English proficiency. For most tests, one-fourth or fewer of educators agreed or strongly agreed that the subject test was valid for students with limited English proficiency. There was no test that a majority of teachers, principals, or superintendents agreed was valid.

As was the case with the corresponding questions regarding special education students, the shares of educators strongly disagreeing that the tests were valid for students with limited English proficiency were relatively high. Among teachers, 27 to 39 percent strongly disagreed, as did 32 to 40 percent of principals and 22 to 25 percent of superintendents.

[6] Excluding those who answered "Not sure," 187 principals and 80 superintendents replied to the question. The question was omitted inadvertently from the teacher questionnaire.

Table 4.2
Percentages of Educators Agreeing That KCCT Subject Tests Were Valid for Students With Limited English Proficiency

Subject Test	Teachers	Principals	Super-intendents
Reading	11.8	16.1	25.4
Math	27.1	14.8	28.6
Science	14.9	30.1	36.7
Social Studies	13.2	17.1	20.9
Arts & Humanities	13.0	17.8	20.9
Practical Living/Vocational Studies	14.0	18.6	22.7

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by subject test. The number of teachers ranged from 150 to 155, the number of principals from 123 to 149, and the number of superintendents from 60 to 71.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

Validity for All Other Students

Teachers, principals, and superintendents were also asked whether six CATS subject tests were valid measures of knowledge of the core content for all other students. At least two-thirds of superintendents agreed that all six tests were valid for elementary, middle, and high school students. This was also true for approximately 60 percent or more of principals. Majorities of teachers responded that most of the tests were valid, but support for this position was less widespread than for the other two groups.

Surveys of Teachers, Principals, and Superintendents. Teachers, principals, and superintendents were asked about the validity of six CATS subject tests for elementary, middle, and high school students.[7] As shown in Table 4.3, educators had more confidence in the validity of some subject tests than others.

At least two-thirds of superintendents agreed or strongly agreed that all six tests were valid for elementary, middle, and high school students. This was also true for approximately 60 percent or more of principals. Majorities of teachers responded that most of the tests were valid, but support for this position was less widespread than for the other two groups.

[7] The question was "FOR ALL OTHER STUDENTS: For each CATS test and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' knowledge of the core content." The types of students were elementary, middle, and high school. The response categories were "Strongly disagree, "Disagree," "Agree," "Strongly agree," and "Not sure."

Table 4.3
Percentages of Educators Agreeing That KCCT Subject Tests Were Valid for All Other Students

Subject Test/Level	Teachers	Principals	Super-intendents
Reading			
Elementary	59.9	79.1	84.8
Middle	57.7	70.5	85.9
High	57.9	71.0	82.1
Math			
Elementary	58.6	74.3	86.9
Middle	57.4	69.2	82.8
High	59.6	69.3	77.9
Science			
Elementary	55.3	73.5	84.7
Middle	55.0	66.7	82.8
High	59.4	68.4	81.1
Social Studies			
Elementary	54.3	74.1	86.6
Middle	55.0	68.3	84.0
High	58.8	69.0	80.6
Arts & Humanities			
Elementary	44.7	57.5	72.8
Middle	46.7	60.2	73.0
High	51.5	63.2	70.3
Practical Living/ Vocational Studies			
Elementary	46.2	67.6	71.3
Middle	49.5	62.2	69.0
High	52.9	62.9	69.9

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by subject test and level. The number of teachers ranged from 101 to 177, the number of principals from 95 to 187, and the number of superintendents from 84 to 99.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

Although the numbers varied by level of student, each group seemed to distinguish the arts and humanities and practical living and vocational tests from the other four tests. For superintendents, typically the percentages responding that the tests were valid were at least 10 percentage points higher for reading, math, science, and social studies. The distinction between the two groups of tests was not as strong for principals, but for any level of student, more principals responded that the first four subjects were valid.

Teachers also made the distinction between the two groups of tests. Given the teachers' overall lower level of agreement than that of principals and superintendents, this meant that for some tests/levels, a majority of teachers did not agree that the tests were valid. Less than half of the teachers agreed or strongly agreed that the arts and humanities test and the practical living and vocational test were valid for elementary or middle school students.

Approximately 60 percent or more of surveyed high school students agreed each test was a good measure of their knowledge of the subject. More than 60 percent of surveyed parents and guardians indicated that CATS tests were fair measures of their children's knowledge of school subjects.

Surveys of High School Students and Parents and Guardians. Students in grades 10 through 12 and parents and guardians were also asked about KCCT. The questions were phrased differently than for educators. For each of six subject tests, students were asked about the extent to which they disagreed or agreed that the test was a "good measure of your knowledge of that subject." Excluding the students who answered "Not sure," 59 percent or more agreed or strongly agreed that each test was a good measure of knowledge. Students did distinguish somewhat among the tests. For the reading, math, and social studies tests, at least 73 percent agreed or strongly agreed that each test was a good measure. For science, 66 percent did so. Consistent with the responses of educators, the lowest levels of agreement or strong agreement were for arts and humanities (59 percent) and practical living and vocational studies (62 percent).

Parents and guardians were asked—"Yes," "No," or "Not sure"—whether the CATS tests were "fair measures of your child's knowledge of school subjects," with no references to specific subject tests. Almost a third answered that they were not sure. Excluding those who were not sure, 61 percent responded "Yes."

Utilization of CATS Test Results

Currently, the CATS testing system is designed to ensure school accountability, not accountability for the individual student. Students' test results, however, are available in detail to school personnel and in summary format for students and parents and guardians. The surveys of all these groups contained questions about how useful the respondents found the information provided.[8]

[8] The questions for educators were whether they agreed or disagreed that "CATS testing provided useful information about how well students are doing" and "CATS testing provides useful information about how well schools are doing." The student question was "CATS testing provides useful information about how well I'm doing in school" ("Agree/Disagree"). The parent questions were "The CATS tests are a good way to measure how my child is doing in school" ("Agree/Disagree") and "Does the School Report Card help you understand how your child's school is doing?" ("Yes/No/Not sure").

Note that respondents were asked about the usefulness of the test results, not whether students should be held accountable based on their test scores. A summary of the responses is below.

Table 4.4
Percentages Agreeing That CATS Tests Provided Useful Information About Students and Schools

How well:	Teachers	Principals	Super-intendents	Students	Parents/ Guardians
student is doing	39.6	55.4	68.2	49.1	45.9
school is doing	29.9	55.1	69.2	Not asked	61.4*

*For consistency with the other results shown in the table, the calculation excludes those who answered "Not sure."
Note: There were 107 superintendents answering each question, and 159 students answered their one question. The number of teachers ranged from 318 to 321, the number of principals from 224 to 227, and the number of parents and guardians from 88 to 98.
Source: *2004 Surveys of Kentucky Teachers, Principals, Superintendents, Students, and Parents and Guardians*

The educators surveyed did not distinguish between the usefulness of CATS test scores for determining how students and schools were doing. Less than 40 percent of teachers agreed that the results were useful for either purpose, but 55 percent of principals and at least 68 percent of superintendents agreed the results were useful for each purpose. Just under half of the parents and guardians who responded indicated that CATS tests were good measures of how their children were doing in school. More than 60 percent replied that school report cards helped them understand how their children's schools were doing.

The educators surveyed did not make much of a distinction in the usefulness of CATS test results for students or schools. Less than 40 percent of teachers agreed that the results were useful for either, as did approximately 55 percent of principals. Approximately two-thirds of superintendents agreed that test results were useful indicators of how students and schools were doing.

Just under half of the students who responded indicated the results provided useful information about how they were doing in school. Parents and guardians were asked the basic question in a different format: whether the tests "are a good way to measure how my child is doing in school." The question about the utility of CATS results for schools asked about the school report card, which provides information in addition to test scores. Keeping in mind the question format, this was the only group that distinguished a difference in the value of test results for students and schools. Just under half indicated that CATS tests were good measures of student performance. Excluding the share of parents and guardians who answered "Not sure," more than 60 percent responded that school report cards helped them understand how their child's school was doing.

Student Effort on the Tests

The level of effort by students on CATS tests is relevant to the usefulness and potential utilization of results. It is possible that any given test of students' knowledge is valid in the sense that the material covered on the test closely matches the material that students were supposed to learn. However, if, for whatever reason, students do not take the test as seriously as educators would like, then the validity, reliability, and usefulness of the results are questionable. The question of the level of student effort on CATS tests is a difficult one to answer. Any results from surveys are likely to be suggestive only.

Educators and students were asked about the level of effort made by students on the tests. The question for teachers, principals, and superintendents was "What share of students try their best on CATS tests?" The student question was "What percent of students try their best on CATS tests?" The complete results are below.

Table 4.5
Educators' and Students' Perceptions of Shares of Students Trying Their Best on CATS Tests

Response	Teachers	Principals	Super-intendents	Students
None or almost none	1.5	0.4	1.9	8.9
Some	48.2	22.7	34.9	52.5
Most	43.9	49.8	54.7	30.4
All or almost all	5.2	27.1	7.5	6.3
Not sure	1.2	0.0	0.9	1.9
Total	100.0 %	100.0 %	100.0 %	100.0 %
Number of respondents	326	225	106	158

Source: *2004 Surveys of Kentucky Teachers, Principals, and Students*

Approximately half of the surveyed teachers and students indicated that some students made their best effort. Less than half replied that most, nearly all, or all students did so.

More so than for almost every other question on the survey, respondents had clear views on this issue, as indicated by the few answering "Not sure." The teachers and students who responded had the more negative perceptions of student effort. Approximately half indicated that some students made their best effort. Less than half replied that most students did so. Less than 7 percent of teachers and students responded that all or almost all made their best effort.

Approximately half or more of the principals and superintendents who responded to the surveys answered that most students try their best. Less than 10 percent of superintendents reported that all or almost all students made their best effort; 27 percent of principals gave this response.

Approximately half or more of the principals and superintendents who responded to the surveys answered that most students try their best. Less than 8 percent of superintendents reported that all or almost all students made their best effort; 27 percent of principals gave this response.

The results differed when students were asked to "describe your effort on CATS tests." Approximately 10 percent indicated that they never, rarely, or usually did not try their best. Approximately 30 percent reported usually trying their best; 60 percent indicated they tried their best for all the tests.

Integration of Writing Portfolios Into a Student's Learning Experience

The degree to which writing portfolios are integrated into a student's learning experience can be thought of as a continuum. The least integrated alternative is that a student would work on the portfolio with one teacher, in one class designated for writing instruction, and during a block of time during the school year set aside for portfolios. If the portfolio were fully integrated, it would be a part of the student's everyday educational experience. The student would be working with more than one teacher in several classes and would be working on the portfolio throughout the school year.

Educators and students were asked three questions about the integration of writing portfolios into a student's learning experience. There was broad agreement that work on portfolios was done throughout the school year, rather than during a block of time set aside for them.

Educators and students were asked three questions to assess these aspects of integration. The wording of the questions and the results are shown in Table 4.6. The results are mixed. There was broad agreement that work on portfolios was done throughout the school year. The groups were divided as to whether work was done under the supervision of one teacher or more than one.

Teachers, principals, superintendents, and students were in agreement about when work on portfolios was done (the second question in the table). More than three-fourths of each group responded that work was done throughout the school year and not during a block of time set aside for emphasis on portfolios.

Table 4.6
Integration of Writing Portfolios
(Surveys of Educators and Students)

	Teachers	Principals	Super-intendents	Students
Students' work on portfolios is mostly done as part of writing assignments in several classes OR in English or writing classes?				
Several classes	56.2%	63.4%	55.2%	79.6%
English or writing classes	43.8%	36.6%	44.8%	20.4%
Total	100.0%	100.0%	100.0%	100.0%
Number of Respondents	315	224	105	157
Students' work on portfolios is mostly done throughout the school year OR during a block of time during the year set aside for emphasis on portfolios?				
School year	89.6%	92.9%	77.1%	80.3%
Block of time	10.4%	7.1%	22.9%	19.7%
Total	100.0%	100.0%	100.0%	100.0%
Number of Respondents	318	225	105	157
A student working on a portfolio is more likely to work with only one teacher OR more than one teacher?				
Only one teacher	33.9%	30.6%	49.0%	51.3%
More than one teacher	66.1%	69.4%	51.0%	48.7%
Total	100.0%	100.0%	100.0%	100.0%
Number of Respondents	319	222	104	156

Source: *2004 Surveys of Kentucky Teachers, Principals, Superintendents, and Students*

Approximately 80 percent of students responded that work on portfolios was done in several classes. More than half of educators agreed, but 37 to 45 percent of each group responded that work was done mostly in English or writing classes.

There was no consensus on either of the other aspects of integration. Approximately 80 percent of students responded that work on portfolios was done in several classes. A majority of each of the other surveyed groups agreed, but verification of the more integrative position was not as widespread. Substantial minorities, from 37 to 45 percent, replied that work was mostly done in English or writing classes.

Approximately two-thirds each of teachers and principals reported that a student working on a portfolio worked with more than one teacher, rather than just one. Students' and superintendents' responses were nearly evenly divided on the question.

For the third question, approximately two-thirds each of teachers and principals reported that a student working on a portfolio worked with more than one teacher, not just one. However, students' and superintendents' responses were nearly evenly divided on the question.

Parents and guardians were also asked about their level of satisfaction with how their children's writing portfolio fit with work done in different classes.[9] Fifteen percent reported that they were not sure. Among those who made an assessment, 55 percent were dissatisfied or strongly dissatisfied with the integration of the portfolio into class work.

Validity and Value of Writing Assessments

The surveys included several questions that called for respondents to address the validity of writing portfolios and their value for students. Students are also assessed on the quality of their writing through the on-demand writing portion of the core content tests, so educators and students were also asked about the validity of this type of assessment. Given that both assessments are designed to evaluate students' writing, the survey results on portfolios and on-demand writing will both be covered in this section of the report.

Validity of On-demand Writing and Writing Portfolios

Educators were asked whether writing portfolios and on-demand writing were valid measures of students' writing. In regard to special education students, educators were also asked about the alternate portfolio.[10]

Validity for Special Education Students. The following table presents the percentages of each group that agreed or strongly agreed that portfolios, on-demand writing, and the alternate portfolio were valid for special education students.

[9] The question was "To what extent are you satisfied or dissatisfied that your child's writing portfolio fits in with work done in several classes?" The response options were "Very satisfied," "Somewhat satisfied," "Somewhat dissatisfied," "Very dissatisfied," and "Not sure." There were 105 responses.

[10] The question was "For each writing assessment and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' writing." The types of students were "Special Education Students," "Students With Limited English Proficiency," and "All Other Students." Response categories ranged from "Strongly disagree" to "Strongly agree," plus "Not sure."

Table 4.7
Percentages of Educators Agreeing That Writing Assessments Were Valid Measures of Special Education Students' Writing

Writing Assessment	Teachers	Principals	Super-intendents
Writing portfolio	26.0	26.4	41.7
On-demand writing	36.0	28.2	43.2
Alternate portfolio	39.5	45.1	44.8

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by assessment. The number of teachers ranged from 162 to 261, the number of principals from 182 to 220, and the number of superintendents from 87 to 96.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

Less than half of teachers, principals, and superintendents agreed that the writing portfolio, on-demand writing, or the alternate portfolio was a valid measure of special education students' writing.

More teachers and principals replied that the alternate portfolio was valid as opposed to the writing portfolio. However, there was not a majority among any of the groups of educators that responded that any of the three assessments was a valid measure of student's writing. More than 30 percent of teachers strongly disagreed that any of the three assessments was valid. At least 18 percent of principals and superintendents also strongly disagreed that any of the three assessments was valid.

Less than 40 percent of teachers, principals, and superintendents agreed that either the writing portfolio or on-demand writing was valid for students with limited English proficiency.

Validity for Students With Limited English Proficiency. The percentages of teachers (18.5 percent), principals (29.4 percent), and superintendents (25.9 percent) indicating that the writing portfolio was a valid measure of the writing of students with limited English proficiency were also low. As shown in Table 4.8, the responses were somewhat more positive for on-demand writing, but there was not more than 40 percent of any group indicating that on-demand writing was a valid measure for these students. As for special education students, the percentages strongly disagreeing were high compared to most other questions on the surveys.

Table 4.8
Percentages of Educators Agreeing That Writing Assessments Were Valid Measures of the Writing of Students With Limited English Proficiency

Writing Assessment	Teachers	Principals	Super-intendents
Writing portfolio	18.5	29.4	25.9
On-demand writing	21.8	37.1	38.8

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by assessment. The number of teachers ranged from 156 to 157, the number of principals from 175 to 177, and the number of superintendents from 81 to 85.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

Validity for All Other Students. The percentages of teachers, principals, and superintendents agreeing or strongly agreeing that a type of writing assessment is appropriate for a particular level of school are displayed in Table 4.9. There are clear distinctions between types of assessment and school level and among the groups of educators.

For all other students, a majority of superintendents rated portfolios and on-demand writing as valid at all levels. More of them had confidence in the validity of on-demand writing, however.

A majority of superintendents rated portfolios and on-demand writing as valid at all levels. Many more of them had confidence in the validity of on-demand writing, however. Just more than half rated portfolios as valid; approximately three-fourths replied that on-demand writing was valid for all levels.

Teachers and principals distinguished between school levels and type of assessment. There was no majority of either group agreeing that portfolios were valid for any school level.

Teachers and principals distinguished between school levels and type of assessment. There was no majority of either group agreeing or strongly agreeing that portfolios were valid for any school level. For each group, the percentages were lower for elementary students than for high school students.

More than half of teachers and principals agreed that on-demand writing was a valid means of assessment for students at all levels. The percentages indicating agreement were higher for the middle and high school levels than for elementary students.

A similar pattern among teachers and principals held for on-demand writing. More agreed that it was a valid means of assessment for middle and high school students than for elementary students. Even for elementary school students, however, more than half of each group responded that on-demand writing was a valid measure of students' writing.

Table 4.9
Percentages of Educators Agreeing That Writing Assessments Were Valid Measures of Writing for All Other Students

Type of Assessment/ Level	Teachers	Principals	Super-intendents
Writing Portfolios			
Elementary	22.7	35.1	52.3
Middle	32.6	37.9	54.5
High	44.3	42.6	52.7
On-demand Writing			
Elementary	54.1	57.4	75.3
Middle	60.5	67.6	78.7
High	62.8	68.0	72.3

Note: Percentages were calculated with those answering "Not sure" excluded. The number of respondents varied by assessment. The number of teachers ranged from 129 to 181, the number of principals from 101 to 169, and the number of superintendents from 86 to 94.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

Two-thirds of students agreed that the writing portfolio was a "good measure" of their writing. The corresponding figure for on-demand writing was 59 percent.

The survey responses of students were much more positive than those of educators or parents and guardians in regard to portfolios and on-demand writing as measures of their own writing.[11] Excluding those who answered "Not sure," 68 percent of students agreed or strongly agreed that the writing portfolio was a "good measure" of their writing. The corresponding figure for on-demand writing was 59 percent.

Parents and guardians were similar to superintendents in their evaluations of portfolios as a measure of writing. Just more than half (53 percent) responded that the portfolio was a "good measure" of writing.

[11] The question to students was "Based on your own experience, to what extent do you agree that the writing portfolios and on-demand writing have been good measures of how well you wrote at that time?" There were separate response sets ("Strongly disagree" to "Strongly agree," plus "Not sure") for each type of writing. Excluding those who answered "Not sure," 151 students responded to the portfolio item and 140 to the on-demand item. The question to parents was "Do you think your child's score on the writing portfolio is a good measure of how well your child writes?" The possible responses were "Yes," "No," and "Not sure." Excluding those who answered "Not sure," 86 parents answered the question.

Value of Writing Portfolios

The discussion of the survey results regarding writing portfolios has centered on the assessments' validity for measuring students' writing. Another relevant dimension is the overall worth of portfolios, which can be distinguished from validity. Respondents may be confident that a type of assessment is a valid and reliable measure, but they may still not approve of it. For example, even a valid assessment may be thought to take up too much time. Educators may also rate an assessment as valid and reliable, but they may disagree with its weight in the school accountability index. Survey recipients were asked questions addressing both these indicators of the overall value of writing portfolios.

Educators were asked whether they agreed or disagreed "that the amount of time it takes to prepare writing portfolios is appropriate to the benefit received by students." More than 80 percent each of teachers and principals and 77 percent of superintendents disagreed or strongly disagreed that portfolios were worth the time. More than one-half of the principals strongly disagreed, and nearly one-half of teachers did so.

Is Benefit Appropriate to Costs? Educators were asked whether they agreed or disagreed "that the amount of time it takes to prepare writing portfolios is appropriate to the benefit received by students." The complete results are presented below.

Table 4.10
Percentages of Educators Disagreeing or Agreeing That Writing Portfolios Were Worth the Time

Response	Teachers	Principals	Super-intendents
Strongly disagree	48.9	54.9	39.4
Disagree	31.6	27.9	37.5
Agree	12.4	11.9	19.2
Strongly agree	1.5	4.0	1.9
Not sure	5.6	1.3	1.9
Total	100.0 %	100.0 %	100.0 %
Number of Respondents	323	226	104

Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents.*

Compared to the results of almost every other question on the survey, this one is distinguished by the low percentages of unsure respondents and the intensity of responses. Large shares of all three groups disagreed or strongly disagreed that portfolios were worth the time: more than 80 percent each of teachers and principals and 77 percent of superintendents. More than half of the principals strongly disagreed, and nearly half of teachers did so.

Parents and guardians shared the negative view of the portfolio's value. They were asked whether the benefit was "worth the amount

of time it takes."[12] Slightly more than one-third agreed or strongly agreed that portfolios were worth the time. The most common single response was "Strongly disagree" at 40 percent.

Students were the one group for which a majority (58 percent) agreed or strongly agreed that the benefit was appropriate given the time portfolios take.[13]

Teachers, principals, superintendents, and school board members were asked to indicate whether the weights for the KCCT subject tests, writing portfolios, and the norm-referenced CTBS/5 test were about right, should be increased, or should be decreased; or whether the item should be eliminated from the index

Assessments' Weights in the Accountability Index. Another indicator of the overall value of the writing assessments is survey respondents' views on how the assessments should be weighted in the school accountability index. Teachers, principals, superintendents, and school board members were asked to indicate whether the weights for the KCCT subject tests, writing portfolios, and the norm-referenced CTBS/5 test were about right, should be increased, or should be decreased; or whether the item should be eliminated from the index. Respondents could also indicate that they were not sure. The following table shows the percentages of each group that responded that the item should be eliminated or decreased.

The responses from teachers, principals, superintendents, and school board members were generally consistent for each item. Based on the survey responses, the index items can be separated into three groups. The writing portfolio was the only item that a majority of each group of respondents would reduce or eliminate. The percentages ranged from 54 percent for school board members to 72 percent for superintendents. For teachers and principals, there were more respondents indicating that portfolios should be eliminated than weighted less.

[12] The question was "To what extent do you agree or disagree that the benefit your child receives is worth the amount of time it takes to prepare writing portfolios?" The possible responses ranged from "Strongly agree" to "Strongly disagree," plus "Not sure." Excluding those answering "Not sure," there were 88 parents who answered the question.

[13] Excluding those who responded "Not sure," 145 students answered the question.

Table 4.11
Percentages of Respondents Agreeing That an Item Should Be Weighted Less or Eliminated From the School Accountability Index

Item	Teachers	Principals	Super-intendents	School Board Members
Writing portfolios	62.0	66.1	72.1	54.1
Arts & Humanities	35.5	40.2	37.9	36.3
On-demand Writing	43.7	45.2	34.0	27.2
Practical Living/ Vocational Studies	27.7	26.6	38.5	28.8
Social Studies	13.3	9.1	6.9	9.6
Science	12.6	10.0	5.8	6.8
CTBS/5	13.1	3.3	4.9	6.2
Reading	10.9	2.7	2.9	0.7
Math	9.6	2.8	1.9	1.4

Note: The number of respondents varied by item. The number of teachers ranged from 283 to 303, the number of principals from 214 to 221, the number of superintendents from 102 to 104, and the number of school board members from 145 to 148.
Source: *2004 Surveys of Kentucky Teachers, Principals, Superintendents, and School Board Members*

The writing portfolio was the only item that a majority of each group of respondents would reduce or eliminate. More than 40 percent each of teachers and principals wanted on-demand writing to be emphasized less in the index or removed. On-demand writing can be distinguished from writing portfolios in that there was more support for reducing the item's weight than eliminating it from the index.

On-demand writing falls within the second group of items for which sizable minorities would reduce the weight or eliminate the item. More than 40 percent each of teachers and principals wanted on-demand writing to be emphasized less in the index or removed. On-demand writing and the other items in the second grouping (arts and humanities, practical living and vocational studies) can be distinguished from writing portfolios in that there was more support for reducing the item's weight than eliminating it from the index.

Compared to the items noted so far, relatively small percentages of respondents indicated support for reducing or eliminating math, reading, CTBS/5, science, and social studies.

Effect of CATS Testing on Curriculum, Instruction, and Learning

Teachers, principals, superintendents, students, parents and guardians, and school board members were asked whether they agreed or disagreed with a series of statements about CATS testing. Some of the statements were positive, such as "CATS

testing helps align the curriculum." Some were negative, such as "Teaching what is to be covered on the CATS tests is too limiting." Table 4.12 shows the percentages of respondents that agreed with each statement. The wording of some of the statements varied by group. These differences are explained in the note below the table.

Table 4.12
Percentages of Survey Respondents Agreeing With Statements About the Effects of CATS Testing

Statement	Teachers	Principals	Super-intendents	Students	Parents/ Guardians	School Board Members
1. Getting ready for or taking the CATS tests takes too much time away from class time.	73.1	55.8	54.7	57.9	61.2	Not asked
2. CATS testing is too stressful and reduces enjoyment of teaching and learning.*	83.7	75.3	54.7	61.4	50.5	56.3
3. Teaching what is to be covered on the CATS tests is too limiting.*	72.8	52.7	42.1	75.9	86.9	71.3
4. Teachers and students are forced to cover material too quickly in order to prepare for CATS tests.	88.8	74.3	50.5	74.1	66.0	65.6
5. CATS testing provides needed focus and organization.*	57.6	70.8	78.3	40.5	41.7	71.0
6. CATS testing helps align the curriculum.	67.1	85.5	88.8	Not asked	Not asked	75.6

Note: For school board members, each statement included "Not sure" as a potential response. Those responding "Not sure" were excluded in calculating the percentages and respondent totals in the table.
*Statement 2: Students and school board members were given two statements ("CATS testing is too stressful" and "CATS testing reduces enjoyment of learning.") The table shows the results for the "too stressful" statement.
Statement 3: The student version was "Being taught only what is to be covered on the test is too limiting."
Statement 5: The version for students and parents and guardians was "CATS testing helps teachers and students focus on what's most important."
The number of respondents varied by statement. The number of teachers ranged from 314 to 326, the number of principals from 224 to 228, the number of superintendents from 106 to 107, the number of students from 158 to 159, the number of parents and guardians from 96 to 99, and the number of school board members from 125 to 131.
Source: *2004 Surveys of Kentucky Teachers, Principals, Superintendents, Students, Parents and Guardians, and School Board Members*

Generally, more than one-half—sometimes much more—of teachers, principals, and superintendents agreed that
• getting ready for or taking CATS tests took too much time,
• CATS testing was too stressful and reduced enjoyment of teaching and learning,
• teaching what was to be covered on the tests was too limiting, and
• teachers and students were forced to cover material too quickly.

The first four statements were worded so that agreeing with them would reflect negatively on CATS testing. With one exception (superintendents, statement 3), more than half of each group of respondents agreed that

- getting ready for or taking the CATS tests took too much time away from class time,
- CATS testing was too stressful and reduced enjoyment of teaching and learning,
- teaching what was to be covered on the CATS tests was too limiting, and
- teachers and students were forced to cover material too quickly in order to prepare for CATS tests.

The level of agreement was highest for the last two statements above. At least 71 percent each of teachers, students, parents and guardians, and school board members agreed that teaching what was to be covered on the CATS tests was too limiting. At least 65 percent of each of those groups and principals responded that teachers and students were forced to cover material too quickly.

Educators also agreed with positive statements about CATS testing. At least 57 percent of teachers, principals, superintendents, and school board members agreed that testing provided needed focus and organization. At least 75 percent of each group agreed that testing helped align the curriculum.

The last two statements in the table were worded so that agreeing with them reflects positively on CATS testing. For the most part, respondents did agree with them. At least 57 percent of teachers, principals, superintendents, and school board members agreed that CATS testing provided needed focus and organization. At least 75 percent of each group agreed that CATS testing helped align the curriculum.

There were three questions to educators addressing the effect of CATS testing on curriculum, instruction, and learning. The questions were similar in intent and educators' answers followed a pattern, so the results for all three questions are included in the following table.

More than three-fourths of principals and superintendents replied that CATS testing had a positive impact on the curriculum their staffs taught and on instruction in the classroom. At least 63 percent of each group rated the effect of CATS testing on what students learned as positive. For each question, one-half or fewer of teachers agreed that the effects of CATS testing were positive.

More than three-fourths of principals and superintendents replied that CATS testing had a positive impact on the curriculum their staffs taught and on instruction in the classroom. At least 63 percent of each group rated as positive the effect of CATS testing on what students learned. For each question, one-half or fewer of teachers agreed that the effects of CATS testing were positive. From 29.6 to 37.3 percent responded that the effect was negative.

Table 4.13
Educators' Ratings of the Effects of CATS Testing on Curriculum, Instruction, and Learning

	Teachers	Principals	Super-intendents
Overall, how does CATS testing affect the curriculum you teach/ your staff teaches?			
Somewhat or very negatively	37.3	17.1	12.1
Neutral or no effect	17.4	5.3	7.5
Somewhat or very positively	45.3	77.6	80.4
	100.0 %	100.0 %	100.0 %
Overall, how does CATS testing affect instruction in your classroom/school/school system?			
Somewhat or very negatively	34.4	15.9	15.0
Neutral or no effect	20.6	6.6	3.7
Somewhat or very positively	45.1	77.4	81.3
	100.0 %	100.0 %	100.0 %
Overall, how does CATS testing affect what students learn?			
Somewhat or very negatively	29.6	14.9	12.1
Neutral or no effect	20.4	8.8	24.3
Somewhat or very positively	50.0	76.3	63.6
	100.0 %	100.0 %	100.0 %

Note: Percentages as shown may not add to 100 percent due to rounding. For each question, the "somewhat" and "very" answers were combined in the table. The number of respondents varied by question. The number of teachers ranged from 326 to 328 and the number of principals from 226 to 228. The number of superintendents was 107 for each question.
Source: *2004 Surveys of Kentucky Teachers, Principals, and Superintendents*

It should also be noted that for these questions, there were more respondents answering "very positively" than "very negatively." The percentages of each group answering very negatively were approximately 5 percent or less. Approximately one-fourth or more of principals and superintendents answered that CATS testing affects curriculum and instruction very positively. The percentage of teachers answering very positively ranged from 7 to 14 percent.

Students were also asked how CATS testing affects what students learned. More than half replied that the effect was positive or somewhat positive.

Students were also asked how CATS testing affects what students learned. The possible responses were the same as for the questions above. More than half (53 percent) replied that the effect was positive or somewhat positive. Approximately 13 percent responded that the impact was negative. The remaining third of students indicated that CATS testing had a neutral effect or no effect.

Finally, an obvious impact of CATS testing on classroom activity is the time taken to prepare for the tests. Such activity may be worthwhile in itself, but it does reduce the time for other work. To get some idea of their perceptions, teachers and students were asked to estimate the amount of time spent preparing for CATS tests.[14] The question noted that respondents should not include time spent on the material to be covered by the test and on writing portfolios. The question was open ended; respondents could give any answer from 0 to 100 percent. Staff coded the responses into 10 categories, ranging from "0 to 10%" to "More than 90%." The table below shows the percentages of teachers and students whose answers were in each category. For each group, there is also a column indicating the cumulative percentage.

Table 4.14
Teachers' and Students' Estimates of the Percentage of Work Time Spent Preparing for CATS Testing

Category	% of Teachers	Cumulative %	% of Students	Cumulative %
0 to 10%	34.3	34.3	31.4	31.2
11 to 20%	17.5	51.8	15.4	46.8
21 to 30%	17.5	69.3	11.5	58.3
31 to 40%	6.8	76.1	9.6	67.9
41 to 50%	8.2	84.3	10.3	78.2
51 to 60%	1.8	86.1	3.8	82.1
61 to 70%	3.6	89.6	4.5	86.5
71 to 80%	4.6	94.3	7.1	93.6
81 to 90%	2.1	96.4	3.8	97.4
More than 90%	3.6	100.0	2.6	100.0
Total	100.0 %		100.0 %	
Teachers: average=29%, median=20%				
Students: average=32.5%, median=25%				

Note: 280 teachers and 156 students answered the questions.
Source: *2004 Surveys of Kentucky Teachers and Students*

[14] The question for teachers was "What percentage of your work time during the entire school year do you spend preparing for CATS testing, including test-taking techniques and practice tests? Do not include time you spend teaching core content or time spent on writing portfolios." The question for students was "About what percentage of class time throughout the school year do you spend preparing for CATS testing, including test-taking techniques and practice tests? Please do not include time spent learning the subject matter covered by the tests or time spent on writing portfolios."

The average teacher's estimate was that 29 percent of work time during the school year was spent preparing for CATS testing. The average student estimate was 32.5 percent. Approximately a third of each group responded that preparation took 10 percent or less of work time. Approximately one-half of teachers and students answered that preparation took 20 percent or less of work time. Fourteen percent of teachers and 18 percent of students answered that preparation for testing took up more than half of work time.

The average teacher's estimate was that 29 percent of work time during the school year was spent on preparing for CATS testing. The median teacher response was 20 percent.[15] The average student estimate was 32.5 percent. The median student response was 25 percent.

As shown in the table, teachers' and students' estimates were consistent. Approximately a third of each group responded that preparation took 10 percent or less of work time. Approximately one half of teachers and students answered that preparation took 20 percent or less of work time. On the other hand, 14 percent of teachers and 18 percent of students answered that preparation for testing took up more than half of work time.

Student CATS Reports and School Report Cards

Parents and guardians were asked a series of questions about whether they received CATS subject test reports for their children, the understandability and usefulness of the reports, and whether anyone from the child's school discussed the report with the parent. For comparison, most of the questions were also asked about school report cards. The results are shown in Table 4.15.

[15] If each teacher's answer was put in order from lowest to highest (or highest to lowest), the median is the middle answer.

Table 4.15
Survey of Parents and Guardians: Distribution and Usefulness of CATS Student Reports and School Report Cards

Student Report % "Yes" (number of respondents)			School Report Card % "Yes" (number of respondents)		
1. Do you usually receive reports about your child's performance on the CATS subject tests?	▶	83.2% (95)	92.2% (102)	◀	5. Do you usually receive the School Report Card for your child's school?
2. Is the information in the CATS tests report presented in a way that makes it easy to understand?	▶	77.6% (76)	85.6% (90)	◀	6. Is the information presented in a way that makes it easy to understand?
3. Does the CATS tests report help you understand how your child is doing in school?	▶	48.0% (75)	61.4% (88)	◀	7. Does the School Report Card help you understand how your child's school is doing?
4. Does someone from your child's school usually discuss the CATS tests report with you?	▶	14.3% (77)	Not asked		

Note: Percentages and number of respondents are calculated excluding those who answered "Not sure."
Source: *2004 Surveys of Kentucky Parents and Guardians*

Almost all the parents and guardians who responded indicated that they received the CATS student report (83 percent) and the school report card (92 percent). More than three-fourths also answered that the information in each type of report was presented in a way that was easy to understand. More than 60 percent answered that the school report card helped them to understand how their children's schools were doing. Just fewer than one-half indicated that the CATS tests reports helped them understand how their children were doing in school. Only 14 percent of the parents and guardians who responded to the survey indicated that someone from their child's school usually discussed the CATS tests reports with them.

Almost all the parents and guardians who responded indicated that they received the student report (83 percent) and the school report card (92 percent). More than three-fourths also answered that the information in each type of report was presented in a way that was easy to understand.[16]

Parents and guardians were less inclined to agree that the information was useful. More than 60 percent answered that the school report card helped them to understand how their children's schools were doing. Just fewer than one-half indicated that the CATS tests reports helped them understand how their children were doing in school. The response to another survey question provides a potential explanation for the results not being useful to most parents and guardians. When asked whether someone from the child's school usually discussed the CATS tests reports with them, only 14 percent of the parents and guardians who responded to the survey said "Yes".

[16] Students were also asked whether their parents received CATS subject test reports and whether the parents found the information easy to understand. Excluding those answering "Not sure," the percentages were virtually the same as the parents' answers: 84.4 percent (received report) and 72 percent (easy to understand). Students were also asked about the usefulness of the reports. Approximately two-thirds answered that their parents found the results useful; 59 percent of students found the results useful for themselves.

Chapter 5

Cost of CATS

Senate Joint Resolution 156 directed that the Office of Education Accountability's study include an examination of the actual costs associated with CATS, on a per student basis. This chapter presents information pertaining to the costs incurred by state government agencies and local school systems for developing, administering, scoring, and disseminating CATS accountability tests and results.

The major tasks involved in enumerating the actual costs associated with CATS are to identify the particular activities that comprise the assessment system, to define the relevant categories of costs that should be charged against the assessment system, and to gather and analyze data regarding the magnitude of the identified costs at both the state and local levels.

State agencies spent approximately $20.8 million, or $43.49 per assessed pupil, in FY 2004 on CATS-related costs.

Based on the available information, the major conclusions are as follows. First, information about the expenditures for CATS incurred by state agencies can be obtained directly by adding the expenditures of the agencies charged with various activities related to the accountability system. State agencies spent approximately $20.8 million in FY 2004 on the functions related to CATS. Given that slightly more than 477,000 pupils participated in CATS that year, the per assessed pupil state cost is estimated to be $43.49.

Local school districts spent an estimated $16.4 million, or $34.28 per assessed pupil, in FY 2004 on CATS-related expenditures.

Second, it was determined that enumerating the costs incurred by local school systems for CATS is not as direct or reliable because of significant variations in the particular expenditures that local school officials attribute to CATS. A cost survey was administered to school district financial officers and similar individuals at individual schools. Using the survey results, it is estimated that local school systems spent approximately $34.28 per student on CATS-related expenditures in FY 2004. This translates into a total statewide estimated local expenditure of $16.4 million. Adding this to the estimated state-level expenditure yields total estimated state and local expenditures related to CATS of $37.2 million in FY 2004, or about $77.77 per assessed pupil.

To provide context, the estimated CATS-related expenditure of $77.77 per assessed pupil is less than 1 percent of the $8,029.84 in total per student spending by local districts.

For comparison purposes, the total estimated per assessed pupil CATS-related expenditure of $77.77 per student is less than 1 percent (0.97 percent) of the $8,029.84 spent per student in Kentucky by local school districts during the 2004 school year

(Commonwealth. Kentucky Dept. of Ed. *State of Kentucky Profile 2004*). While the $8,029.84 does not contain state-level assessment expenditures, the comparison does provide some context to the magnitude of estimated total CATS-related spending.

In addition to actual expenditures of dollars, another resource that could be considered in estimating costs is the time school officials devote to CATS-related activities. As discussed in Chapter 4, teachers, principals, and superintendents were sent surveys regarding their experiences with CATS.

As an indicator of the amount of time devoted to CATS, teachers were asked to report what percentage of their time during the year is spent preparing students for CATS, not including teaching the core content and preparing writing portfolios. On average, the 281 teachers responding to the question indicated that they spent about 29 percent of their time preparing students for CATS: about half of the teachers reported spending 20 percent of their time or less in CATS preparation, while about half reported spending more than 20 percent of their time on such tasks. In a related measure, approximately 56 percent of the 108 superintendents who responded to a survey offered the opinion that CATS takes too much instructional time away from teaching, and approximately 56 percent of the 229 principals who responded indicated a similar opinion.

It is important to not only consider the costs of an accountability testing system but also the **value** of the information it provides.

While this chapter addresses the issue of the costs of CATS, it is important to remember that cost represents only one side of the equation. The fundamental question is about the **value** of what is purchased compared to its costs. For example, consider the purchase of a television. A cost of $1,000 would be considered exorbitant for a 13-inch black and white set but dirt cheap for a thin-screen 70-inch high-definition set.

For assessment systems, the essential value to be considered is the validity of the test in accurately measuring schools' progress toward state education performance goals. An assessment system that suffers from validity problems is a poor value no matter what its cost. Thus, while this chapter discusses costs in isolation, it is critical to keep in mind that the value of the assessment system can only be evaluated in light of its validity, which is covered in other parts of this report.

Organization of the Chapter

The remainder of this chapter presents a brief description of the required components of the Commonwealth Accountability Testing System, including specific statutory citations. The chapter also provides a discussion of how CATS-related costs were defined for the purposes of this analysis. The estimate of CATS-related costs incurred at the state level is presented next, followed by the estimate of CATS-related local-level costs. The local cost estimates are presented in total and are also broken into costs reported by individual schools and costs reported by central offices. The last section provides a summary of the information obtained about total state and total local CATS-related costs and presents the final statewide estimates.

Required Components of CATS

CATS was created by statute (KRS 158.6453) to be a statewide assessment program.

CATS is established in KRS 158.6453. Under that statute, the Kentucky Board of Education is directed to create and implement CATS as a statewide assessment program to ensure that schools can be held accountable for students' achievement of academic performance goals set forth in KRS 158.645 and 158.6451. The goal most relevant to CATS is that schools should develop the ability of students to understand, integrate, apply, and communicate basic principles from science, mathematics, arts, humanities, social sciences, and practical living.

CATS has five specific statutorily required components.

The statute requires that CATS include five specific components. Table 5.1 lists each of the required components, identifies the specific assessment instrument the Board of Education has adopted to meet the requirement, and notes the grades and subject matter covered by each instrument.[1]

[1] The specific instruments incorporated into CATS and their uses are described fully in other chapters of this report. That level of detail is not repeated here.

Table 5.1
Required Components of CATS

Required Component and Statute	CATS Component	Subjects Assessed	Grades Assessed
Customized or commercially available norm-referenced test KRS 158.6453 (2)(a)	CTBS/5 (Terra Nova Test): a multiple-choice norm-referenced test that allows comparisons nationally between Kentucky students and their peers	Reading, Language Arts, and Mathematics	3,6,9
Open-response or multiple-choice items, or both KRS 158.6453 (2)(b)	KCCT: questions test how well a student has mastered the core content of various subjects	Reading	4,7,10
		Math	5,8,11
		Science	4,7,11
		Social Studies	5,8,11
		Practical Living and Vocational Studies	5,8,10
		Arts and Humanities	5,8,11
On-demand assessment of student writing KRS 158.6453 (2)(b)	On-Demand: writing tests that measure skills of extemporaneous writing	Writing	4,7,12
Writing portfolios consisting of samples of student work KRS 158.6453 (2)(c)	Writing Portfolio: a collection of a student's edited writing	Writing	4,7,12
Performance assessment events for schools that have students enrolled in performing arts organizations KRS 158.6453 (2)(d)	No performance assessment in place	Not Applicable	Not Applicable
A technically sound longitudinal comparison of the assessment results for the same students KRS 158.6453 (2)(e)	No longitudinal assessment in place	Not Applicable	Not Applicable

Source: Compiled by staff

Definition of CATS-related Costs

CATS-related expenditures are defined as those that would not have occurred had CATS not existed.

In theory, the only expenditures that should be attributed to CATS are those that would not have been incurred if the accountability system did not exist. Instructions for the cost survey of local school officials emphasized that respondents should only report CATS-related costs judged to be over and above regular instructional costs.

It is not clear that all survey respondents classified CATS-related expenditures in the same way.

However, it is not clear that all school officials who responded to the cost survey utilized similar criteria for determining what expenditures they only made because of the existence of CATS. For example, when asked to provide expenditures related to CATS, one local school included $14,000 (out of a total reported CATS-related expenditure of $16,000) for the purchase of *SuccessMaker,* a software program for individualized student instruction and

assessment. The implication is that, absent CATS, the school would not have chosen to use this particular approach with its students. Conversely, another school reported a total CATS-related expenditure of $31 for miscellaneous testing supplies. It is not clear whether the wide variation in these reported costs reflects differences in the schools' definition of what costs are related to CATS or differences in the particular methods they use to prepare their students because of the accountability system, which they imply would not be used otherwise.

Given the major caveat regarding the variations in the apparent manner that local officials define the particular expenditures related to CATS, this report presents the information in two ways. First, a simple estimate of local per student CATS-related expenditures is made taking the reported cost figures from school officials at face value. This estimate would be accurate under the assumption that the variation in expenditure categories reflects real differences in how schools conduct their operations because of CATS.

In order to provide the most information, an estimate of local costs is provided as well as a detailed listing of the components of local costs.

However, if the variations in the reports of local CATS-related expenditures are a reflection of differences in how costs are categorized, then they might not be an accurate reflection of true CATS-related costs. To allow readers to judge whether particular types of reported expenditures should properly be charged to CATS, details are presented regarding reported school expenditures for various categories of costs. This will allow the reader to independently evaluate whether some particular categories of local expenditures should not have been charged to CATS. Note that a limitation of this approach is that there is no way to know the level of underreporting of expenditures that the reader might believe **should** be charged to CATS, but were not reported as such by every local school official.

This study does not consider "opportunity" costs in the estimates.

An additional note is that this report does not consider indirect, "opportunity" costs. The concept of an opportunity cost refers to the fact that by using resources for one purpose, they are no longer available to be used for a second purpose. The opportunity cost is the benefit given up by not directing the resources to the second purpose. For example, if a business uses resources to purchase new equipment, then those resources are no longer available to purchase additional advertising. The opportunity cost of buying the equipment is measured by the extra profit that would have been generated by the additional advertising.

In the case of CATS, the opportunity costs of resources devoted to CATS is best understood as the gain in validity in the assessment

system that could be achieved under an alternate system, assuming that in the absence of CATS another assessment system would exist. The point is that in order to assess the opportunity costs of CATS, it is necessary to have a measure of the validity of the assessment system that would be under consideration for replacing it.

Prior Cost Studies

There have been prior estimates of CATS-related costs.

Two prior studies of the direct costs of CATS have been published. Hoxby (2002) considered only state-level expenditures in FY 2001 in estimating a cost of $18 per student in school. Using Hoxby's total estimated FY 2001 expenditure of $11.6 million and dividing by the number of students tested in that year yields an estimated state-level cost of $25 per student tested. Another report, completed by the LRC Program Review and Investigations Committee in 2003, estimated total state-level costs of $10.3 million in FY 2003 and local district costs of $10.6 million in FY 2002. Taken together, these yield an estimated cost of about $44 per student tested (Commonwealth. Legislative.).

Cost Definitions for Current Study

To estimate the direct costs associated with CATS, staff examined CATS expenditures at both the state and local levels. State-level costs were obtained by examining financial information provided by the Kentucky Department of Education (KDE), extracting figures from the Management Reporting Database (MRDB), and reviewing state contracts and invoices for CATS services. Local costs are estimated using data obtained from a survey sent to all school districts and schools in the Commonwealth.

CATS-related expenditures are grouped into three categories:
- advisory and research;
- implementation and administration; and
- accountability.

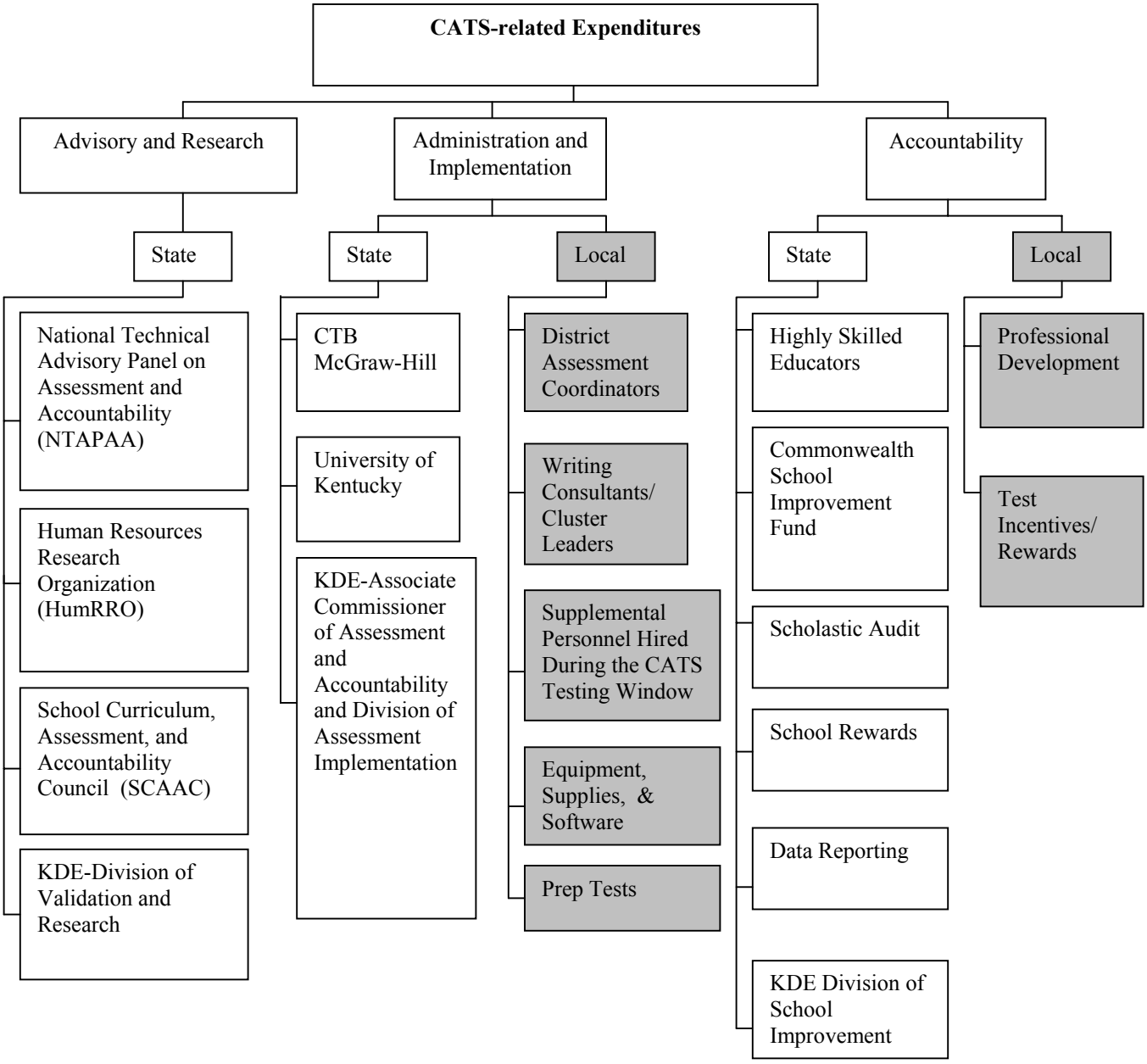
For this analysis, CATS-related expenditures are assigned to one of three categories: advisory and research, implementation and administration, and accountability. Figure 5.A illustrates the specific expenditures included within each category. The relevant state and local CATS-related expenditure categories are identified.

The advisory and research category includes expenditures made by the state for expert advice provided to the Kentucky Board of Education, KDE, and LRC regarding assessment and accountability issues. The implementation and administration category includes costs to develop and produce the assessment instruments and administer them to students. The accountability category includes expenditures for financial and educational support to schools and districts that are directly tied to CATS scores. Many of the rewards and sanctions listed are under

programs instituted before CATS but which are now integral to the current assessment system by either rewarding schools that are achieving desired accountability scores or sanctioning and assisting schools that are not.

A legitimate question is whether expenditures in the accountability category should properly be charged to CATS. From the standpoint that the rewards and remediation expenditures would not exist if there were no accountability system, including those expenditures is appropriate. However, it could be argued that the most pertinent question is what it costs to obtain the accountability information that allows the determination of which schools should be rewarded and which sanctioned. The decision in this study was to provide information regarding expenditures in the accountability category to allow the interested reader to deduct them from the totals if that is believed to be a more accurate representation of the costs of interest. Note that expenditures in the accountability category were not included in the cost estimates in the 2003 Program Review and Investigations study cited above. That is the major reason for the large difference in the estimates between this report and that one.

Figure 5.A
Categories of CATS-related Expenditures



Source: Staff analysis.

State-level Expenditures for CATS in FY 2004

State-level CATS-related spending was estimated to be about $20.8 million, or about $43.49 per assessed pupil.

Examination of state FY 2004 financial records for the expenditure categories identified in Figure 5.A indicates that state agencies expended a total of $19.1 million in state general funds and $1.6 million in federal funds for CATS-related expenditures in that year (Table 5.2). Nearly all of the total expenditures were divided between the implementation/administration and accountability categories, with advisory and research activities accounting for only 3 percent of the total.

The state expenditure from the General Fund for the approximately 477,000 students who took the CATS test was $40.11 per student. When both state and federal funds were combined, the expenditure was $43.49.

Table 5.2
State-level CATS-related Expenditures
FY 2004

Expenditure Category	Federal Funds	State General Fund	Total Funds	Percent of Total*
Advisory and Research		$696,638	$696,638	3%
Implementation and Administration	$1,601,652	$9,252,524	$10,854,176	52%
Accountability	$13,907	$9,195,571	$9,209,478	44%
Total	**$1,615,559**	**$19,144,733**	**$20,760,292**	**100%**

*Does not total 100 percent due to rounding.
Source: Compiled by staff using financial data obtained from KDE, LRC staff, and MRDB.

The remainder of this section presents detailed descriptions of the specific expenditures included in each of the three major categories and reflected in the totals above.

Advisory and Research Expenditures

Advisory and research expenditures related to CATS totaled slightly more than $696,000.

Several advisory and research entities perform research functions for Kentucky Board of Education and LRC (Table 5.3). These entities offer advice and conduct research regarding CATS. Together, state expenditures for the CATS-related activities of these entities totaled slightly more than $696,000 in FY 2004. This represents about 3 percent of the state-level CATS expenditures.

The National Technical Advisory Panel on Assessment and Accountability (NTAPAA) is a panel of national testing experts appointed by LRC, in accordance with KRS 158.6454, to advise the Kentucky Board of Education, KDE, and LRC on issues related

to CATS. The total state expenditure for NTAPAA in FY 2004 was $107,300.[2]

KDE contracts with the Human Resources Research Organization (HumRRO), a Louisville consulting firm, to perform on-going technical studies relating to CATS. These studies relate to the validity and reliability of the assessment system. According to KDE, state payments to HumRRO in FY 2004 were $239,000.

The School Curriculum, Assessment and Accountability Council is charged with advising the Kentucky Board of Education and LRC on issues related to the creation and implementation of CATS. State expenditures for the council were $4,195 in FY 2004 according to KDE.

KDE funds the Division of Validation and Research, which is within the Office of Assessment and Accountability. State expenditures for the division were $346,143 for FY 2004, as calculated by staff using figures obtained from MRDB. Expenses included travel, payroll, and office supplies.

Table 5.3
State-level Advisory and Research Expenditures Related to CATS
FY 2004

Expenditure	State General Fund
National Technical Advisory Panel on Assessment and Accountability KRS 158.6454	$107,300
Human Resources Research Organization	$239,000
School Curriculum, Assessment, and Accountability Council KRS 158.6452	$4,195
KDE Division of Validation/Research	$346,143
Total	**$696,638**

Source: Compiled by staff using financial data obtained from KDE, LRC staff, and MRDB.

[2] Expenditures for NTAPAA are made in the legislative budget.

Implementation and Administration Expenditures

Implementation and administration expenditures related to CATS were approximately $10.9 million at the state level.

Slightly more than half (52 percent) of the state-level expenditures related to CATS are for implementation and administration of the accountability tests. Approximately $10.9 million was spent by state agencies for activities in this category in FY 2004 (Table 5.4).

Table 5.4
State-level Implementation and Administration Expenditures
Related to CATS
FY 2004

Expenditure	Federal	State	Total	Percent
CTB McGraw-Hill (assessment contractor)	$1,601,652	$8,154,700	$9,756,352	90%
University of Kentucky (alternate portfolio contractor)		$249,400	$249,400	2%
KDE-Associate Commissioner of Assessment and Accountability and Division of Assessment Implementation		$848,424	$848,424	8%
Total	**$1,601,652**	**$9,252,524**	**$10,854,176**	**100%**

Source: Compiled by staff using financial data obtained from KDE, LRC, and CTB staff.

CTB McGraw-Hill. Approximately 90 percent of the state-level expenditures for implementation and administration were payments to CTB McGraw-Hill (CTB), the test contractor hired by the Kentucky Board of Education to carry out such functions related to the assessment. CTB's responsibilities include developing the assessment, providing assistance in administering the assessment instruments, scoring tests, and reporting the results. In FY 2004, KDE reported paying CTB $9.8 million. Approximately $8.2 million was paid from state funds and $1.6 million was paid from federal funds. A detailed listing of the services provided by CTB is shown in Table 5.5.

Table 5.5
Itemized Costs From the CTB McGraw-Hill CATS Contract
FY 2004

Service	Expenditure	Percent of Total Contract
CATS Test Development	$1,971,698	20.2%
CATS Test Materials	$1,940,434	19.9%
CATS Test Processing	$3,143,337	32.2%
On-demand Writing Prompts	$127,588	1.3%
Writing Portfolio	$1,354,996	13.9%
Customer Service Center	$21,753	0.2%
CATS Score Reporting	$383,636	3.9%
Other Costs	$199,327	2.0%
KCCT Subtotal	$9,142,769	
CATS Technical Specifications/Research	$338,801	3.5%
NRT TerraNova Assessment	$274,682	2.8%
Total	**$9,756,252***	**100%**

*The February 24, 2004, invoice submitted by CTB shows $2,391,731.25, but KDE paid $2,391,831.25, accounting for the $100 difference in the total of the CTB contract in Table 5.5 and Table 5.6.
Note: State and federal dollars included.
Source: Compiled by staff using financial data obtained from KDE and CTB-McGraw Hill officials.

Alternate Portfolio Program. KDE contracts with the University of Kentucky to perform work on the alternate portfolio component of CATS. The Kentucky Alternate Portfolio Program at the university provides support for districts and teachers in administering an alternate portfolio to students who cannot take the regular CATS assessments because of a significant disability. Between 1,200 and 2,000 students each year participate in this method of assessment. According to KDE, the FY 2004 contract for support of the alternate portfolio program was $249,400.

KDE Associate Commissioner. The KDE associate commissioner of assessment and accountability is responsible for overseeing the implementation of CATS. The associate commissioner supervises the work of the Division of Assessment Implementation, the Division of Validation and Research, and the work of various assessment contractors and panels.[3] Expenditures of the associate commissioner's office, including personnel expenditures, and the Division of Assessment Implementation are included in the implementation and administration category. Their combined

[3] Although the Division of Validation and Research reports to the associate commissioner of Assessment and Accountability, expenditures by the division were included in the advisory and research category rather than the implementation and administration category.

expenditures were $848,424 in FY 2004, as calculated by staff using figures extracted from MRDB.

Accountability Expenditures

Accountability expenditures totaled about $9.2 million at the state level.

In addition to the expenditures for development and implementation of components of the assessment system, KDE also makes expenditures based on the assessment results. These include expenditures for reporting results to schools and the public and any rewards given to schools that perform well or for assistance directed to schools that do not perform well. Some of the accountability programs, such as Highly Skilled Educators, Scholastic Audits and Reviews, and Commonwealth School Improvement Funds, were enacted prior to CATS. They could still be considered a cost of the assessment system because they would not exist in the same manner in its absence. However, it is also reasonable to argue that costs associated with actions arising from information obtained in an accountability system should not be directly charged to the accountability system itself. In this report, the effort is to provide the most complete information available, so the accountability expenditures are included in the total. Those who believe this overstates the true cost of CATS should deduct the accountability amounts from the totals provided.

In FY 2004, KDE expended $9.2 million for accountability activities related to CATS (Table 5.6). These expenditures were for programs to disseminate information about CATS scores and to assist schools with scores below those deemed acceptable.

Table 5.6
State-level Accountability Expenditures Related to CATS
FY 2004

Expenditure	Federal	State	Total	Percent
Highly Skilled Educators KRS 158.782	$0	$6,126,292	$6,126,292	67%
Commonwealth School Improvement Fund KRS 158.805	$0	$2,029,329	$2,029,329	22%
Scholastic Audit and Review 703 KAR 5:120	$13,907	$361,489	$375,396	4%
School Rewards	$0	$0	$0	0%
Data Reporting (report cards)	$0	$294,754	$294,754	3%
KDE Division of School Improvement	$0	$383,707	$383,707	4%
Total	**$13,907**	**$9,195,571**	**$9,209,478**	**100%**

Source: Compiled by staff using financial data obtained from KDE, LRC staff, and MRDB.

Highly Skilled Educators. Slightly more than two-thirds of the expenditures in the accountability category were made for the Highly Skilled Educators program in FY 2004. First implemented in 1998, the Highly Skilled Educator program is governed by KRS 158.782. Schools designated as Level 3, ("in need of assistance") as a result of CATS scores are required to receive assistance from a teacher who has been designated to have exceptional teaching ability (703 KAR 5:120 Section 3). Pursuant to KRS 158.782, KDE is responsible for designating and training highly skilled educators and absorbing the cost of their leaves of absence from their home schools while they serve for two years at a school designated as being in need of assistance. They provide various services including training staff, making personnel assignments, and completing evaluations of personnel. There were 47 highly skilled educators assigned to schools in FY 2004 at a cost of more than $6.1 million, as calculated by staff using figures extracted from MRDB.

Commonwealth School Improvement Funds. KRS 158.805 established the Commonwealth School Improvement Fund to provide financial assistance to local schools so they can pursue innovative strategies to meet the educational needs of their students and, thereby, attempt to raise the performance level. State spending for the fund was $2.0 million in FY 2004, accounting for slightly less than a fourth of state-level accountability expenditures according to staff calculations using figures extracted from MRDB. Approximately $1.8 million went to school districts, and the remaining $200,000 went to KDE personnel and operating expenses. The funds distributed to school districts through Commonwealth School Improvement Fund were included in the state section of this report rather than the local section. The total appropriation could be obtained at the state level, whereas the amount spent at the local level was less than the total since not all districts responsed to the survey.

Scholastic Audits and Reviews. Under KRS 158.6455 and 703 KAR 5:120 and 130, schools that fail to meet specified CATS goals are subject to a scholastic audit or review. In this case, a team of educators reviews the personnel and academic operations of the school and makes recommendations regarding how these can be improved to enhance students' academic performances. In FY 2004, $375,396 was expended for this program, according to KDE.

KDE has an alternating schedule for scholastic audits and reviews. For the years KDE performs scholastic audits, outside help is hired under contract. For scholastic reviews, KDE staff and highly skilled educators are used. Generally, scholastic audits are more

costly. FY 2004, the year covered by this study, KDE conducted scholastic reviews. The portion of time the highly skilled educators and KDE staff spent conducting reviews is included in this section.

School Rewards. KRS 158.6455 requires that the Kentucky Board of Education establish a system to reward schools that perform well on the CATS assessment. Administrative regulations 703 KAR 5:020, 703 KAR 5:040 and 703 KAR 5:060 indicate that schools that meet or exceed their specified CATS goals are to be rewarded with funds paid from the Kentucky Successful Schools Trust Fund.

While the statutory language has not been repealed, beginning in 2003, the General Assembly has chosen to direct amounts contained in the trust fund to uses other than school rewards. Therefore, no state expenditures were made in FY 2004 for rewards to schools.

Data Reporting. The cost of school and district CATS report cards has been included in this category of expenditures. In FY 2004, the state spent $294,754 in printing and distributing these report cards to the public, as calculated by staff using figures extracted from MRDB.

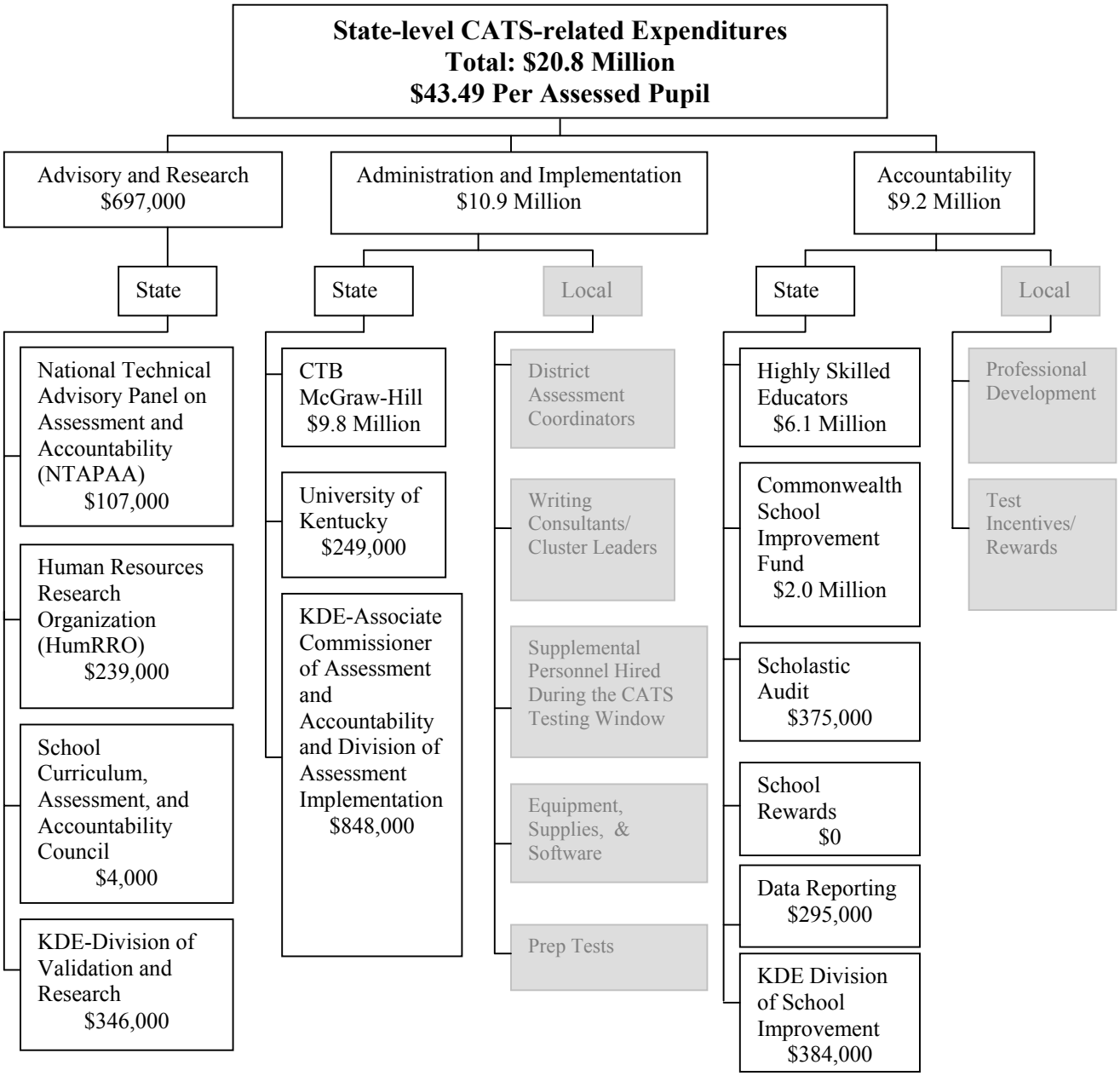
MAX Database. The 2000-2002 budget passed by the 2000 General Assembly provided funding to the Education Professional Standards Board to establish a new enterprise database. The database warehouses student, financial, and related data and provides teacher quality data in a publicly accessible format. The MAX system does provide an online link to public reports regarding CATS, Kentucky Core Content Test, and No Child Left Behind performance, but staff who support that system indicated that the electronic links take the viewer to the site supported by KDE. Although suggested by KDE, this report does not classify expenditures associated with the MAX database as being related to CATS. Therefore, the $1.6 million expenditure, obtained through MRDB and KDE, on the MAX database in FY 2004 is not included in the totals above.

KDE Division of School Improvement. The KDE Division of School Improvement oversees the Highly Skilled Educators program, the Commonwealth School Improvement Fund program, and Scholastic Audits/Reviews. Through these programs, the division assists those schools identified by the CATS assessment and the Kentucky Board of Education as low-performing schools. State expenditures, obtained from MRDB, for this division were $383,707 in FY 2004.

Summary of State-level Expenditures

Figure 5.B presents a summary of the state-level expenditures related to CATS in FY 2004. Note that the boxes denoting local-level costs are lightened to indicate that they are not considered in this particular figure: they are discussed in the next section.

Figure 5.B
Estimate of State-level CATS-related Expenditures*
FY 2004



Note: Boxes denoting local-level costs are lightened to indicate that they are not considered in this figure.
*Columns may not total due to rounding.
Source: Staff analysis.

Local-level CATS-related Expenditures

Obtaining local CATS-related expenditures is more difficult than obtaining state-level expenditures. There is no systematic accounting of costs incurred for CATS by local school districts. To collect information about the cost of CATS at the local level, LRC staff surveyed personnel at district central offices and individual schools. The results of these surveys provided a snapshot of costs attributed to CATS by local officials.

Total CATS-related local-level expenditures were estimated to be $16.4 million.

Using the survey results, total estimated expenditures at the local level attributed to CATS for 2003-2004 were approximately $16.4 million. Local CATS-related expenditures fall into the categories of implementation and administration and of accountability. The estimated local CATS-related expenditures for these categories were $11.1 million and $5.3 million, respectively.

Total local-level per student spending was estimated to be $34.28.

On a per assessed student basis, the average CATS-related expenditure by local school districts was estimated to be $34.28. This is composed of an estimated per student expenditure of $17.93 reported by schools and of $16.35 reported by central offices.

This section provides more explanation of these local cost estimates. It also presents information to allow the reader to investigate further the differences in reported costs by both schools and central offices.

Survey

Two surveys were developed and administered: one for central offices and another for individual schools.

In order to gather local-level CATS costs, two Web-based surveys were conducted during November 2004. Initially, LRC staff planned to survey only district central offices. However, after feedback from a group of district finance officers, it was determined that some costs would only be known at the school level. The central office survey collected information on local costs tracked by the district office. The school survey collected information that would be difficult for the central office to quantify. To minimize the risk of the same costs being reported at both levels, each survey displayed all questions, but the ones that did not apply to the respondent were grayed out. Two e-mails were sent to the finance officer in each of the 176 school district offices. The district finance officer was asked to complete the survey regarding district-level expenditures related to CATS. That individual was also asked to forward the e-mail regarding school-level costs to the official in each school believed most able to provide accurate information about CATS-related expenditures for

that school. Responses to these two surveys were used to calculate local-level CATS-related costs.

Survey Response

Of the 176 central offices, 132 responded to the survey; 796 schools responded.

Of the 176 central offices, 132 responded to the survey. In addition, 796 individual schools responded. Table 5.7 shows summary descriptions of the responding districts and schools.

Table 5.7
Summary Characteristics of Responding Central Offices and Schools

Characteristics	Central Office	School
Number Responding	132	796
Total Number in Kentucky	176	1,249*
Number of Assessed Students Covered by Responding Local Entities	376,071	294,438
Total Number of Assessed Students in Kentucky (2003-2004)	477,361	477,361
Average Number of Assessed Students for Sample	2,849	370
Average Number of Assessed Students for State (2003-2004)	2,712	382

*This number includes all schools in Kentucky. However, some of these schools do not have CATS testing, such as elementary schools that are K-3.
Sources: LRC Survey to schools and central offices and KDE.

Central office responses covered about 79 percent of the assessed student population. School responses covered about 62 percent of the assessed student population.

The assessed student population of responding central offices was 376,071, about 79 percent of the total 477,361 students assessed in Kentucky in 2004. The number of assessed students represented by individual school respondents was 294,438, approximately 62 percent of the total assessed student population. These response rates show that the surveys collected information covering the majority of the assessed student population in Kentucky. As such, the information collected, if accurate, should give a realistic picture of the cost of CATS at the local level.

It is not clear how much variation should be expected in the expenditures reported by school and central offices related to CATS.

The surveys showed a significant amount of variation in reported CATS-related expenditures. While it is not unexpected that different schools and districts could incur different costs related to CATS, it is not clear how much variability should be expected. Some portion of reported differences may stem from actual differences in CATS-associated costs. However, it is also possible that reported differences stem from respondents classifying spending differently. The goal of the surveys was to isolate expenditures that would not have been incurred if CATS did not

exist. However, this can be difficult for a number of reasons. For example, if a salaried district employee performs work preparing for CATS but has other responsibilities, it is not completely clear how the expense of that staff member should be categorized. Care was taken to minimize the possibility of misclassification and to have respondents focus on costs that stem exclusively from CATS; however, it is not clear to what degree this attempt was successful since there is no formal accounting procedure for expenditures related to CATS.

To provide the most information, total cost estimates are provided for both schools and central offices as well as more detailed information about individual cost categories.

In order to present the results of the survey in the most effective manner, they are presented in two ways. The first provides a summary of overall reported costs and an estimate of the total expenditures reported by central offices and schools related to CATS. The second breaks the expenditure results into smaller categories by schools and central offices. In addition, further information about the variation in answers is provided to allow the reader to judge the reported expenditures that they believe to be directly related to CATS.

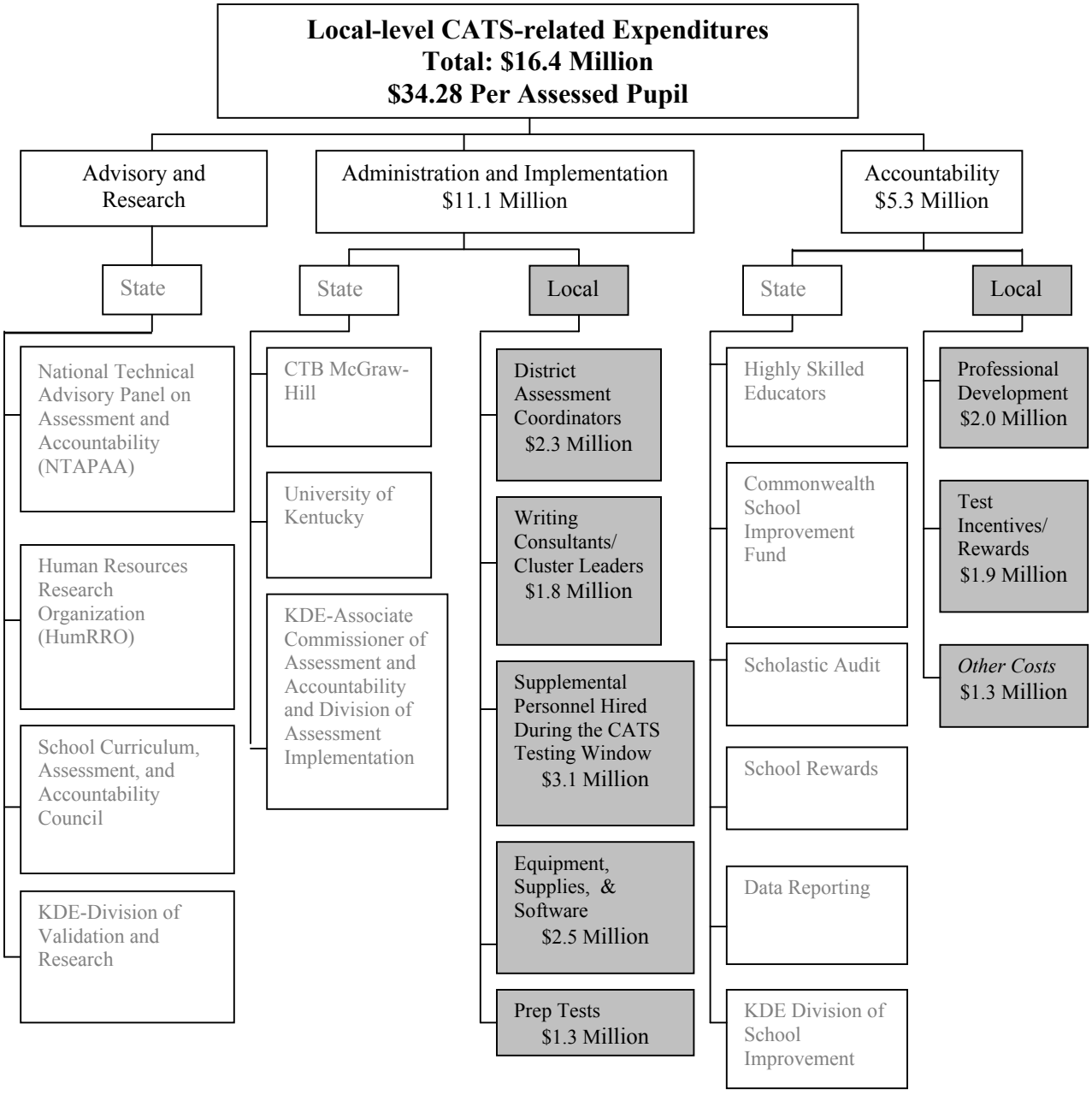
Overall Estimates of Local Expenditures Related to CATS

A total of $11.1 million was spent by local entities on CATS for administration and implementation. Another $5.3 million was spent for accountability.

Central offices reported average CATS-related expenditures of $16.35 for each assessed student in the 2003-2004 school year. Schools reported an average expenditure of $17.93 per assessed student. In all, local CATS-related expenditures per assessed student were estimated to be $34.28. Using this estimate of per student spending with the total number of students assessed during the 2003-2004 school year (477,361), total CATS-related expenditures by local units was estimated to be almost $16.4 million.

This $16.4 million in CATS-related expenditures is broken down into approximately $11.1 million for administration and implementation and $5.3 million for accountability. Figure 5.C shows further detail in reported expenditures.

Figure 5.C
Estimate of Local-level CATS-related Expenditures*
FY 2004



Note: Text in boxes denoting state-level costs is lightened to indicate that they are not considered in this figure.
*Columns may not total due to rounding.
Source: Staff analysis.

Almost 68 percent of local CATS-related expenditures were for administration and implementation. These expenditures are split among five categories: district assessment coordinators; writing consultants/cluster leaders; supplemental personnel; equipment, supplies, and software; and prep tests. Short descriptions of these categories are provided in Table 5.8 along with total local-level expenditure estimates.

Table 5.8
Local-level Administration and Implementation Expenditures
FY 2004

Category	Description	Estimated Local-level Expenditures
District Assessment Coordinators	Personnel who coordinate and facilitate the implementation of assessment and accountability programs	$2,278,000
Writing Consultants/Cluster Leaders	Personnel who serve as mentors and resources for teachers concerning writing portfolios	$1,782,000
Supplemental Personnel	Personnel hired during CATS testing window to cover workload, such as substitute teachers	$3,146,000
Equipment, Supplies, and Software	Equipment, supplies, and software for CATS (examples: pens, paper)	$2,541,000
Prep Tests	Preparatory or supplemental tests for students	$1,317,000
Total		**$11,065,000**

Source: Staff analysis

The remaining 32 percent of estimated CATS-related expenditures fall under accountability and are split among three categories: professional development; test incentives and rewards; and other expenses. This represents an estimated $5.3 million in expenditures. Table 5.9 provides short descriptions of these categories as well as total local-level estimated expenditures.

Table 5.9
Local-level Accountability Expenditures
FY 2004

Category	Description	Estimated Local-level Expenditures
Professional Development	Training for teachers and staff, including training to score portfolios and alternate portfolios	$2,036,000
Test Incentives and Rewards	Pre- and post-CATS incentives for students	$1,936,000
Other Costs	Miscellaneous expenses reported by local officials not classified elsewhere	$1,3267,000
Total		**$5,299,000**

Source: Staff analysis

Per Student CATS-related Expenditures

There was significant variation between schools and between central offices in reported CATS-related expenditures.

There was significant variation in categories of CATS-related expenditures reported by schools and central offices. Table 5.10 provides a listing of the estimated per assessed student expenditures and how they sum to the total estimated per student expenditure for both schools and central offices. The largest single expenditure per assessed student reported by central offices was for district assessment coordinators—individuals who help implement assessment and accountability programs. These positions accounted for approximately 29 percent of central office CATS-related expenditures. The second largest category of expenditures was writing consultants/cluster leaders, accounting for approximately 23 percent of reported expenditures. Together, these two categories represent more than 50 percent of CATS-related expenditures reported by central offices.

On average, school-reported CATS-related expenditures were more evenly divided into five of the eight categories. The largest expenditure component was for test incentives and rewards, accounting for 23 percent of reported spending. Supplemental personnel and equipment, supplies, and software each made up another 20 percent of reported CATS-related expenditures.

Table 5.10
Reported Local CATS-related Expenditures Per Assessed Pupil
FY 2004

Category	Total		Schools		Central Offices	
	$	*%*	*$*	*%*	*$*	*%*
District Assessment Coordinators	4.77	14%	0.00	0%	4.77	29%
Writing Consultants/ Cluster Leaders	3.73	19%	0.00	0%	3.73	23%
Supplemental Personnel Hired During the CATS Testing Window	6.59	2%	3.56	20%	3.03	19%
Equipment, Supplies, & Software	5.32	16%	3.60	20%	1.73	11%
Prep Tests	2.76	8%	2.76	15%	0.00	0%
Professional Development	4.27	12%	3.22	18%	1.04	6%
Test Incentives/ Rewards	4.06	12%	4.06	23%	0.00	0%
Other Costs	2.78	8%	0.74	4%	2.04	12%
Total	**$34.28**	**100%**	**$17.93**	**100%**	**$16.35**	**100%**

Source: Staff analysis

Variation in CATS-related Expenditures Reported by Schools and Central Offices

Reported spending variations can represent actual differences in spending or can indicate differences in how spending is reported.

It is not clear if there should be significant differences in reported expenditures between schools and differences between central offices. If reported expenditures are substantially different between schools or between districts for the same expenditure categories, this could mean a number of things. First, it could be that districts are actually spending different amounts as a result of CATS. Another explanation could be that districts are reporting expenditures differently because of differences in determining whether certain expenditures are related to CATS or not. A third possibility is that districts are accounting for and categorizing CATS-related expenditures differently, leading to similar expenditures being reported in different categories by different districts.[4] This could be the case for both responding schools and

[4] It should be noted that there is an additional explanation that reported numbers are not wholly accurate. However, even though there is no official accounting structure to isolate costs associated with CATS, there is no reason to suspect the costs reported are invalid.

central offices. It is important to note that even when the size of schools and districts is taken into account, significant variation in reported costs persists. There is no way to verify the precise cause of differences.

In order to provide readers the most information and to allow them to judge expenditures for themselves, a more detailed breakdown of reported expenditures is provided. School responses are presented first followed by central office responses. The previous expenditure estimates presented were aggregated estimated expenditures from the surveys and included all reported expenditures. This section provides a further breakdown of these aggregated expenditure estimates and allows a more detailed examination of differences in reporting.

CATS-related Expenditures Reported by Schools

Variation in spending is evident in every category of reported school-level spending.

A total of 796 schools responded to the school survey. Among those respondents, there was significant variation in reported CATS-related expenditures. This variation is seen in every category of expenditure. Table 5.11 presents categories of expenditures contained in the survey along with the average per student spending over the entire assessed student sample population. In addition, the minimum and maximum per student expenditures are reported as well as the percentage of schools that reported no expense for a given expenditure category.

In every category of expenditure, there is a wide range between the reported highest and lowest per student CATS expenditures. For example, in the expenditure category food for students, the average reported per student spending across schools is about $1.18. However, school responses ranged from nothing being spent to $29 being spent per assessed pupil. Table 5.11 also shows that, for this category, almost 31 percent of responding schools reported no expense.

Ten categories of costs had more than 50 percent of schools reporting no expenditures.

In addition to there being a significant range of responses, there are 10 expenditure categories in which more than half of responding schools claimed no expense at all for CATS. Notable are the categories other school personnel and preparatory materials. In each, almost 78 percent of schools reported no expenditure associated with CATS, while the average for the sample for these two expenditure categories was $2.15 and $1.07, respectively. Together, they comprise about $3.22 (about 18 percent) of the estimated average school per student expenditure of $17.93. Thus, while some schools report that these are major expenditure categories, others report that they are not at all important.

Table 5.11
Reported CATS-related Expenditures Per Assessed Student Reported by Schools
FY 2004

Expenditure Category	Detailed Expenditure Category	Average Per Student Expenditure	Minimum Reported Per Student Expenditure	Maximum Reported Per Student Expenditure	Percent Reporting No Expenditure
Supplemental Personnel	Substitute Teachers	1.41	0.00	48.58	34.9
	Other School Personnel	2.15	0.00	314.18	77.9
Professional Development	Scoring Portfolios Training	1.30	0.00	22.87	38.3
	Alternate Portfolio Training	0.23	0.00	6.29	69.2
	Other CATS-related PD	1.70	0.00	86.84	69.3
Equipment, Supplies, and Software	Calculators	1.55	0.00	81.08	54.5
	Miscellaneous Supplies	0.56	0.00	25.86	29.8
	Preparatory Materials	1.07	0.00	136.67	78.0
	Other Equipment or Software	0.42	0.00	47.11	89.9
Prep Tests	CATS Preparatory Tests	1.81	0.00	78.14	53.5
	CATS Supplemental Tests	0.95	0.00	50.91	82.7
Test Incentives and Rewards	Food for Students	1.18	0.00	29.03	30.8
	Pre-test Incentives	0.89	0.00	25.58	53.3
	Post-test Rewards	1.99	0.00	43.74	27.9
Other Costs	Other CATS Expenses	0.74	0.00	121.21	82.9
Total		**$17.93**	**$0.00***	**$453.62****	

*Four schools reported no CATS-related expenditures in the school survey.
**This is the highest total expenditure per assessed student reported by schools.
Source: Staff analysis

The school reporting an average of $314.18 spent on other school personnel reported total CATS-related expenditures of about $110,000, or $421 per assessed student. Of this total, $82,000 was spent on other school personnel. More specifically, the school reported that these other school personnel CATS-related expenditures were for "academic performance specialists, assistants in fifth grade, and fourth grade writing resource teachers."

The school that reported CATS-related expenditures of $136.67 per student for preparatory materials reported total expenditures per student of $154.45, spending a total of $16,063. The spending on preparatory materials was attributed to purchasing the software *SuccessMaker*, a software program used for individualized student instruction, for $14,214.

The smallest 20 percent of schools, in student population, spent almost three and a half times as much per assessed pupil as did the largest 20 percent of schools on CATS-related expenditures, on average.

Important to note is that the average spending reported by schools is higher than the average of those costs over the entire assessed student population. This implies that larger schools have generally lower per student costs than do smaller schools. This is not surprising as larger schools may be able to spread a single expenditure over more students than a small school could. For example, if a school decides to purchase computer software for CATS preparation and uses the program for 100 students, the average cost per student will be lower than if a school uses the same software for only 20 students. Further investigation of survey data showed that the smallest 20 percent of schools, in student population, have total average per student expenditures attributed to CATS almost three and a half times greater than that of the largest 20 percent of schools.[5]

With wide differences in reported spending, it is difficult to determine a true level of expenditure per student in Kentucky related to CATS. In an ideal world, schools would report similar expenditures across the sample. However, as noted previously, it can legitimately be the case that different schools spend significantly different amounts per student based on how they prepare for and implement CATS. The differences could also stem from different judgments as to what is a CATS cost and how to categorize the cost. There is no way to verify with certainty the reason behind differences in the reported per student costs of CATS to schools.

CATS-related Expenditures Reported by Central Offices

The wide variation in reported expenditures in schools was also seen in spending reported by central offices.

Seventy-five percent of central offices responded to the cost survey. As with the schools, there was significant variation in the responses. Table 5.12 provides a listing of expenditure categories along with the average per student spending over the entire student population. In addition, Table 5.12 also presents the minimum and maximum per student spending reported along with the percentage of central offices that reported no expense for a given category.

The results for central offices are similar to that of the schools. A number of expenditure categories show considerable differences in the maximum reported per student expenditures compared to the minimum reported per student expenditures. In addition, 10 of the expenditure categories have more than half of responding central offices reporting no expenses, further demonstrating the differences among reporting entities.

[5] Some of this difference may also be attributed to differences among school types (elementary, middle, and high school) as well as among school populations (different socio-economic statuses, backgrounds, etc).

The smallest 20 percent of districts, in student population, reported spending twice as much per student as the largest 20 percent of districts toward CATS-related expenditures.

It is also true that the average of the central offices' per student expenditures is higher than the average for all students. Similar to schools, this implies that smaller districts tend to have higher expenditures attributed to CATS per student than do larger districts. However, the difference between average per student expenditure for the smallest 20 percent and largest 20 percent of responding districts, in student population, is smaller than that seen for schools: about double for central offices compared with almost triple for schools.

Table 5.12
Reported CATS-related Expenditures Per Assessed Student
Reported by Central Offices
FY 2004

Expenditure Category	Detailed Expenditure Category	Average Per Student Expenditure	Minimum Reported Per Student Expenditure	Maximum Reported Per Student Expenditure	Percent Reporting No Expenditure
Supplemental Personnel	Readers Per Individual Education Plan	0.08	0.00	3.96	82
	Interpreters Per Individual Education Plan	0.00	0.00	0.05	97
	Scribes Per Individual Education Plan	0.05	0.00	8.15	86
	Testers of homebound & expelled students	0.15	0.00	8.00	64
	Other supplemental personnel	0.48	0.00	39.69	83
	Other school personnel	2.27	0.00	105.61	62
District Assessment Coordinators	District assessment coordinators' salaries	4.77	0.00	66.84	8
Writing Consultants/ Cluster Leaders	Writing consultants	2.98	0.00	79.62	67
	Cluster leaders	0.75	0.00	16.91	34
Professional Development	Scoring portfolios training	0.38	0.00	11.56	47
	Alternate portfolio training	0.20	0.00	5.28	50
	Other CATS-related professional development	0.47	0.00	21.46	52
Equipment, Supplies, and Software	Software to analyze scores	0.41	0.00	6.31	77
	Other equipment or software	1.32	0.00	45.70	80
Other Costs	Other CATS expenses	2.04	0.00	32.38	47
Total		**$16.35**	**$0.00***	**$198.00****	

*Two central offices reported no CATS-related expenditures in the central office survey.
**This is the highest total expenditure per assessed student reported by central offices.
Source: Staff analysis

The highest reported level of CATS-related expenditures by expense category was for other supplemental personnel. The highest level of expenditure was reported to be $105.61 per assessed student. The central office that reported this figure reported the expenditures were for "CATS implementation." The total per student reported by this central office was almost $198 per assessed student.

Reported Funding Sources

The largest source of funds used to pay CATS-related expenditures was district general funds.

Of the estimated $16.4 million in expenditures by local schools and districts, approximately 64.7 percent was reportedly spent out of district general funds. The remaining 35.3 percent was spent from a variety of other sources. The survey results showed that central offices reported school district general fund dollars were used to cover approximately 72.9 percent of CATS-related expenditures, while schools reported a wider variety of sources being used. Schools still reported that district general fund monies accounted for more than half of expenditures. Table 5.13 lists all fund sources and the total percentage of spending they made up for in schools and central offices, as well as the percentage of total spending.

Table 5.13
Funding Sources of Reported Local CATS Costs
as a Percent of Expenditures
FY 2004

Fund Source	Central Office	Schools	Total
District General Funds	72.9%	55.0%	64.7%
Title I	7.8%	9.2%	8.5%
Other	5.2%	6.9%	6.0%
Professional Development	2.2%	10.3%	5.9%
Title II	6.2%	1.7%	4.1%
Extended School Services	0.7%	4.6%	2.5%
Individuals with Disabilities Act, Part B	4.2%	0.4%	2.4%
School Activities Fund	0.0%	4.6%	2.1%
Support Organizations	0.0%	4.6%	2.1%
Family Resource and Youth Services Centers	0.1%	1.7%	0.8%
Not Identified	0.7%	1.0%	0.9%
Total	**100%**	**100%**	**100%**

Source: Staff analysis of LRC survey to central offices and schools.

Total State and Local CATS-related Expenditures

Total state and local spending for CATS-related costs was estimated to be $37.2 million FY 2004.

Together, estimated state and local CATS-related expenditures total $37.2 million. This means that, on average, almost $78 was spent per assessed pupil as a result of CATS for fiscal year 2004.

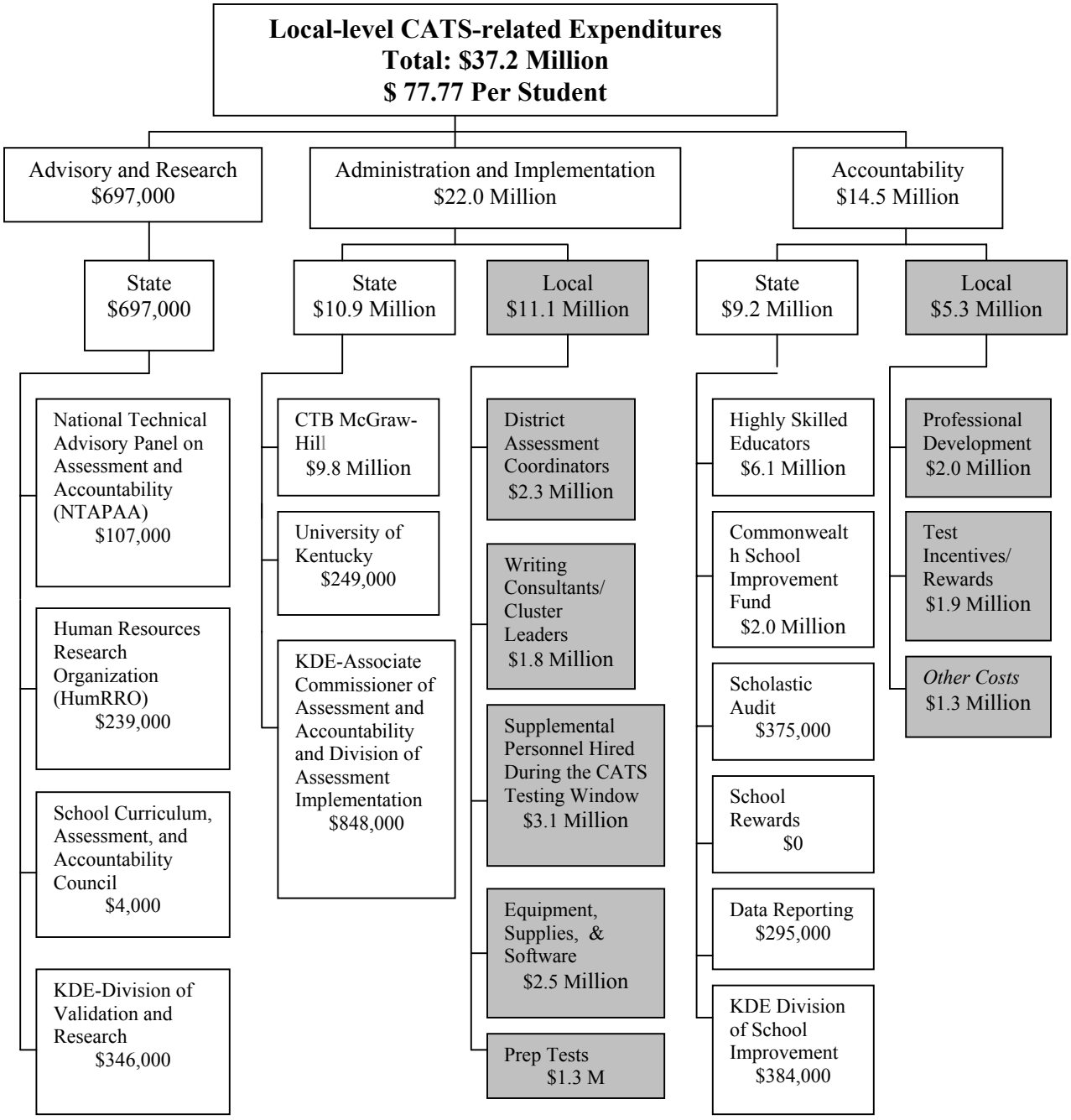
Figure 5.D presents a summary of the state-level and local-level expenditures related to CATS in FY 2004. Figure 5.D combines previous Figures B and C to show where CATS-related expenditures are made.

The largest component of reported expenditures is for administration and implementation, at about $22 million, which is fairly evenly divided between state and local expenditures. Accountability expenditures come to $14.5 million. For this category, state spending is almost twice that of local spending. The smallest category, which is entirely funded by state expenditures, is advisory and research, at less than $1 million.

While state-level expenditures were straightforward to obtain, local-level expenditures were not. In collecting local-level CATS-related expenditures through surveys of local school officials, significant variations in expenditures were reported for central offices as well as schools. While, as discussed previously, the cause of this variation is unclear, it is important to note when considering reported local-level expenditures.

To provide a context of per student CATS-related spending, the total estimated per assessed pupil CATS-related expenditure can be compared to total per student expenditures. KDE reports that in FY 2004, $8,029.84 was spent per student by local districts on average, exlcuding state-level assessment expenditures. Thus, the total per student estimated CATS-related expenditure is less than 1 percent (0.97 percent) of the total per student district-level spending.

Figure 5.D
Estimate of State-level and Local-level CATS-related Expenditures*
FY 2004



*Columns may not total due to rounding.
Source: Staff analysis.

Works Cited

Abedi, Jamal, and Carolyn Hofstetter. “Assessment Accommodations for English Language Learners: Implications for Policy-Based Empirical Research.” Review of Educational Research 74.1 (2004) 1-28.

Abedi, Jamal, Carolyn Hofstetter, Eva Baker, and Carol Lord. *NAEP Math Performance and Test Accommodations: Interactions With Language Background.* CSE Technical Report 536. Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing. Feb. 2001.

AEL. *Analysis of Kentucky Commonwealth Accountability Testing System Based on a Review of the Literature and Kentucky Documents.* Charleston, Feb. 11, 2005.

Brennan, Robert L. Coefficients and Indices in Generalizability Theory. Center for Advanced Studies in Measurement and Assessment. CASMA Research Report No. 1. Aug. 2003.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: 1999.

Commonwealth of Kentucky. Education Professional Standards Board. "Identifying 'Highly Qualified' Teachers Pursuant to the No Child Left Behind Act of 2001." Nov. 14, 2003. <http:kyepsb.net/documents/nclb/NCLBDEF9.doc> (accessed Feb. 21, 2005).

---. Kentucky Board of Education. "Issues Brief -- Assessment, Accountability and NCLB." Oct. 2003. <http://www.education.ky.gov/NR/rdonlyres/emt56by3whuj24jxzv7cibk2yw olpo3yrtcravl6dtu2d5zhimmunzetsatrtjt6g4qptpod4va44vbg3bezvju2ylc/IssuesBriefsAssesAccountNCLB.pdf> (accessed Feb. 21, 2005).

---. ---. "Issues Brief — Improving CATS for Optimum Results." Feb. 2004. <http://www.education.ky.gov /NR/rdonlyres/ew23ue4l6nerx6rpvrgjcg4rbf3uourzjromnvwquvxaw2qsz4ehyfqtng446mc7g44hlo6uza4pe4 yihbjykgrpi5e/IBfeb04cats.pdf> (accessed Feb. 21, 2005).

---. Kentucky Department of Education. *2004 CATS Interpretative Guide, Version 2.01.* <http://www.education.ky.gov/NR/rdonlyres/estue3du34uop7gly654o4qk254zn2s7hrsg4d6mow2sdepzmrs myvqes44pl7mbch5n2ha77mqgrl77vksjoemncze/InterpretiveGuide2004.pdf> (accessed March 17, 2005).

---. ---. “2004 No Child Left Behind (NCLB) Briefing Pack - Final.” Feb. 17, 2005. <http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/Reports/No+Child+Left+Behind+Reports/2004+No+Child+Left+Behind+-NCLB-Briefing+Packet--Final.htm> (accessed Feb. 21, 2005).

---. ---. *2004-2005 Nonacademic Data Report Guidelines*. Sept. 2004. <http://www.education.ky.gov /NR/rdonlyres/ewxh4sdukm2ll5t2rq6jdpe5r67no2eb6vdihpwl6cuznj54vdereohjen2ec4s5ico3bpmtur34il7bf 6n6jax4tjg/200405NonacademicGuidelines.pdf> (accessed March 17, 2005).

---. ---. *Alternate Portfolio State Summary 2004*. <http://www.education.ky.gov/NR/rdonlyres /ed54zzx4kzr5yqow6eqlbwajhtl4lwlln4f6jplynbelqfhwuaqa6o6nr7b2zwaljmgejcnxnako2cq57iqfaejr5d/AP State2004.pdf> (accessed March 17, 2005).

---. ---. *Alternate Portfolio State Summary 2003.* <http://www.education.ky.gov/NR/rdonlyres /ejcwd5d2ohvj55lsu6ij5xgt4fgrts5lx3d4reuupujjfxn6vydtgcitoxc7rha2gp5g4lahhfma66pnxkh7md722jh/AP State2003.pdf> (accessed March 17, 2005).

---. ---. *Grade 7 Sample Released Questions*. Jan. 2004. <http://www.education.ky.gov/NR/rdonlyres/eodfouo2tgcxpfcmlrt5oiqprs3v2khoxuxl74fdmkjmfs3el6tr24zbrti6krs7lgefffheub25cu6t2q6tkx5oada/2004ReleaseGrade7FINAL.pdf> (accessed March 17, 2005).

---. ---. *Grade 12 Sample Released Writing Tasks.* Jan. 2004. <http://www.education.ky.gov/NR/rdonlyres/e7cc2cbvch27m24xpeyv2bvibjypmsoj25kfei3luud4ukxrlaenva5ndqx2gawbr23epetznpkztomcm2mx6j46awb/2004ReleaseGrade12FINAL.pdf> (accessed March 17, 2005).

---. ---. "Inclusion of Special Populations in the State-Required Assessment and Accountability Programs" Feb. 2004. < http://www.education.ky.gov/NR/rdonlyres/ezi7gp7333hme4yl6aixinfvft3d7gzkrjxpcvd6c3tr6y7y3ggyxjm7ll7lf5a6t6joht2abroczxdfia2k6c3dvfb/InclusionsDocument.pdf> (accessed May 5, 2005).

---. ---. "Kentucky's Consolidated State Application Revised Accountability Workbook." June 15, 2004. <http://www.education.ky.gov/NR/rdonlyres/eqqehvdpisroibpr4rlzuzkgw6gqchvg3zyo6z7wa67h6kca7y4nbmt3jnm3da7c2sutsa3vgmxgfmnj4y24hjesrye/NCLB+PLAN+-+REVISED+june+15.pdf> (accessed Feb. 21, 2005).

---. ---. *Kentucky Core Content Test Scores Released.* Oct. 13, 2004.

---. ---. *Kentucky Core Content Tests 2002 Technical Report.* March 2004. <http://www.education.ky.gov/NR/rdonlyres/eyd2oftmvzjozuigtzp4cwhhlc3t7bzo4dylpqiss656xuyrjge7xmlunsi7jqqkfidmoelu7yvrfccvuzl7lw3jfhf/Kentucky+Core+Content+Tests+2002+Technical+Report+Final.pdf> (accessed March 17, 2005).

---. ---. "No Child Left Behind -- Unsafe School Choice Option." April 28, 2003. <http://www.education.ky.gov/NR/rdonlyres/ehilba4grsixv5pxv7gpbej3bunxwdbmqnwi7b4pprjyqvw63ujverszr5btmt3wdxyhrsgo5i6nkhs2qpqer2r3cwe/NoChildLeftBehindUSCODraft2April282003wpg.pdf> (accessed Feb. 21, 2005).

---. ---. *Nonacademic Data: 1993 to 2003 State Totals.* Aug. 16, 2004. <http://www.education.ky.gov/NR/rdonlyres/et5x5i2ystlrqqqf3ifv5oajp2cy5hxi7rgj26zpg77bwhtejce63h6c7mv3ozbyp7bkub6plekokwqboywd5sqpbad/NonacademicBrief2003.pdf> (accessed March 17, 2005).

---. ---. *Program of Studies.* <http://www.education.ky.gov/KDE/Instructional+Resources/Curriculum+Documents+and+Resources/Program+of+Studies> (accessed March 17, 2005).

---. ---. *Sample Released Questions with Annotated Student Responses, 2004 Grade 7.* <http://www.education.ky.gov/NR/rdonlyres/ewx4xuxhquyy4rc6zwqcryfxeql6yczl7zb4cn4las4hg7tumsq3rsx5yt3w32gsqgbucolniwtonjd4wwi7dt3hfgf/Phase22004ReleaseGrade7.pdf> (accessed March 17, 2005).

---. ---. *Spring 2004 Kentucky Performance Report.* <http://apps.kde.state.ky.us/secure_cats_reports_04/index.cfm?fuseaction=main.show_report&report=KPR04_999.PDF> (accessed March 17, 2005).

---.---. "Writing Development Teacher's Handbook." 2003. <http://www.education.ky.gov/NR/rdonlyres/ekgi53n3hb7xoo4ij6mzds3ao4a7dhisy6olmqvkoneguqqxm7ukqzkgscpbbu7dtxtwjpedaj6nqtv562n7k3njazc/WritingTeachersHandbook2003rev.pdf> (accessed May 5 2005).

---. Legislative Research Commission. Program Review and Investigations Committee. *The Commonwealth Accountability Testing System.* Research Report No. 312. Frankfort: LRC, 2003.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). "The dependability of behavioral measurements: Theory of generalizability for scores and profiles." *Journal of Educational Measurement* 30 (1972): 215–32.

CTB McGraw-Hill. *2003-2004 Writing Portfolio Audit Rationale and Procedures.* Kentucky Department of Education 2004.

Deffendall, Lisa. "Ratings Upend School Transfer Eligibilities – No Child Left Behind Scores Updated." *Lexington Herald-Leader* Oct. 17, 2004:C1.

Education Commission of the States. "No State Left Behind: The Challenges and Opportunity of ESEA 2001." Denver: ECS, 2002.

ETS. The Praxis Series. <http://www.ets.org/praxis/prxaboutII.html> (accessed April 18, 2005).

Hambleton, R.K., R.M. Jaeger, D. Koertz, R. Linn, J. Millman, and S.E. Phillips. *Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991-1994.* A report prepared for the Office of Education Accountability. Kentucky General Assembly: Frankfort, 1995.

Herman, Joan L., P.R. Aschbacher, and L. Winters. *A Practical Guide to Alternative Assessment.* Alexandria: Association for Supervision and Curriculum Development. 1992.

Herman, Joan L., Maryl Gearhart, and Eva L. Winters. "Assessing Writing Portfolios: Issues in the Validity and Meaning of Scores." *Education Assessment*. 1.3 (1993): 201-224.

Hoxby, Caroline M. "The Cost of Accountability." School Accountability: An Assessment by the Koret Task Force on K-12 Education. Williamson M. Evers and Herbert J. Walberg, eds. Stanford: Hoover Press, 2002.

HumRRO. *The Accuracy of School Classifications for the 2002 Accountability Cycle of the Kentucky Commonwealth Accountability Testing System.* Authored by R. Gene Hoffman and Lauress L. Wise. FR-03-06. Jan. 2003.

---. *The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2001 and 2002 Kentucky Core Content Tests.* Authored by R. Gene Hoffman. FR-02-46. Sept. 2002.

---. *The Accuracy of Students' Novice, Apprentice, Proficient, and Distinguished Classifications for the 2003 Kentucky Core Content Tests.* Authored by R. Gene Hoffman. FR-03-73. Sept. 2003.

---. *Variations on a Theme: Programs That Contribute to Success at 14 Kentucky Schools.* Authored by Arthur A. Thacker, Lisa E. Koger, Milton E. Koger, and R. Gene Hoffman. FR-01-31. Nov. 2001.

Gearhart, Maryl, Joan L. Herman, Eva L. Baker, and Andrea K. Whittaker. *Whose work is it? A Question for the Validity of Large-Scale Portfolio Assessment.* CSE Technical Report 363. Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing. July 1993.

Koretz, Daniel, Brian Stecher, Stephen Klein, Daniel McCaffrey, and Edward Deibert. "Can Portfolios Assess Student Performance and Influence Instruction? The 1991-92 Vermont Experience." CSE Technical Report 371. Los Angeles: University of California and the National Center for Research on Evaluation, Standards, and Student Testings. Dec. 1993.

Kosciolek, S. and J.E. Ysseldyke. *Effects of a Reading Accommodation on the Validity of a Reading Test.* Technical Report 28. Minneapolis: National Center on Educational Outcomes, Dec. 2000. <http://education.umn.edu/NCEO/OnlinePubs/Technical28.htm> (accessed March 24, 2005).

National Technical Advisory Panel for Assessment and Accountability. Memorandum to Marcia Seiler, deputy director of the Office of Education Accountability. Feb. 22, 2005.

Reckase, M.D. "Statistical Test Specifications for Performance Assessments: Is this an Oxymoron?" National Council on Measurement. Chicagom, 1993.

Shavelson, R.J., and N.M. Webb. *Generalizability theory: A primer.* Newbury Park: Sage Publications, 1991.

Simon, Ray. United States. Department of Education. Assistant secretary for Elementary and Secondary Education. Letter to Chief State School Officers. Oct. 22, 2004.

Sireci, Stehpen G. "Validity Issues in Accommodating NAEP Reading Tests." NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP. Jan. 2004. <www.nagb.org/pubs/conferences/sireci.pdf> (accessed March 22, 2005).

Swartz, C., Stephen R. Hooper, James W. Montgomery, Melissa B. Wakely, Renee E. L. de Kruif, Martha Reed, Timothy T. Brown, Melvin D. Levine, Kinnard P. White. *Using Generalizability Theory to Estimate the Reliability of Writing Scores Derived From Holistic and Analytical Scoring Methods*. Educational and Psychological Measurement, 59.3, SAGE Publications 1999, 492-506.

United States. Department of Education. "Accountability and AYP." National Title I Directors' Conference. Feb. 2003. < http://www.ed.gov/admins/lead/account/ayp203/edlite-index.html> (accessed Feb. 21, 2005).

---. ---. Title I, Part A Program. <http://www.ed.gov/programs/titleiparta/index.html> (accessed Feb. 21, 2005).

---. Government Accountability Office. *THE NO CHILD LEFT BEHIND ACT: Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions*. GAO 04-734. Sept. 2004.

Appendix A

Summary of Statutes and Regulations Governing CATS

Statutes

Legislative Intent

KRS 158.645 expresses recognition that public education involves responsibilities shared among the local communities, parents, students, school employees, and state government, and that the cooperation of all is needed to achieve desired outcomes.

This statute is also a declaration of legislative intent to create a system of public education that will allow and assist all students to acquire the capacities below.

1. Communication skills
2. Knowledge to make economic, social, and political choices
3. Core values and qualities of good character
4. Understanding of governmental processes
5. Sufficient knowledge of self and mental and physical wellness
6. Sufficient grounding in the arts to appreciate one's cultural and historical heritage
7. Sufficient preparation to intelligently choose and pursue one's life's work
8. Skills to enable one to compete favorably with students in other states

Goals for Schools

KRS 158.6451 is a declaration of goals for the Commonwealth's schools.

Schools shall expect a high level of achievement of all students and shall develop their students' ability to

1. use basic communication and math skills;
2. apply core concepts and principles learned in classes to real-life situations;
3. become self-sufficient individuals of good character;
4. become responsible members of a family, work group, or community;
5. think and solve problems in a variety of situations; and
6. connect and integrate experiences and new knowledge from all subject fields.

Schools shall also increase their students' attendance rates and reduce their dropout and retention rates. Schools shall be measured on the proportion of students who make a successful transition to work, post-secondary education, and the military.

This statute further requires the Kentucky Board of Education to disseminate to local districts a model curriculum framework directly tied to the goals, outcomes, and assessment strategies set forth in the statutes. The framework shall provide direction to local districts and schools as they develop their curricula. The framework shall identify teaching and assessment strategies, instructional material resources, ideas regarding incorporating the resources of the community, a directory of model teaching sites, alternative ways of using school time, and strategies to incorporate character education throughout the curriculum.

School Curriculum, Assessment, and Accountability Council

KRS 158.6452 creates the School Curriculum, Assessment, and Accountability Council to study, review, and make recommendations regarding setting academic standards, assessing learning, holding schools accountable, and assisting them to improve. The council shall advise the Kentucky Board of Education and the Legislative Research Commission on issues related to developing and communicating academic expectations and core content for assessment, developing and implementing the assessment and accountability program, distributing rewards and imposing sanctions, and assisting schools to improve performance under KRS 158.6453, 158.6455, 158.782, and 158.805.

The composition of the council shall be 17 voting members appointed by the governor, drawn from various defined classifications, with broad geographical representation of all school levels, equal representation of both sexes, and minority representation that reflects the composition of the Commonwealth. The council is attached to the Department of Education (KDE) for administrative purposes.

CATS and Its Components

KRS 158.6453 defines and creates the Commonwealth Accountability Testing System (CATS) to ensure school accountability for student achievement of goals set forth in KRS 158.645 and KRS 158.6451 and places the responsibility for creating and implementing CATS with the Kentucky Board of Education. The assessment program shall not measure a student's ability to become a self-sufficient individual or to become a responsible member of a family, work group, or community.

CATS shall include

1. a valid and reliable, customized, or commercially available norm-referenced test that measures the core content for assessment;
2. questions to assess student skills in reading, mathematics, science, social studies, the arts, the humanities, and practical living and vocational studies, and an on-demand assessment of student writing;
3. writing portfolios consisting of samples of student work;
4. performance assessment events for schools with students enrolled in performing arts organizations; and
5. a technically sound longitudinal comparison of the assessment results for the same students.

Role of Teachers in Designing CATS

Kentucky teachers shall have a significant role in designing CATS. CATS shall be designed to

1. measure grade-appropriate core academic content, basic skills, and higher-order thinking skills and their application;
2. provide valid and reliable scores; and
3. minimize time spent by teachers and students on assessment.

CATS results must be reported to districts and schools no later than 150 days following the first day the assessment can be administered.

Establishing Validity of CATS

KDE must gather information to establish the validity of CATS and must develop a biennial plan for validation studies. The statute sets forth the minimum studies that must be included in the plan.

In addition to creating and implementing CATS, the Kentucky Board of Education also has responsibility for assisting local school districts and schools in developing and using continuous assessment strategies needed to assure student progress.

School Report Cards

The Kentucky Board of Education is also required to promulgate regulations to establish the components of a school report card that clearly communicates with parents and the public about school performance. The report card shall be sent to the parents of the students of the districts, and a summary of the results for the district shall be published in the newspaper with the largest circulation in the county. The report card is required to include the following information, at a minimum, reported by race, gender, and disability when appropriate:

1. Student academic achievement, including the results from each of the assessments administered pursuant to this statute;
2. Nonacademic achievement, including the school's attendance, retention, dropout rates, and student transition to adult life; and
3. School learning environment, including measures of parental involvement.

Note: Language in the enacted budget for FY 2006 allows districts the alternative of publishing the information on their Web sites or having a printed copy available at a public library within the school district. If either of these methods is chosen, notification must be given in the newspaper with the largest circulation in the county.

National Technical Advisory Panel on Assessment and Accountability

KRS 158.6454 provides for the appointment of a National Technical Advisory Panel on Assessment and Accountability consisting of no fewer than three professionals with a variety of expertise in education testing and measurement. The panel shall provide advice regarding the implementation of KRS 158.6453 and 158.6455.

Legislative Intent for School Accountability System

KRS 158.6455 is a declaration of legislative intent and provides the framework for the school accountability system.

It is the intent of the General Assembly that "schools succeed with all students and receive the appropriate consequences in proportion to that success."

School Accountability Index

The Kentucky Board of Education is given responsibility for promulgating administrative regulations to establish a system for identifying and rewarding successful schools. Rewards are to be distributed to successful schools based on the number of certified staff employed in the school on the last working day of the year of the reward. The Board of Education shall identify reports, paperwork requirements, and administrative regulations from which high-performing schools shall be exempt. Effective July 1, 2006, the Board of Education shall reward schools that exceed their improvement goals and have annual average dropout rates below 5 percent. In calculating the dropout rate, schools must include students who were enrolled in the school for at least 30 days that school year. Schools shall not include students enrolled in district-affiliated alternative programs leading to a certificate of completion or GED diploma, or students who withdraw and obtain a GED by October 1 of the following year. Students in district-

affiliated alternative programs must participate in the appropriate CATS assessments.

The Kentucky Board of Education is given responsibility for promulgating regulations that set forth the formula for a school accountability index to classify schools every two years based on whether they have met their threshold levels for school improvement. The formula must reflect the academic goals set forth in KRS 158.6451.

A student's test scores must be counted in the accountability index of the school at which the student was enrolled for at least 100 days of the school year prior to the beginning of the statewide testing period. The scores are included in the district's accountability index if the student was enrolled in the district for at least 100 days of the school year prior to the beginning of the testing period. The scores are included in the state's index if the student is enrolled in a Kentucky public school prior to the beginning of the testing period.

The Kentucky Board of Education is further required to promulgate regulations to establish appropriate consequences for schools failing to meet their thresholds. The consequences shall be designed to improve teaching and learning and may include, but not be limited to,

1. a scholastic audit process to determine the appropriateness of a school's classification and to recommend needed assistance;
2. school improvement plans;
3. eligibility to receive Commonwealth school improvement funds under KRS 158.805;
4. education assistance from highly skilled certified staff under KRS 158.782;
5. evaluation of school personnel; and
6. student transfer to successful schools.

Scholastic Audits

The Kentucky Board of Education is also required to promulgate regulations establishing guidelines for conducting scholastic audits, which shall include the process for

1. appointing and training audit team members;
2. reviewing a school's learning environment and efficiency, students' academic performances, and the quality of the school council's data analysis and planning;
3. evaluating each certified staff member; and
4. making a recommendation to the Kentucky Board of Education about the appropriateness of the school's

classification and a recommendation concerning the assistance required by the school to improve teaching and learning.

For information purposes, the Kentucky Board of Education is also required to conduct scholastic audits in a sample of schools that achieved their goals and report the findings to the public.

The statute further requires the Kentucky Board of Education to promulgate regulations that establish a formula for school accountability and a school improvement goal for each school for the 1998-1999 and 1999-2000 school years. Rewards and audits were to be assessed for those years as set forth in the statute.

The Kentucky Board of Education was given discretion to promulgate regulations that created a system of district accountability similar to the system of school accountability, including a formula for accountability, goals for improvement, rewards for leadership in improving teaching and learning, and consequences for a district's failure to achieve its goals.

Appeals Process

The Kentucky Board of Education is also required to promulgate regulations establishing a process by which a school can appeal a performance judgment it considers grossly unfair. The Board of Education may adjust a performance judgment when evidence warrants the conclusion that the judgment is based on fraud or a mistake in computations, is arbitrary, is lacking any reasonable basis, or if there are significant new circumstances occurring during the biennial assessment period that are beyond the control of the school.

Definitions

KRS 158.6457 provides the following definitions:

1. "Accountability index" means the statistic that combines a school's academic and nonacademic factors;
2. "Core content for assessment" means the content identified for all students to know that is to be included on the state assessment; and
3. "Nonacademic factors" means the statistic that describes school success on increasing attendance and decreasing retention and dropout rates.

KDE To Develop Plan To Implement CATS

KRS 158.6458 requires KDE to develop a plan to implement CATS and to provide quarterly reports to the Interim Joint Committee on Education on its progress in the following nine areas:

1. Establishing a consistent structure of test components, test distribution, and test administration procedures;
2. Beginning a new cycle of equating procedures and conducting appropriate equating analyses;
3. Publishing informative guides for interpreting school accountability index score changes;
4. Reviewing school accountability classifications to assure their construct validity;
5. Maintaining and strengthening the annual audit of portfolio scores;
6. Developing and implementing a validity research plan;
7. Establishing additional routine audits of key processes;
8. Maintaining a library of technical documents and producing an annual technical report; and
9. Maintaining a vigorous ongoing program of research and documentation.

Kentucky Institute for Education Research Board

KRS 158.646 creates the Kentucky Institute for Education Research Board and defines its purpose, duties, and membership. Its two-fold purpose is to

1. solicit and raise funds to support the independent evaluation of KERA and related activities; and
2. serve as a stimulus and clearinghouse for KERA-related research projects.

The duties of the board are to

1. cause an in-depth evaluation of the impact of KERA to be performed;
2. make recommendations regarding enhancement of benefits of KERA and expansion and improvement of services to students;
3. develop the capacity to manage and coordinate research, conduct research, and design and implement a comprehensive educational data information system; and
4. prepare an annual report of its activities.

The board is to consist of 10 members initially appointed by the governor. The members are required to enact bylaws governing membership making the board self-perpetuating.

Education Assessment and Accountability Review Subcommittee

KRS 158.647 creates the permanent Education Assessment and Accountability Review Subcommittee and defines its membership and purpose.

It is to be composed of eight members, four from each legislative chamber. This subcommittee has responsibility for reviewing

administrative regulations and advising the Kentucky Board of Education concerning the implementation of CATS and for advising and monitoring the Office of Education Accountability.

Achievement Gaps

KRS 158.649 defines the term "achievement gap" and requires KDE to provide each school with performance data on its students disaggregated by race, gender, disability, English proficiency, and participation in the free and reduced-price lunch program.

The statute requires each local board of education to adopt a policy for reviewing the academic performance for various groups of students. It further requires the local boards to establish biennial targets no later than December 1, 2002, for each school for reducing identified achievement gaps.

By February 1, 2003, and each February 1 in odd-numbered years thereafter, the school-based decision-making council shall set the school's biennial targets for eliminating achievement gaps and submit the targets to the superintendent for consideration. With the agreement of the superintendent, the targets are submitted to the local board of education for adoption.

By April 1, 2003, and each April 1 in odd-numbered years thereafter, the school council, with the involvement of parents, faculty, and staff, shall review the data and revise the consolidated plan to include the biennial targets, strategies, activities, and a time schedule calculated to eliminate the achievement gap. The principal shall convene a public meeting at the school to present and discuss the plan prior to its submission to the superintendent and local board of education. At a minimum, the plan shall address the following areas:

1. Curriculum alignment;
2. Evaluation and assessment strategies to monitor and modify instruction to meet student needs and support proficient student work;
3. Professional development;
4. Parental communication and involvement;
5. Attendance improvement and dropout prevention; and
6. Technical assistance.

Local boards shall determine if each school has achieved its biennial targets for each group of students. If a school has not met its target, the board shall require the council to submit its revisions to the consolidated plan describing the use of funds to reduce the school's achievement gap. The plan must address how the school will meet the academic needs of the group of students at issue.

If a school fails to meet its targets to reduce the achievement gap for any student group for two successive biennia, the superintendent shall report that failure to the commissioner for education. The school's consolidated plan shall be subject to review and approval by KDE, and the school shall submit an annual status report. If a school meets its biennial target for reducing the achievement gap, the council is no longer required to seek approval of its plan.

Regulations

Writing Portfolio Procedures

703 KAR 5:010 establishes procedures to reduce the teacher and student time involved in preparing a writing portfolio.

A five-piece portfolio shall be produced in 12th and 7th grades, and a four-piece portfolio shall be produced in 4th grade. Schools and districts shall also develop a procedure to collect writing pieces at nonaccountability levels that are appropriate types of writing for portfolio categories. These pieces may serve as rough drafts that can be edited for inclusion in the accountability portfolio or they may be included as finished products.

Each school and district shall provide support for teachers to attend professional development focused on the types of writing assessed in the portfolios. Each school and district shall develop procedures for scoring portfolios that include an adequate number of teacher scorers to limit the number of portfolios scored by any one teacher to 30. Conferencing on portfolio pieces shall be focused on one or two areas of need, addressing patterns of errors that occur frequently.

Teachers shall allow students to use word processing during the development of writing pieces or allow students to submit pieces in their own handwriting. Teacher-assigned writing tasks shall relate to the content being studied. Teachers shall choose content area readings that represent the kinds of writings the students are asked to include in the portfolios, allowing the discussion of content and writing form to occur at the same time.

Assessment Systems

703 KAR 5:020 establishes a single assessment system with two accountability dimensions: one addressing the statutory CATS requirements and one addressing the federal requirements under the No Child Left Behind Act of 2001 (NCLB).

Section 1. This section establishes the testing timetable. The Kentucky Department of Education shall administer the Kentucky Core Content Tests and norm-referenced tests. Core content tests are to be administered as follows:

a. Reading at grades 4, 7, and 10;
b. Math at grades 5, 8, and 11;
c. Science at grades 4, 7, and 11;
d. Social studies at grades 5, 8, and 11;
e. Arts and humanities at grades 5, 8, and 11;
f. Practical living/vocational studies at grades 5, 8, and 10;
g. Writing at grades 4, 7, and 12;
h. Writing portfolio at grades 4, 7, and 12; and
i. Alternate portfolio at 4, 8, and the last anticipated year of attendance at the high school level.

Norm-referenced tests shall be administered in reading/language arts and math at the end of primary, grade 6, and grade 9.

In order to comply with NCLB, KDE shall augment the norm-referenced test to appropriately measure Kentucky's core content in reading and math at grades 3 and 6. Additional augmented norm-referenced tests shall be administered in reading at grades 5 and 8, and in math at grades 4 and 7.

If a school is selected by the U.S. Department of Education or its contractors to participate in the state National Assessment of Educational Progress in reading, math, and science at grades 4 and 8, the school shall participate fully.

Calculating Academic Indices

Section 2. This section provides instructions in calculating the academic indices. A school shall be held accountable based on an aggregated average of the performance of the students who have been enrolled in the school for a full academic year in the accountability grades. The points awarded for students at each achievement level are as follows:

1. Nonperformance – 0 points;
2. Medium novice – 13 points;
3. High novice – 26 points;
4. Low apprentice – 40 points;
5. Medium apprentice – 60 points;
6. High apprentice – 80 points;
7. Proficient – 100 points; and
8. Distinguished – 140 points.

Values for attendance rate and successful transition to adult life rate shall be the actual percentage reported. The values for

retention rate and dropout rate shall be 100 minus the actual percentage reported. Alternate portfolio scores shall be included so that they contribute the same weight as do scores for other students. The same requirement shall apply to calculations required by NCLB.

Section 3. This section defines the two components of the accountability index and explains how to calculate those components. It assigns various weights by content area and grade level that are to be used in calculating component one. Component two is to be calculated using a national norm-referenced test and the scores assigned by achievement level set forth above.

Section 4. This section addresses how scoring should be undertaken for those schools that do not conform to standard grade configurations. Schools may request a waiver from scoring requirements and specify other combinations of schools and assessment data so long as all students in an accountability grade are included.

Section 5. This section provides that if a school has more than one accountability level (elementary, middle, or high school), that school's accountability index shall be the average of the academic and nonacademic data for the school, both under CATS and NCLB.

Section 6. This section defines "reconfigured schools" and provides for their treatment in accountability decisions. Reconfigured schools are those in which less than 80 percent of the school's population is stable. A reconfigured school shall have the performance judgment that would have applied to the district at that level. In the alternative, a school district may submit a plan for reconstituting baseline data taking into consideration the changes in service areas and assuring that local district calculations are accurate and include all student data. Under NCLB, a reconfigured school must meet the annual measurable objectives only in reading and math to show adequate yearly progress for the first three years after reconfiguration, provided it has a participation rate of at least 95 percent. In the alternative, a school district may submit a plan for reconstituting data as mentioned above.

Section 7. This section provides how to establish expected levels of growth for each school and establishes five points of school recognition for the purpose of recognizing school standing.

School Classifications and Rewards

Section 8. This section defines the following school classifications:

1. "Meets goal,"
2. "Progressing,"
3. "In need of assistance," and
4. "Commonwealth pace-setter."

A school's classification is determined through a comparison of the school's growth accountability index for a biennium with a corresponding goal point and assistance point. This section also sets forth requirements schools must meet in order to receive rewards. A pace-setter school shall be entitled to one share of rewards if not otherwise receiving rewards for growth.

Section 9. This section establishes the levels of rewards available to schools. A school classified as meets goal shall earn three shares of rewards. A progressing school shall earn one-half share of rewards. Additionally, schools that meet or exceed school recognition points are entitled to a one-time reward of one share for each point met or exceeded. Numbers of shares earned are multiplied by the number of certified staff to determine the final reward amount.

NCLB Accountability System

Section 10. This section describes the NCLB accountability system requirements and consequences. It requires KDE to establish a single starting point in reading and math for elementary, middle, and high schools by looking at data from the 2001-2002 school year and determining the percentage of students at or above the proficient level who are in the school at the 20^{th} percentile in the state.

In determining adequate yearly progress, a school is held accountable based on an aggregated average of the performance of the students who have been enrolled in the school for a full academic year in the accountability grades.

A school has made adequate yearly progress if

a. the school and subpopulations of sufficient size have met district annual measurable objectives in reading and math, or have met the "safe harbor" provisions;
b. the school showed progress or met the CATS accountability index goals at the elementary and middle school accountability levels;
c. the school showed progress or met the annual goal for graduation rate; and

d. the school had a participation rate of at least 95 percent of the enrolled students and subpopulations of sufficient size.

A school that fails to make adequate yearly progress in reading and math for two consecutive school years shall be identified as a "NCLB improvement school." A school that makes adequate yearly progress in both reading and math for two consecutive years shall receive a reward or recognition from KDE.
A school district shall provide the school with an opportunity to review the relevant data before identifying a school as an improvement school and implementing consequences.

Students enrolled in a school on the first day of the testing window for the school shall be included in the calculation of the participation rate. Students enrolled in a school for a full academic year shall be included in the school calculation of the percentage of students performing at the proficient level or above in both reading and math.

The regulation prescribes the annual measurable objectives in reading and math for the years 2003-2014.

Section 11. This section describes the consequences for failing Title I schools under NCLB.

If a Title I school is identified as an improvement school, the local district must provide parental notification as required by federal law, including information that the students have the option to transfer to another public school in the district that is not an improvement school.

If a Title I school fails to make adequate yearly progress in both reading and math by the end of one full year after being identified as an improvement school, the local district shall provide the notice described above and shall also require the school to provide supplemental services.

If a Title I school fails to make adequate yearly progress by the end of two full years after being identified, the local district shall provide the services required above and shall also take corrective action as required by federal law.

After a third year of failure to make adequate yearly progress, the local district shall plan for alternative school governance as required by federal law. After a fourth year of failure, the alternative governance plan shall be implemented.

Section 12. This section addresses the duration of consequences. A school that makes adequate yearly progress in both reading and math for two consecutive years after identification shall no longer be identified as an NCLB improvement school and shall not be subject to federal consequences.

School/Program Categories and Accountability

703 KAR 5:040 defines the characteristics of A1 schools and A2-A6 programs and explains how accountability indices and rewards will be determined for the various categories.

Section 1. This section addresses the types of schools and programs included in the accountability indices. All A1 schools and A3, A5, and A6 programs must be included if the programs are final programs of placement. A final program of placement is one that has at least 10 students in each NCLB assessment grade (at least 60 students combined); and is one in which the student is placed to complete the elementary, middle, or high school program and is not expected to transition back to the A1 school.

Note: This regulation previously defined A1-A6 schools (e.g. technical schools, special education schools, preschools, alternative schools) but, as amended September 13, 2004, it no longer defines those schools or programs.

Section 2. This section establishes which schools and districts are accountable for which students.

A student enrolled in an A1 school and those enrolled in an A1 school but attending an A2 program shall be counted in the membership of the A1 school and attributed to it for accountability purposes. Students enrolled in an A4 program do not participate in the state accountability system.

Students enrolled in A3, A5, or A6 programs shall be assigned to A1 schools, districts, or the state for accountability purposes as follows:

a. If an A1 school places a student in an A3, A5, or A6 program, or if the placement is the result of local district policies, the student's assessment data is attributed to the sending A1 school and district. If it is a final placement, it is attributed to the sending school, the program, and the district.
b. If a student is placed in a program by a court, agency, other A1 school, or is self-placed, and the student has been enrolled in a single A1 school or district for a full year prior to placement, the data shall be attributed to the A1 school

or district where the student was enrolled for the full academic year.

Section 3. This section addresses accountability for those students not attributed to an A1 school or A3, A5, or A6 program. Those students' data shall be aggregated into a state-level accountability report. The commissioner of education shall appoint a committee to review the performance of these programs.

Section 4. This section discusses accountability for nonacademic data. All schools and programs must observe nonacademic-data collection procedures. Data collected from A4 programs shall not be included in accountability calculations. Students in A3, A5, or A6 programs who earn a diploma or certificate shall be monitored by staff for purposes of reporting data on transition to adult life.

Section 5. This section discusses placement of students in A3, A5, and A6 programs. Except for those students placed by courts and other government agencies, students shall first be entered into an appropriate A1 school, which must determine the most appropriate means of delivering services to the students and can then place the students in the appropriate programs if warranted.

Section 6. This section outlines how rewards and assistance are determined for A2, A3, A4, A5, and A6 programs. Accountability indices are calculated for final programs of placement as described in Section 1. Other programs are considered attached to the district central office for purposes of rewards and assistance. Programs that serve multiple school districts receive rewards if more than 10 percent of the students in accountability grades come from districts that receive rewards. They are subject to assistance if more than 50 percent come from districts that are subject to assistance.

Appeals of Performance Judgments

703 KAR 5:050 establishes procedures for a school to appeal a performance judgment it considers to be grossly unfair.

It defines the following terms:

1. “Baseline accountability index;”
2. “Growth accountability index;” and
3. “Performance judgment.”

A school may request a data review by submitting the request to the commissioner of education within 14 days after KDE officially releases the performance judgments to the public. If the performance judgment is revised or the school is not satisfied with

the results of the data review, the school shall submit a written appeal of the performance judgment to the commissioner within 30 days after the school has received notification of the review results.

A school may appeal a performance judgment by submitting a written appeal to the commissioner within 45 days after KDE officially releases the performance judgments to the public. The appeal must clearly identify the basis for the wrongful effect on the accountability indices and shall detail the requested adjustment to be made to one or more of these indices.

The commissioner shall appoint a committee to review the pending appeals and to make recommendations to the commissioner as to whether to dispute an appeal. A hearing officer shall conduct a hearing and shall submit a written recommended order to the Kentucky Board of Education for the board's consideration in rendering its final order.

Special Student Populations

703 KAR 5:070 establishes procedures for including special student populations in the state-required assessment and accountability programs.

It incorporates by reference the document titled, "Inclusion of Special Populations in the State-Required Assessment and Accountability Programs" February 2004.

Special populations include

1. students with disabilities;
2. students who attend schools classified as A2-A6 and state agency children;
3. students whose primary language is not English;
4. students receiving instruction in home/hospital settings (for example, homebound instruction, not home schools); and
5. students who have temporary medical conditions that necessitate accommodations or modifications or both for participation.

The provisions of this regulation apply to all students except those 21 years of age or older who are part-time students attending less than 6 hours per day and those enrolled in the Adult General Education Diploma Program (GED). Students enrolled in a Secondary GED Program offered by a public high school are not considered dropouts and are therefore subject to the inclusion policies.

All students with disabilities shall participate in the state-required Assessment and Accountability Programs. A small percentage of students shall participate in the Alternate Portfolio Assessment Program.

Schools shall assess all students with limited English proficiency enrolled on the first day of the testing window.

Students receiving instruction in home/hospital settings shall participate in the state-required Assessment and Accountability Programs unless participation would jeopardize a student's physical, mental, or emotional well-being and a school or district has submitted a request for medical exemption describing the medical condition.

Students With Disabilities

Section 1 – Inclusion of Students With Disabilities

Students with disabilities are students who meet the criteria under KRS 157.200 and the Kentucky Administrative Regulations related to Exceptional Children (707 KAR Chapter 1) or Section 504 of the Rehabilitation Act of 1973.

For students with disabilities, the Admissions and Release Committee or Individual Education Program (IEP) or 504 Committee shall determine on an individual basis how the student will be included in the state-required Assessment and Accountability Programs. Below is a description of the three options for inclusion and which students shall participate under each option.

1. Participation with no accommodations or modifications. This option includes students referred for evaluation but for whom the evaluation process has not been completed, and those students with disabilities who are not receiving special education and related services.

2. Participation with accommodations or modifications or both. This option includes students who meet the eligibility criteria for one of the disability categories under the Kentucky Administrative Regulations, have a current IEP, and are receiving specially designed instruction. The students must also currently be using accommodations or modifications as part of their regular instructional routines. The accommodations or modifications must be related to the individual student’s needs and the impact of the disability and must be specified in the student’s IEP.

Students who meet the eligibility criteria in Section 504 of the Rehabilitation Act of 1973 for having a physical or mental disability that substantially limits one or more major life activities and have a current 504 Plan may use accommodations or modifications for the assessment if the accommodations or modifications are part of the student's regular instructional routine, are related to the student's needs and the impact of the disability, and are specified in the student's 504 Plan.

Examples of conditions that meet the definition under the 1973 act include cerebral palsy; visual, speech, and hearing impairments; epilepsy; and cancer.

Alternate Portfolio Assessment Program

3. Participation in the Alternate Portfolio Assessment Program. This option includes only those students who meet all the criteria for the certificate program as stated in 707 KAR Chapter 1 related to Exceptional Children and the Program of Studies (704 KAR 3:303).

The results of the alternate portfolio assessment shall be included in the accountability calculations and shall be equivalent to the impact of a student participating in the regular Assessment and Accountability Programs process.

Alternate portfolios shall be completed once each in elementary, middle, and high school. The elementary portfolio shall be completed in the 4th grade, the middle school portfolio in the 8th grade, and the high school portfolio in the last full year of school.

Students with disabilities in nongraded programs who do not participate in the Alternate Portfolio Assessment Program shall be required to participate in the state-required Assessment and Accountability Programs at their present appropriate grade assignment with accommodations or modifications if appropriate.

A student with disabilities who turns 21 during a school year may "age out" of school without completing the school year and participating in the assessment program. If this is a possibility, the student shall be included in the assessment during the school year prior to turning 21.

Students who skip a grade within the accountability system shall still participate in the assessment components associated with the grade being passed through.

With the exception of students who participate in the Alternate Portfolio Assessment Program, schools that serve disabled students shall be held accountable for these students' transitions to adult life using the same standards applied for all other students.

Students who participate in the Alternate Portfolio Assessment Program may be considered to have made successful transitions to adult life if they

1. make successful transitions as defined for the regular population;
2. enroll as full- or part-time students at postsecondary vocational schools or adult education programs preparing students for integrated work;
3. work in an integrated setting at least 10 hours per week;
4. participate in supported employment; or
5. transition to community rehabilitation and their training or employment takes place in an integrated environment.

Students in A2-A6 Schools

Section 2 – Inclusion of Students in A2-A6 Schools and State Agency Children

Students attending schools classified as A2-A6 shall be included in the overall accountability program. These students' scores shall be tracked back to the A1 schools that would have served them had they not required the services offered by the A2-A6 schools they attend.

State agency children shall have the same assessments administered and the results shall be included in the accountability index of the last A1 school the student attended or the school the student would have attended in that district.

Students With Limited English Proficiency

Section 3 – Inclusion of Students Whose Primary Language Is Not English

For accountability purposes, students who have attained English proficiency (based on a state-approved English proficiency assessment in conjunction with professional judgment) shall remain in the limited English proficient (LEP) subgroup for two years but not be counted in determining whether the group meets the state-defined minimum group size.

State-required Assessment and Accountability Programs

Schools shall assess and be held accountable for all students with limited English proficiency enrolled on the first day of the testing

window unless the students are in their first year of enrollment in a United States school.

LEP students in their first year of enrollment in a U.S. school shall be required to take an English language proficiency assessment. Students who enroll in a grade in which an NCLB-required math test is given must take the test with accommodations and/or modifications as appropriate.

All LEP students enrolled on the first day of the testing window shall be included in the school and district's participation rate. LEP students in their first year in U.S. schools shall be included in the participation rate based on their participation in the NCLB-required math test if administered in that grade, and based on participation in the English language proficiency assessment if in another grade (or the NCLB reading test if administered). LEP students in their first year in U.S. schools are not required to participate in the state-required reading, science, social studies, practical living/vocational studies, arts and humanities, or writing-on-demand assessments.

Under both CATS and NCLB, schools (and districts) shall be held accountable based on an aggregated average of the academic performance of students enrolled for a full academic year in the accountability grades. This applies to LEP subgroups of sufficient size except for students in the first year in U.S. schools.

Schools and districts may choose to include results for LEP students in their first year in U.S. schools if they have been enrolled for a full academic year as defined in 703 KAR 5:001. If the option is exercised, the decision shall be consistent across all content areas for the student.

It may be necessary to permit instructionally consistent accommodations or modifications or both for LEP students taking assessments. It must be based on an assessment of proficiency, consistent with the normal ongoing delivery of instructional services, and stated in the student's Program Services Plan.

If an LEP student is not in the first year in a U.S. school and has been in an English language instructional environment for at least two full school years, the student shall be required to submit a writing portfolio and shall be included in the accountability calculations. Other LEP students may be allowed to develop portfolios, but they shall not be included in the accountability calculations.

State-required English Language Proficiency Assessment

A local school district shall administer a home language survey to students enrolled in the district as the first screening process to identify students with limited English proficiency. If the answers indicate a language other than English is spoken at home, the district shall administer an annual assessment of English proficiency approved by KDE. The results in conjunction with professional judgment shall determine whether the student is identified as limited in English proficiency.

School personnel shall determine on an individual basis whether students with limited English proficiency will participate in state-required assessments with or without accommodations or modifications or both. A student with limited English proficiency may use accommodations or modifications or both if the student

1. has been assessed and meets the criteria as a student with limited English proficiency;
2. has evaluation data that demonstrate a need for accommodations or modifications;
3. has a current Program Services Plan that includes accommodations or modifications; and
4. is participating in instructional programs and services to meet the student's language and academic content needs.

Accommodations or modifications or both shall be related to the individual student's needs and the impact on the student's ability to access the curriculum and demonstrate knowledge on a test written in English. Decisions shall be made in the best interest of the student. Specific documentation must be provided if accommodations or modifications or both are needed.

Accommodations or modifications or both may include various administration strategies such as reading the text in English, paraphrasing instructions, or translating the text of questions. Every effort should be made to distribute test forms randomly, but if sufficient numbers of interpreters are not available for LEP students, schools may allow students to use the same test form so that one interpreter per foreign language may be used. Various response strategies may be allowed such as allowing students access to a foreign language dictionary and use of grammar and spell-check systems. Accommodations or modifications shall not be made solely for the state-required assessment.

A student with limited English proficiency may submit a portfolio in a language other than English if the student's daily instruction

and class work are conducted in the student's native language and the local scorer or a scorer hired by the district is both fluent in that language and trained to score the portfolio.

Home and Hospital Students

Section 4 – Inclusion of Students Receiving Instruction in Home/Hospital Settings

School personnel shall determine on an individual basis how each such student will participate in the state-required assessment programs. A student shall either participate fully or is exempted medically if participation would jeopardize a student's physical, mental, or emotional well-being and the school submits a request for medical exemption subject to approval by KDE. An identified disability or handicapping condition alone is not sufficient reason for granting medical exemption. If a student with disabilities is receiving instruction temporarily or long term in a home/hospital setting, the procedures described in Section 1 shall be followed.

Students With Temporary Medical Conditions

Section 5 – Inclusion of Students With Temporary Medical Conditions That Necessitate Accommodations or Modifications or Both for Participation

Students who become injured or develop an ailment before or during the testing window may be allowed appropriate accommodations or modifications or both to allow their participation in the state-required assessment programs.

Appropriate Accommodations and Modifications

Section 6 – Conditions for Implementing Accommodations or Modifications or Both

Accommodations or modifications or both shall be age appropriate and related to the student's verified disability and to specially designed instruction. They shall be based on the student's individual needs and not a disability category or designation as LEP. They shall be part of the student's ongoing instructional program and not introduced for the first time during assessment. They shall not inappropriately impact the content being measured.

A variety of accommodations and modifications may be appropriately used for students with disabilities, including readers, scribes, paraphrasing, use of technology and special equipment, extended time, reinforcement and behavioral modification, manipulatives, prompting or cueing, and interpreters.

Any individual who provides assistance to a student with disabilities during the assessment shall be trained in that role and

shall abide by confidentiality laws, ethics provisions, and the conditions for use as described in the student's IEP. Particular forms of the test should be randomly distributed just as with students without disabilities, with the exception of students with limited English proficiency as described in Section 3.

Readers. If listening to a reader is the normal mode by which the student is presented regular print materials, reading assessments may be read to a student. The Admissions and Release Committee (ARC) or 504 Committee shall have considered under what conditions a student will use a reader on a routine basis during instruction. On-demand tasks may be read to students under certain, specified conditions. A reader shall not inappropriately impact the content being measured. A reader shall read information as written and shall not use information to lead the student to specific information needed to answer questions and shall not point out parts of a task or question skipped by the student. A reader shall reread portions only if specifically requested by the student.

Scribes. The ARC or 504 Committee shall consider under what conditions a student will use a scribe on a routine basis during instruction. A scribe may be used for state assessments under certain, specified conditions. Technology and natural supports shall be used prior to the more intrusive process of using a scribe. A scribe shall not inappropriately impact the content being measured. A scribe's role shall be to record the student's work to allow the student to reflect what the student knows and is able to do while providing the student with an alternative means to express thoughts and knowledge.

Paraphrasing. The ARC or 504 Committee shall consider under what conditions a student will use paraphrasing on a routine basis during instruction. On-demand tasks may be paraphrased under specified conditions. Paraphrasing for the state assessment shall be consistent with classroom instruction and can include repeating or rephrasing the directions, prompt, or situation, but shall not include defining words or concepts or telling a student what to do and the order in which it should be done. Stories and content passages may not be paraphrased. A paraphraser shall not inappropriately impact the content being measured.

Use of Technology and Special Equipment. The ARC or 504 Committee shall consider under what conditions a student may use technology on a routine basis during instruction. During the state-required assessment, a student with a disability may use special equipment that is part of the student's regular instructional routine.

If it is necessary for a student with special needs to complete written work on a computer and this procedure is routinely used in the student's regular instructional program and noted in his or her IEP or 504 Plan, it may be used when responding to open-response questions while participating in the state-required assessment under specified conditions intended to protect the security, confidentiality, and integrity of the assessment.

Extended Time. Students with disabilities shall be allowed extended time to complete items on state-required tests if they have IEPs or 504 Plans that stipulate extra time is needed and if extended time is an accommodation for assessments and completion of assignments as part of their daily instructional routines.

Reinforcement and Behavior Modification Strategies. Students with disabilities who have IEPs or 504 Plans that stipulate the use of reinforcement or behavior modification strategies and for whom the use of such strategies is implemented during routine instruction, may use these strategies on the state-required assessment. They may also be implemented for any student who displays aggressive or disruptive behavior during testing.

Manipulatives. Manipulatives may be used to complete the state-required tests and the development of portfolios if they are a strategy used by the student to solve problems routinely during instruction and if the use of manipulatives is described in the student's IEP or 504 Plan.

Prompting or Cueing. The ARC or 504 Committee shall consider under what conditions a student will use prompting and cueing on a routine basis during instruction. If a student uses a cue card or other strategy on a daily basis during instruction as stipulated by the student's IEP or 504 Plan, the student may use the cue card or strategy during the state-required assessment. The teacher shall not draw figures, suggest leading sentences, or provide content during the administration of the state-required standardized assessment.

Interpreters for a Student With a Hearing Impairment. The state-required tests may be translated to the student in sign language under specified conditions, including the use of sign language in the student's IEP or 504 Plan and the use of signing as part of the student's regular instructional routine. Signing shall not be a replacement for technology or reading instruction. The interpreter shall not indicate correct answers to test items.

Oral Interpreters for an LEP Student. The state-required tests may be interpreted orally into the primary language of an LEP student only if

1. the student has been officially designated as LEP;
2. the student's Program Services Plan includes specific goals and objectives related to reading, communication, and language development;
3. evaluation information supports the needs for the accommodation; and
4. the Program Services Plan documents that translation is part of the student's ongoing delivery of instruction and is necessary.

Administration Code

703 KAR 5:080 establishes an Administration Code for Kentucky's Educational Assessment Program of appropriate testing practices for state-required tests. The "Administration Code for Kentucky's Educational Assessment Program" dated June 1999 is adopted and incorporated by reference. Sections I to V of the code are summarized below.

I. Rationale

This document describes the practices considered appropriate in preparing students for assessments, in administering them, and in providing for proper security of the assessment materials. No test preparation practice shall violate the ethical standards of the education profession in 704 KAR 20:680. No test preparation practice shall increase students' test scores on the statewide assessment components without simultaneously increasing students' abilities to apply the content tested to real-life situations. All assessment work shall be done entirely by the student.

II. Appropriate Assessment Practices

Each individual involved in any component of the assessment must read, sign, and comply with the Administration Code and receive training on it. Each test administrator or proctor shall sign a verification form stating he or she has received and read this Administration Code and instruction manual.

Test Security. District assessment coordinators, administrators, and teachers shall ensure the security of the assessment materials before, during, and after test administration. It is appropriate for teachers to know and teach the concepts measured by the statewide assessment, but secure test materials shall not be reproduced in any way nor shall notes be taken regarding any secure test item. Tests

shall be distributed in the order in which they are received in shrink-wrapped packages. No one may have test booklets without authorization. No one may show items in the test booklets to anyone not administering the test. No one may reveal the content of any secure test item or use that knowledge to prepare students for the assessment. Test administrators must destroy any notes, drafts, or scratch paper produced by students and must ensure that any testing materials reused from previous years are free of any marks.

In those instances in which computer technology must be used to provide access to tests, secure materials shall be scanned to and stored on floppy disks that shall be returned to the contractor. When space requirements are prohibitive, the material may be scanned to larger disk drives if the district staff can assure the security of the assessment. Student responses that reveal that the students may cause harm to themselves or others or are suffering abuse may be copied in relevant part and turned over to appropriate authorities.

Procedures for Reporting Errors in Assessment Materials. The test item containing the error shall not be reproduced. Rather, the location of the error shall be identified, and the error shall be summarized for the district assessment coordinator who shall notify KDE.

Classroom Materials. Materials may be placed on classroom walls and bulletin boards for instructional purposes, and lesson plans shall indicate the relationship between the materials and instruction. Staff shall not place materials specifically designed for assessment on classroom walls or bulletin boards.

Dictionaries and thesauri may be used only on the writing-on-demand subtest. Students shall have access to calculators as designated in the administration manuals. Blank writing or graph paper and highlighters or markers may be made available. Other information and materials not sent as part of the assessment materials shall not be made available to students. Students shall not leave the testing area to gain access to calculators or other resources. Materials placed on classroom walls for instructional purposes shall not be moved to other locations for assessment purposes.

Administration Practices. Building personnel and district assessment coordinators shall schedule test administration; arrange for adequate staff to administer the assessment; prepare an accurate

student accountability roster; and ensure that assessment materials are kept secure before, during, and after testing sessions.

During testing, words of encouragement and general instructions that do not imply evaluation of student work are permissible. Accommodations or modifications or both may be provided if consistent with a student's IEP, 504 Plan, or LEP Plan and the routine delivery of instructional services. Tests should be scheduled to avoid conflicts with lunch. Interval or restroom breaks may be conducted, but the integrity of testing shall not be affected. Test sections shall be administered in the order in which they appear in the test booklets. Time limits and specific directions in the manuals shall be observed. Test administrators shall circulate throughout the testing site to monitor students as they work. When students need extended time to complete a test session, the additional time shall begin immediately following the initial administration.

During testing, test administrators shall not answer student questions that would aid the student in responding to any item on the test nor shall they assist the student in understanding the question. Test administrators shall not encourage students to edit their responses by providing any evaluation of student work. Students shall not take more than a single school day to complete a testing session except where there is a documented student illness or emergency. A student shall not be left alone in a room to take the test nor shall a student be allowed to take a test booklet or answer booklet out of the testing area without supervision.

Disciplinary Practices and Student Motivation. Administrators may direct students to apply themselves to the task at hand but shall not give any direction that would enable a student to better understand the task or to gain advantage in responding to the task. Student responses may be visually scanned after the testing session to determine disciplinary problems. If disciplinary problems are determined to exist, students shall not be allowed to modify their initial response. If a student's responses are found to contain inappropriate language, the student may be instructed to answer the questions again for disciplinary purposes. Both the original and rewritten responses shall be submitted to the testing contractor with the rewritten ones clearly marked NOT TO BE SCORED—ITEMS RETAKEN FOR DISCIPLINARY PURPOSES. Student responses may be visually scanned during or after the testing session to determine good-faith efforts, but no evaluative statements shall be made until the entire assessment has been administered and submitted to the district assessment coordinator.

Writing Portfolios. Each portfolio entry is to be evaluated by certified personnel trained to apply the same set of standards in the same manner from student-to-student, from school-to-school, and from year-to-year.

Teachers may provide opportunities for writing that is appropriate for inclusion in the portfolio, may allow ample time for preparation of portfolio entries in the classroom, and may allow student work outside of class. Teachers may ask questions to clarify aspects of the student's work, may indicate where errors occur, and may ask questions about the errors. Teachers may share and discuss with students the portfolio scoring criteria and samples of student work and may discuss a student's best pieces and choices for inclusion in the portfolio. Teachers may assist students in identifying a variety of tasks that address the required types of portfolio entries and may assign peer tutors and others to assist students with portfolio development. Students must write, type, or word process portfolio pieces by themselves, unless otherwise allowed as accommodations.

Teachers and others may not provide any assistance that diminishes personal ownership of the portfolio and may not alter documentation attesting that the portfolio contents were produced by the student. No one shall make direct corrections or revisions of portfolio entries except for the student. No changes shall be made to portfolio contents after the completion date.

Writing Portfolio Scoring. Only certified school personnel who have received current KDE training may provide accountability scores. Scorers should use current scoring materials and apply the scoring standards accurately and consistently. Scoring judgments are made on the basis of the scoring guide, benchmarks, and reference to high-end portfolios to resolve decisions about performance level. The district shall maintain documentation that all scorers of writing portfolios have been appropriately trained. No individual shall instruct or encourage teachers to assign higher or lower scores than warranted, and scoring accuracy should not be compromised by lack of adequate training or inappropriate scoring conditions.

Inclusion of Special Populations. An individual who provides any accommodation to a student with disabilities on any component of the statewide assessment shall be trained in the role and responsibilities and abide by confidentiality laws, the Administration Code, and the conditions as described in the

student's IEP, 504 Plan, or LEP Plan. Any accommodations or modifications shall also be consistent with 703 KAR 5:070.

Alternate Portfolios. A student who meets all the eligibility requirements for the Alternate Portfolio Assessment Program may submit an alternate portfolio. Any intervention from teachers, peers, or others should enhance a student's ownership of the portfolio. Teachers, parents, friends, and peers may assume support roles as listeners, responders, and encouragers. Only certified school personnel who have received current KDE training may provide accountability scores. Scorers should refer to the terms used in the Alternate Portfolio Program Holistic Scoring Guide and score only evidence seen. The district shall maintain documentation showing that scorers have been appropriately trained. No teacher-authored materials shall be included other than the entry cover page. The student may use an accommodation or assistive device only if it is a regular part of that student's instruction. No additions, subtractions, or revisions may be made after the completion deadline.

III. Violations of the Administration Code for Kentucky's Educational Assessment Program

The following steps shall be taken for any alleged state testing violation:

1. An allegation of inappropriate testing practices received at KDE shall be referred to the Bureau of Management Support Services, Division of Management Assistance (DMA).
2. DMA staff shall manage the process for investigating each allegation of inappropriate testing practice.
3. DMA staff shall report all findings for each allegation to the Board of Review consisting of members appointed by the commissioner of education.
4. The Board of Review shall review the findings and make a recommendation to the commissioner.
5. The commissioner shall make a final determination and then notify the school district superintendent and the school board chairperson. If an allegation is determined to be valid and warrants invalidation or change of scores, the commissioner shall direct the deputy commissioner of learning support services to make appropriate adjustments in a school's or district's scores. If it appears a school district employee is guilty of wrongdoing, the local district superintendent shall report within 45 days in writing to the commissioner whether disciplinary action was taken or

considered necessary and shall comply with the reporting responsibility pursuant to KRS 161.120. If school or district accountability indices are adjusted as a result of the commissioner's final determination, individual student reports shall not be changed, but changes to school or district accountability indices shall be reflected. Scores used to calculate the affected growth indices shall be adjusted and may be reduced to nonperformance for accountability purposes.

6. After the local district receives the letter from the commissioner of the action to be taken by the department, the school may challenge the action by appealing the next performance judgment it receives, as described in 703 KAR 5:050.

IV. Review of Secure Assessment Components by Local District and Other Certified Staff, Parents, and Persons Not in the Employment of a Kentucky Public School District

While KDE does not require individual student participation in the statewide testing program, KDE shall hold schools accountable for the performance of all students. In the absence of assessment information about the performance of a student, the school shall be assigned a nonperformance (low novice) level for that student.

Local district and other certified staff shall not be permitted routine and systematic access to the assessment. If a district chooses to assist in the review of secure testing materials, the review shall take place in the presence of the local district assessment coordinator. If a district chooses not to permit the review of secure materials under its auspices, KDE may permit review based on the availability of appropriate staff to supervise the review activities.

V. Proper Reporting of Nonacademic Indicators (Attendance, Retention, Dropout, and Transition to Adult Life)

Local districts shall be responsible for submitting this data as accurately as possible and are responsible for informing KDE of any known errors in the data reported. Reporting incorrect data shall be considered a violation of the Administration Code and shall be treated as described in Section III.

Assistance to Schools, Scholastic Audits

703 KAR 5:120 establishes standards for assistance to schools and for conducting scholastic audits.

Section 1. A Level 1 school shall conduct a scholastic review and self-study facilitated by the district's professional development coordinator with assistance provided by KDE staff. A Level 1 school may be eligible to receive Commonwealth school improvement funds.

Section 2. A Level 2 school shall receive a scholastic review facilitated and chaired by a designee of the commissioner of education with assistance from the district's central office staff. A Level 2 school may be eligible to receive Commonwealth school improvement funds.

Section 3. A Level 3 school shall receive education assistance from a highly skilled educator under KRS 158.782 and a scholastic audit. A Level 3 school shall be eligible to receive Commonwealth school improvement funds.

Section 4. Evaluation of school personnel in a Level 3 school shall address specific issues including the district's evaluation plan and process for certified staff and the need for additional staff evaluations.

Section 5. If a school is classified as a Level 3 school for two consecutive biennia, a student attending the Level 3 school may transfer to a school with an accountability index above its assistance line. The superintendent shall select the receiving successful school in the home district or make arrangements with a neighboring district. The school district in which the student is enrolled shall retain the Support Education Excellence in Kentucky funding and the student's resident district shall be responsible for all transportation costs incurred as a result of a student transferring.

Section 6. If a school is classified as Level 3, a scholastic audit team may request the commissioner to recommend to a local board of education the removal of a school council member under KRS 160.347.

Section 7. Members of the scholastic audit team shall be selected and trained from a pool of candidates who have submitted an application to KDE. This section of the regulation describes topics that must be included in the training that the team members shall receive, such as developing a comprehensive school improvement plan, building capacity for school leadership, organizing the school to maximize use of resources, developing an effective learning community, conducting professional growth and evaluation of

certified personnel, and assessing and advising compliance with statutes and regulations.

The scholastic audit team shall consist of the following six members from which the commissioner shall name the chairperson:

1. A highly skilled educator;
2. An active or retired teacher from another district;
3. An active or retired principal from another district;
4. An active or retired administrator from another district;
5. A parent or legal guardian; and
6. An active or retired university faculty member.

Prior to the scholastic audit, the school principal shall prepare a school portfolio for use in creating a profile of the strengths and limitations of the school's instructional and organizational effectiveness. The scholastic audit team shall evaluate a school's learning environment, efficiency, and student academic performance by using "Standards and Indicators for School Improvement." The audit team shall make recommendations for assistance, share a draft report with the school faculty and council members prior to departure, and submit a final exit report within three weeks following the site visit.

The school principal and other school council members shall notify parents and interested community members of the findings and recommendations of the audit team. The audit findings shall be presented and discussed on the agenda of the next school council meeting and at a local board of education meeting.

School improvement plans shall be based on recommendations from the audit team's exit report and research-based standards and indicators of quality. Amending a school plan shall be a local decision of which the district is notified.

Section 8. A principal of a school classified as Level 1, 2, or 3 shall participate in at least 12 hours of professional development activities within 12 months of the classification of the school.

Section 9. KDE shall conduct scholastic audits in a random sample of schools.

Section 10. The "Standards and Indicators for School Improvement" is incorporated by reference. It establishes nine standards with multiple indicators relevant to the attainment of each standard. The nine standards are listed below.

1. Academic Performance – Curriculum
2. Academic Performance – Classroom Evaluation/Assessment
3. Academic Performance – Instruction
4. Learning Environment – School Culture
5. Learning Environment – Student, Family and Community Support
6. Learning Environment – Professional Growth, Development & Evaluation
7. Efficiency – Leadership
8. Efficiency – Organizational Structure and Resources
9. Efficiency – Comprehensive and Effective Planning

District Rewards

703 KAR 5:130 establishes eligibility for district rewards and establishes procedures for determining assistance and consequences for local school districts having schools in need of assistance as defined in 703 KAR 5:020.

Section 1. If a district is selected by the U.S. Department of Education to participate in the state National Assessment of Educational Progress in reading, math, and science at grades 4 and 8, the district shall participate fully.

Section 2. Dropout data generated at an A2-A6 school shall be attributed to the school district in which the A2-A6 school is located unless the district can identify the A1 school that the student would have attended. In that case, the dropout data shall be assigned to the A1 school.

Section 3. A local district in which all schools are classified as progressing or meets goal under 703 KAR 5:020 and meet the dropout criteria to earn rewards in 703 KAR 5:020 shall be declared an exemplary growth district and shall receive rewards determined by the Kentucky Board of Education.

Section 4. A district meeting adequate yearly progress in both reading and math for two consecutive years shall receive a reward or recognition as determined by KDE.

Section 5. A local school district shall be held accountable for providing its schools appropriate instructional leadership and instructional support. A local school district containing a school classified as Level 3 shall modify its district consolidated plan by including a specific support plan designed to assist each Level 3 school in improving its academic achievement. If a school is classified as Level 3 for two or more consecutive accountability

cycles, the school district shall be subject to a district audit conducted by a district evaluation team.

Section 6. A local school district shall address particular areas in its school support plan, including instructional staff access to curriculum-related materials and training, professional development planning process, structure for instructional improvement, financial services and support, adequate maintenance of facilities, and an effective certified evaluation program.

Section 7. The district evaluation team shall submit a report, including its recommendations, to the commissioner of education, the district superintendent, and the local board of education within two weeks of its review. The report shall be presented by a member of the district evaluation team at a local board of education meeting with opportunity for public comment.

Section 8. For the purpose of determining whether a district has met the annual measurable objectives in reading or math under NCLB, KDE shall establish a single starting point for each content area at each accountability level using 2001-2002 data.

To determine adequate yearly progress, a district shall be held accountable based on an aggregated average of students enrolled in the district for a full year and producing accountability statistics.

A district has made adequate yearly progress if

1. the district and subpopulations of sufficient size met the annual measurable objectives in reading and math or met the "safe harbor" conditions;
2. it showed progress or met criteria on the elementary and middle school accountability indices;
3. it made progress or met the graduation rate annual goal; and
4. it had a participation rate of at least 95 percent of students and subpopulations of sufficient size.

A district shall be identified as an NCLB improvement district if it fails to make adequate yearly progress in the same content area for two years.

A district has met the annual measurable objective if the objective falls within the 99 percent confidence interval placed around the district's percent of students proficient and above.

A student enrolled in a Kentucky public school on the first day of the testing window at each accountability level shall be included in the calculation of the district's participation rates. A student enrolled for a full academic year shall be included in the district calculations of the percent of students performing at the proficient level or above under NCLB. Annual measurable objectives for reading and math for a district are established in 703 KAR 5:020, §10(11).

A district identified as an NCLB improvement district shall develop or revise its improvement plan within three months. The district shall implement the plan no later than the beginning of the following school year. If the district fails to make adequate yearly progress by the end of the second full school year after identification, the district shall be subject to corrective action. If a district makes adequate yearly progress in both reading and math for two consecutive years after identification, it shall no longer be an improvement district and shall not be subject to federal consequences.

School and District Report Cards

703 KAR 5:140 establishes the standards for a school and district report card.

The following terms are defined:
- Average student/teacher ratio;
- Average years of experience;
- Base year;
- Certified teacher;
- Content-focused professional development;
- District report card (base);
- District report card (expanded);
- School;
- School report card (base);
- School report card (expanded);
- School safety data;
- Spending per student – district;
- Spending per student – school;
- Spending per student – state; and
- Total enrollment.

A school report card (base) shall be sent to the parents or guardians of each student in a school by U.S. mail, unless a waiver is granted by KDE allowing a school to use a method of distribution that is equally effective. A school report card shall provide specific required information: relevant contact information, the school's total enrollment, results of all components of CATS, teacher

qualification information, school safety data, student resource data, parental involvement information, and a narrative describing actions being taken to address issues regarding the equity of the delivery of educational services to all students.

A school report card (expanded) shall be available for viewing on request in the office of the school. It shall include the specific information required by this regulation: data disaggregation pages, documentation of plans for assisting students at risk of failure, number of students participating in special education programs, number of students and percentage of student population receiving accommodations, executive summary from the school consolidated plan, a listing of average class size, a school technology report, and the number of students enrolled for a fifth year in the primary program.

Upon the implementation of a statewide student database, the expanded school report card shall include additional information regarding the advanced placement subjects offered by a high school, an indication of whether a Commonwealth diploma is offered, and the total number of students enrolled in the gifted and talented program.

A district report card (base) shall include a district-level summary of all school data required on the school report card (base) and shall be the aggregation of the school report cards by grade level.

A district report card (expanded) shall be available for viewing in the district central office. It shall include information required by the regulation: data disaggregation pages, documentation of plans to assist students at risk of academic failure, number of students participating in special education, number of students receiving instructional accommodations, executive summary from the district consolidated plans, average class sizes, technology report, number of students enrolled for a fifth year in the primary program, and copies of all base school report cards.

Upon implementation of a statewide student database, the expanded district report card shall include additional information as described above in the expanded school report card.

The school council shall review and approve the school report card (base and expanded) before it is printed. A school report card (base) shall be printed and sent to all parents no later than 77 calendar days from the release of data to the schools. A school

report card (expanded) shall be available in the schools by the same date that the school report card (base) is sent.

KDE shall make district and school data available electronically no later than November 1 of each year. A district has 21 days to report inaccuracies and request the data be changed. KDE then has 21 days to correct the data or determine that no change will be made.

A district report card (base) shall be published in the newspaper with the largest circulation in the county no later than the second Sunday in February. By the date that the district report card (base) is publish, a district report card (expanded) must be available in the district central office.

KDE shall conduct an audit of school and district report cards for compliance with this regulation. If a school district fails to meet the timelines for publication, it shall communicate by letter to KDE identifying the component that has not been published and indicating when it was or will be communicated to the appropriate public. If a school district intentionally publishes incorrect information, alters data, or refuses to produce a required component of a school report card, the matter shall be referred to the Division of Management Assistance.

The "Calculation Procedures for Data Included in the School Report Card," August 2000 is incorporated by reference. This document is a compilation of all calculation procedures used in the school report card components.

Appendix B

Survey Results

This appendix summarizes the results of the closed-ended questions from the surveys of teachers, principals (163-183), superintendents (pages 163-183), high school students (183-190), parents (190-193), and school board members (194-198). The results from the open-ended questions included here are labeled as such. Each survey was conducted in 2004 by AEL, Inc. of Charleston, West Virginia. Because the questionnaires were similar for teachers, principals, and superintendents, the results from those surveys are presented using common tables whenever feasible.

Surveys of Teachers, Principals, and Superintendents

Note: Survey questions for teachers are indicated by "T," principals by "P," and superintendents by "S." Question numbers are in parentheses (Q#).

Grade(s) taught (T=Q1)

	Number of Respondents	% of Total
Kindergarten	51	15.2%
1	62	18.5%
2	65	19.3%
3	72	21.4%
4	85	25.3%
5	69	20.5%
6	58	17.3%
7	54	16.1%
8	57	17.0%
9	82	24.4%
10	87	25.9%
11	89	26.5%
12	84	25.0%
Other	14	4.2%
Number of respondents	336	*

*Total adds to more than 100% because multiple answers are possible; 336 teachers answered the question.

Recoded as:		
K-5	156	48.4%
6-8	97	30.1%
9-12	106	32.9%
Number of respondents	322	*

*Total adds to more than 100% because multiple answers are possible; 322 teachers answered the question.

What subject(s) do you teach? (T=Q2)
[Open question, coded by staff]

Four or more subjects	68	21.1%
Arts and Humanities	24	7.5%
Foreign Language	6	1.9%
Language Arts	101	31.4%
LEP/ESL	14	4.3%
Math	69	21.4%
Physical Education/Health	10	3.1%
Practical Living	4	1.2%
Science	42	13.0%
Social Studies	40	12.4%
Special Education	27	8.4%
Other	20	6.2%
Number of respondents	322	*

*Total adds to more than 100% because multiple answers are possible; 322 teachers answered the question.

What is your overall opinion of CATS testing? (T=Q3, P&S=Q1)

	Teacher		Principal		Superintendent	
Very dissatisfied	59	17.7%	25	10.9%	7	6.5%
Somewhat dissatisfied	113	33.8%	52	22.6%	17	15.9%
Neutral	57	17.1%	18	7.8%	12	11.2%
Somewhat satisfied	97	29.0%	125	54.3%	64	59.8%
Very satisfied	8	2.4%	10	4.3%	7	6.5%
Totals	334	100.0%	230	100.0%	107	100.0%

Overall, how does CATS testing affect what students learn? (T=Q4, P&S=Q2)

	Teacher		Principal		Superintendent	
Very negatively	15	4.6%	3	1.3%	3	2.8%
Somewhat negatively	82	25.0%	31	13.6%	10	9.3%
Neutral or no effect	67	20.4%	20	8.8%	26	24.3%
Somewhat positively	140	42.7%	137	60.1%	58	54.2%
Very positively	24	7.3%	37	16.2%	10	9.3%
Totals	328	100.0%	228	100.0%	107	100.0%

Assuming they are taught the same content of the tested subjects as other students, are any of the following types of students disadvantaged by CATS testing? (T=Q5, P&S=Q3)

Below-average readers

	Teacher		Principal		Superintendent	
None or almost none	13	4.1%	12	5.3%	18	17.0%
A few	70	21.9%	66	29.1%	46	43.4%
Many	236	74.0%	149	65.6%	42	39.6%
Totals	319	100.0%	227	100.0%	106	100.0%

Below-average writers

	Teacher		Principal		Superintendent	
None or almost none	11	3.4%	8	3.5%	16	15.2%
A few	59	18.5%	56	24.8%	40	38.1%
Many	249	78.1%	162	71.7%	49	46.7%
Totals	319	100.0%	226	100.0%	105	100.0%

Students with limited English proficiency

	Teacher		Principal		Superintendent	
None or almost none	20	6.7%	11	5.3%	10	9.9%
A few	52	17.3%	37	17.9%	22	21.8%
Many	228	76.0%	159	76.8%	69	68.3%
Totals	300	100.0%	207	100.0%	101	100.0%

Special Education students

	Teacher		Principal		Superintendent	
None or almost none	30	9.8%	28	12.4%	19	17.9%
A few	78	25.4%	67	29.8%	29	27.4%
Many	199	64.8%	130	57.8%	58	54.7%
Totals	307	100.0%	225	100.0%	106	100.0%

Overall, how does CATS testing affect the curriculum you teach? (T=Q6)
Overall, how does CATS testing affect the curriculum your staff teaches? (P&S=Q4)

	Teacher		Principal		Superintendent	
Very negatively	17	5.2%	4	1.8%	3	2.8%
Somewhat negatively	105	32.1%	35	15.4%	10	9.3%
Neutral or no effect	57	17.4%	12	5.3%	8	7.5%
Somewhat positively	102	31.2%	118	51.8%	54	50.5%
Very positively	46	14.1%	59	25.9%	32	29.9%
Totals	327	100.0%	228	100.0%	107	100.0%

Overall, how does CATS testing affect instruction in your classroom? (T=Q7)
Overall, how does CATS testing affect instruction in your school/school system? (P&S=Q5)

	Teacher		Principal		Superintendent	
Very negatively	17	5.2%	4	1.8%	5	4.7%
Somewhat negatively	95	29.1%	32	14.2%	11	10.3%
Neutral or no effect	67	20.6%	15	6.6%	4	3.7%
Somewhat positively	109	33.4%	122	54.0%	60	56.1%
Very positively	38	11.7%	53	23.5%	27	25.2%
Totals	326	100.0%	226	100.0%	107	100.0%

Please indicate whether you agree or disagree with these statements about CATS testing. (T=Q8, P&S=6)

Getting ready for or taking the CATS tests takes too much time away from class time.

	Teacher		Principal		Superintendent	
Agree	236	73.1%	126	55.8%	58	54.7%
Disagree	87	26.9%	100	44.2%	48	45.3%
Totals	323	100.0%	226	100.0%	106	100.0%

The CATS testing system assures that goals are the same for everyone.

	Teacher		Principal		Superintendent	
Agree	196	60.5%	154	68.1%	83	77.6%
Disagree	128	39.5%	72	31.9%	24	22.4%
Totals	324	100.0%	226	100.0%	107	100.0%

CATS testing is too stressful and reduces enjoyment of teaching and learning.

	Teacher		Principal		Superintendent	
Agree	273	83.7%	171	75.3%	58	54.7%
Disagree	53	16.3%	56	24.7%	48	45.3%
Totals	326	100.0%	227	100.0%	106	100.0%

CATS testing provides useful information about how well teachers are doing.

	Teacher		Principal		Superintendent	
Agree	57	17.4%	102	45.1%	62	57.9%
Disagree	270	82.6%	124	54.9%	45	42.1%
Totals	327	100.0%	226	100.0%	107	100.0%

Too often, CATS test content is inappropriate for the students taking a test.

	Teacher		Principal		Superintendent	
Agree	202	63.1%	100	44.6%	32	30.2%
Disagree	118	36.9%	124	55.4%	74	69.8%
Totals	320	100.0%	224	100.0%	106	100.0%

CATS testing provides needed focus and organization.

	Teacher		Principal		Superintendent	
Agree	181	57.6%	160	70.8%	83	78.3%
Disagree	133	42.4%	66	29.2%	23	21.7%
Totals	314	100.0%	226	100.0%	106	100.0%

Too often, the format of CATS test questions is inappropriate for the students taking the test.

	Teacher		Principal		Superintendent	
Agree	197	63.3%	107	47.6%	32	30.2%
Disagree	114	36.7%	118	52.4%	74	69.8%
Totals	311	100.0%	225	100.0%	106	100.0%

CATS testing helps align the curriculum.

	Teacher		Principal		Superintendent	
Agree	214	67.1%	195	85.5%	95	88.8%
Disagree	105	32.9%	33	14.5%	12	11.2%
Totals	319	100.0%	228	100.0%	107	100.0%

Teaching what is to be covered on the CATS tests is too limiting.

	Teacher		Principal		Superintendent	
Agree	230	72.8%	118	52.7%	45	42.1%
Disagree	86	27.2%	106	47.3%	62	57.9%
Totals	316	100.0%	224	100.0%	107	100.0%

CATS testing provides useful information about how well students are doing.

	Teacher		Principal		Superintendent	
Agree	126	39.6%	124	55.4%	73	68.2%
Disagree	192	60.4%	100	44.6%	34	31.8%
Totals	318	100.0%	224	100.0%	107	100.0%

Teachers and students are forced to cover material too quickly in order to prepare for CATS tests.

	Teacher		Principal		Superintendent	
Agree	286	88.8%	168	74.3%	54	50.5%
Disagree	36	11.2%	58	25.7%	53	49.5%
Totals	322	100.0%	226	100.0%	107	100.0%

CATS testing provides useful information about how well schools are doing.

	Teacher		Principal		Superintendent	
Agree	96	29.9%	125	55.1%	74	69.2%
Disagree	225	70.1%	102	44.9%	33	30.8%
Totals	321	100.0%	227	100.0%	107	100.0%

What percentage of your work time during the entire school year do you spend preparing for CATS testing, including test-taking techniques and practice tests? Do not include time you spend teaching core content or time spent on writing portfolios. (T=Q9)

[Open question, coded by staff]

% of Time		
0-10	96	34.3%
11-20	49	17.5%
21-30	49	17.5%
31-40	19	6.8%
41-50	23	8.2%
51-60	5	1.8%
61-70	10	3.6%
71-80	13	4.6%
81-90	6	2.1%
More than 90	10	3.6%
Total	280	100.0%

Average=29, Median=20

For each CATS test and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' knowledge of the core content. (T=Q10, P&S=7)

Special Education students: Reading

	Teacher		Principal		Superintendent	
Strongly disagree	89	31.9%	50	22.2%	15	14.6%
Disagree	91	32.6%	92	40.9%	36	35.0%
Agree	56	20.1%	62	27.6%	34	33.0%
Strongly agree	5	1.8%	7	3.1%	4	3.9%
Not sure	38	13.6%	14	6.2%	14	13.6%
Totals	279	100.0%	225	100.0%	103	100.0%

Special Education students: Math

	Teacher		Principal		Superintendent	
Strongly disagree	72	25.8%	41	18.3%	16	15.5%
Disagree	97	34.8%	85	37.9%	30	29.1%
Agree	65	23.3%	75	33.5%	37	35.9%
Strongly agree	4	1.4%	7	3.1%	4	3.9%
Not sure	41	14.7%	16	7.1%	16	15.5%
Totals	279	100.0%	224	100.0%	103	100.0%

Special Education students: Science

	Teacher		Principal		Superintendent	
Strongly disagree	71	25.5%	40	17.9%	14	13.6%
Disagree	103	37.1%	87	38.8%	34	33.0%
Agree	60	21.6%	74	33.0%	34	33.0%
Strongly agree	4	1.4%	7	3.1%	5	4.9%
Not sure	40	14.4%	16	7.1%	16	15.5%
Totals	278	100.0%	224	100.0%	103	100.0%

Special Education students: Social Studies

	Teacher		Principal		Superintendent	
Strongly disagree	72	26.2%	39	17.6%	14	13.6%
Disagree	97	35.3%	86	38.7%	36	35.0%
Agree	62	22.5%	74	33.3%	34	33.0%
Strongly agree	3	1.1%	7	3.2%	4	3.9%
Not sure	41	14.9%	16	7.2%	15	14.6%
Totals	275	100.0%	222	100.0%	103	100.0%

Special Education students: Arts and Humanities

	Teacher		Principal		Superintendent	
Strongly disagree	73	26.3%	47	21.0%	16	15.5%
Disagree	100	36.0%	90	40.2%	37	35.9%
Agree	51	18.3%	64	28.6%	29	28.2%
Strongly agree	4	1.4%	6	2.7%	4	3.9%
Not sure	50	18.0%	17	7.6%	17	16.5%
Totals	278	100.0%	224	100.0%	103	100.0%

Special Education students: Practical Living/Vocational

	Teacher		Principal		Superintendent	
Strongly disagree	63	23.2%	45	20.0%	14	13.6%
Disagree	92	33.8%	86	38.2%	35	34.0%
Agree	61	22.4%	74	32.9%	33	32.0%
Strongly agree	4	1.5%	6	2.7%	4	3.9%
Not sure	52	19.1%	14	6.2%	17	16.5%
Totals	272	100.0%	225	100.0%	103	100.0%

Special Education students: Alternate Portfolio

	Teacher		Principal		Superintendent	
Strongly disagree	*		42	19.5%	15	14.7%
Disagree			53	24.7%	27	26.5%
Agree			65	30.2%	28	27.5%
Strongly agree			27	12.6%	10	9.8%
Not sure			28	13.0%	22	21.6%
Totals			215	100.0%	102	100.0%

*This item was omitted from the teacher questionnaire inadvertently.

For each writing assessment and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' writing. (T=Q11, P&S=Q8)

Special Education students: Writing portfolios

	Teacher		Principal		Superintendent	
Strongly disagree	97	33.7%	63	27.8%	18	17.5%
Disagree	94	32.6%	99	43.6%	38	36.9%
Agree	58	20.1%	51	22.5%	37	35.9%
Strongly agree	9	3.1%	7	3.1%	3	2.9%
Not sure	30	10.4%	7	3.1%	7	6.8%
Totals	288	100.0%	227	100.0%	103	100.0%

Special Education students: On-demand writing

	Teacher		Principal		Superintendent	
Strongly disagree	81	27.8%	47	20.7%	17	16.5%
Disagree	86	29.6%	111	48.9%	37	35.9%
Agree	84	28.9%	57	25.1%	38	36.9%
Strongly agree	10	3.4%	5	2.2%	3	2.9%
Not sure	30	10.3%	7	3.1%	8	7.8%
Totals	291	100.0%	227	100.0%	103	100.0%

Special Education students: Alternate Assessment

	Teacher		Principal		Superintendent	
Strongly disagree	50	18.2%	47	21.3%	18	17.5%
Disagree	48	17.5%	53	24.0%	30	29.1%
Agree	54	19.6%	61	27.6%	28	27.2%
Strongly agree	10	3.6%	21	9.5%	11	10.7%
Not sure	113	41.1%	39	17.6%	16	15.5%
Totals	275	100.0%	221	100.0%	103	100.0%

For each CATS test and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' knowledge of the core content. (T=Q10, P&S=7)

Students with limited English proficiency: Reading

	Teacher		Principal		Superintendent	
Strongly disagree	60	24.4%	56	27.5%	17	17.0%
Disagree	75	30.5%	69	33.8%	36	36.0%
Agree	14	5.7%	24	11.8%	16	16.0%
Strongly agree	4	1.6%	0	0.0%	2	2.0%
Not sure	93	37.8%	55	27.0%	29	29.0%
Totals	246	100.0%	204	100.0%	100	100.0%

Students With Limited English Proficiency: Math

	Teacher		Principal		Superintendent	
Strongly disagree	42	16.9%	59	28.9%	16	16.0%
Disagree	71	28.6%	68	33.3%	34	34.0%
Agree	37	14.9%	22	10.8%	17	17.0%
Strongly agree	5	2.0%	0	0.0%	3	3.0%
Not sure	93	37.5%	55	27.0%	30	30.0%
Totals	248	100.0%	204	100.0%	100	100.0%

Students With Limited English Proficiency: Science

	Teacher		Principal		Superintendent	
Strongly disagree	55	22.1%	39	19.3%	13	13.0%
Disagree	76	30.5%	47	23.3%	25	25.0%
Agree	20	8.0%	32	15.8%	18	18.0%
Strongly agree	3	1.2%	5	2.5%	4	4.0%
Not sure	95	38.2%	79	39.1%	40	40.0%
Totals	249	100.0%	202	100.0%	100	100.0%

Students With Limited English Proficiency: Social Studies

	Teacher		Principal		Superintendent	
Strongly disagree	56	23.0%	47	23.2%	16	16.0%
Disagree	75	30.7%	74	36.5%	37	37.0%
Agree	17	7.0%	24	11.8%	12	12.0%
Strongly agree	3	1.2%	1	0.5%	2	2.0%
Not sure	93	38.1%	57	28.1%	33	33.0%
Totals	244	100.0%	203	100.0%	100	100.0%

Students With Limited English Proficiency: Arts and Humanities

	Teacher		Principal		Superintendent	
Strongly disagree	55	22.2%	48	23.8%	17	17.0%
Disagree	79	31.9%	72	35.6%	36	36.0%
Agree	16	6.5%	25	12.4%	12	12.0%
Strongly agree	4	1.6%	1	0.5%	2	2.0%
Not sure	94	37.9%	56	27.7%	33	33.0%
Totals	248	100.0%	202	100.0%	100	100.0%

Students With Limited English Proficiency: Practical Living/Vocational

	Teacher		Principal		Superintendent	
Strongly disagree	54	22.1%	47	23.4%	15	14.9%
Disagree	75	30.7%	71	35.3%	36	35.6%
Agree	17	7.0%	26	12.9%	14	13.9%
Strongly agree	4	1.6%	1	0.5%	1	1.0%
Not sure	94	38.5%	56	27.9%	35	34.7%
Totals	244	100.0%	201	100.0%	101	100.0%

For each writing assessment and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' writing. (T=Q11, P&S=8)

Students with limited English proficiency: Writing portfolios

	Teacher		Principal		Superintendent	
Strongly disagree	60	24.0%	43	21.9%	18	18.8%
Disagree	68	27.2%	82	41.8%	42	43.8%
Agree	25	10.0%	42	21.4%	20	20.8%
Strongly agree	4	1.6%	10	5.1%	1	1.0%
Not sure	93	37.2%	19	9.7%	15	15.6%
Totals	250	100.0%	196	100.0%	96	100.0%

Students with limited English proficiency: On-demand writing

	Teacher		Principal		Superintendent	
Strongly disagree	51	20.3%	32	16.4%	15	15.5%
Disagree	71	28.3%	78	40.0%	37	38.1%
Agree	29	11.6%	56	28.7%	30	30.9%
Strongly agree	5	2.0%	9	4.6%	3	3.1%
Not sure	95	37.8%	20	10.3%	12	12.4%
Totals	251	100.0%	195	100.0%	97	100.0%

FOR ALL OTHER STUDENTS: For each CATS test and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' knowledge of the core content. (T=Q12, P&S=Q9)

Elementary School students: Reading

	Teacher		Principal		Superintendent	
Strongly disagree	19	7.7%	7	3.3%	3	2.9%
Disagree	52	21.1%	32	15.0%	12	11.4%
Agree	98	39.8%	132	62.0%	68	64.8%
Strongly agree	8	3.3%	16	7.5%	16	15.2%
Not sure	69	28.0%	26	12.2%	6	5.7%
Totals	246	100.0%	213	100.0%	105	100.0%

Middle School students: Reading

	Teacher		Principal		Superintendent	
Strongly disagree	14	7.1%	6	3.8%	4	3.8%
Disagree	33	16.8%	25	16.0%	9	8.7%
Agree	58	29.4%	66	42.3%	65	62.5%
Strongly agree	6	3.0%	8	5.1%	14	13.5%
Not sure	86	43.7%	51	32.7%	12	11.5%
Totals	197	100.0%	156	100.0%	104	100.0%

High School students: Reading

	Teacher		Principal		Superintendent	
Strongly disagree	13	6.5%	9	6.0%	6	5.8%
Disagree	32	15.9%	20	13.3%	11	10.7%
Agree	56	27.9%	62	41.3%	64	62.1%
Strongly agree	6	3.0%	9	6.0%	14	13.6%
Not sure	94	46.8%	50	33.3%	8	7.8%
Totals	201	100.0%	150	100.0%	103	100.0%

Elementary School students: Math

	Teacher		Principal		Superintendent	
Strongly disagree	20	8.3%	10	4.7%	3	2.9%
Disagree	50	20.7%	37	17.5%	10	9.5%
Agree	92	38.2%	118	55.7%	71	67.6%
Strongly agree	7	2.9%	18	8.5%	15	14.3%
Not sure	72	29.9%	29	13.7%	6	5.7%
Totals	241	100.0%	212	100.0%	105	100.0%

Middle School students: Math

	Teacher		Principal		Superintendent	
Strongly disagree	15	7.6%	8	5.2%	4	3.8%
Disagree	31	15.7%	24	15.5%	12	11.5%
Agree	58	29.4%	66	42.6%	64	61.5%
Strongly agree	4	2.0%	6	3.9%	13	12.5%
Not sure	89	45.2%	51	32.9%	11	10.6%
Totals	197	100.0%	155	100.0%	104	100.0%

High School students: Math

	Teacher		Principal		Superintendent	
Strongly disagree	12	6.0%	9	6.0%	5	4.9%
Disagree	30	15.0%	22	14.7%	16	15.5%
Agree	56	28.0%	64	42.7%	60	58.3%
Strongly agree	6	3.0%	6	4.0%	14	13.6%
Not sure	96	48.0%	49	32.7%	8	7.8%
Totals	200	100.0%	150	100.0%	103	100.0%

Elementary School students: Science

	Teacher		Principal		Superintendent	
Strongly disagree	22	9.0%	13	6.1%	3	2.9%
Disagree	54	22.1%	36	16.9%	12	11.4%
Agree	87	35.7%	120	56.3%	70	66.7%
Strongly agree	7	2.9%	16	7.5%	13	12.4%
Not sure	74	30.3%	28	13.1%	7	6.7%
Totals	244	100.0%	213	100.0%	105	100.0%

Middle School students: Science

	Teacher		Principal		Superintendent	
Strongly disagree	18	9.0%	9	5.8%	4	3.8%
Disagree	32	16.1%	25	16.1%	12	11.5%
Agree	57	28.6%	62	40.0%	65	62.5%
Strongly agree	4	2.0%	6	3.9%	12	11.5%
Not sure	88	44.2%	53	34.2%	11	10.6%
Totals	199	100.0%	155	100.0%	104	100.0%

High School students: Science

	Teacher		Principal		Superintendent	
Strongly disagree	12	6.0%	9	6.1%	5	4.9%
Disagree	29	14.4%	22	14.9%	13	12.7%
Agree	54	26.9%	62	41.9%	65	63.7%
Strongly agree	6	3.0%	5	3.4%	12	11.8%
Not sure	100	49.8%	50	33.8%	7	6.9%
Totals	201	100.0%	148	100.0%	102	100.0%

Elementary School students: Social Studies

	Teacher		Principal		Superintendent	
Strongly disagree	20	8.4%	13	6.1%	4	3.8%
Disagree	55	23.1%	35	16.4%	9	8.6%
Agree	82	34.5%	121	56.8%	70	66.7%
Strongly agree	7	2.9%	16	7.5%	14	13.3%
Not sure	74	31.1%	28	13.1%	8	7.6%
Totals	238	100.0%	213	100.0%	105	100.0%

Middle School students: Social Studies

	Teacher		Principal		Superintendent	
Strongly disagree	17	8.6%	9	5.9%	4	3.8%
Disagree	33	16.7%	23	15.0%	11	10.6%
Agree	57	28.8%	63	41.2%	67	64.4%
Strongly agree	4	2.0%	6	3.9%	12	11.5%
Not sure	87	43.9%	52	34.0%	10	9.6%
Totals	198	100.0%	153	100.0%	104	100.0%

High School students: Social Studies

	Teacher		Principal		Superintendent	
Strongly disagree	12	6.0%	8	5.3%	6	6.0%
Disagree	30	15.0%	23	15.3%	12	12.0%
Agree	55	27.5%	63	42.0%	63	63.0%
Strongly agree	5	2.5%	6	4.0%	12	12.0%
Not sure	98	49.0%	50	33.3%	7	7.0%
Totals	200	100.0%	150	100.0%	100	100.0%

Elementary School students: Arts and Humanities

	Teacher		Principal		Superintendent	
Strongly disagree	24	10.0%	25	11.8%	5	4.8%
Disagree	64	26.7%	52	24.5%	20	19.2%
Agree	65	27.1%	96	45.3%	56	53.8%
Strongly agree	6	2.5%	8	3.8%	11	10.6%
Not sure	81	33.8%	31	14.6%	12	11.5%
Totals	240	100.0%	212	100.0%	104	100.0%

Middle School students: Arts and Humanities

	Teacher		Principal		Superintendent	
Strongly disagree	21	10.6%	11	7.1%	5	4.8%
Disagree	35	17.6%	28	18.2%	19	18.3%
Agree	45	22.6%	55	35.7%	54	51.9%
Strongly agree	4	2.0%	4	2.6%	11	10.6%
Not sure	94	47.2%	56	36.4%	15	14.4%
Totals	199	100.0%	154	100.0%	104	100.0%

High School students: Arts and Humanities

	Teacher		Principal		Superintendent	
Strongly disagree	13	6.6%	10	6.7%	6	5.9%
Disagree	36	18.2%	25	16.8%	21	20.6%
Agree	48	24.2%	56	37.6%	53	52.0%
Strongly agree	4	2.0%	4	2.7%	11	10.8%
Not sure	97	49.0%	54	36.2%	11	10.8%
Totals	198	100.0%	149	100.0%	102	100.0%

Elementary School students: Practical Living/Vocational

	Teacher		Principal		Superintendent	
Strongly disagree	23	9.7%	17	8.0%	5	4.8%
Disagree	62	26.2%	42	19.8%	22	21.2%
Agree	67	28.3%	111	52.4%	55	52.9%
Strongly agree	6	2.5%	12	5.7%	12	11.5%
Not sure	79	33.3%	30	14.2%	10	9.6%
Totals	237	100.0%	212	100.0%	104	100.0%

Middle School students: Practical Living/Vocational

	Teacher		Principal		Superintendent	
Strongly disagree	18	9.1%	10	6.6%	5	5.1%
Disagree	34	17.3%	27	17.8%	21	21.2%
Agree	47	23.9%	56	36.8%	47	47.5%
Strongly agree	4	2.0%	5	3.3%	11	11.1%
Not sure	94	47.7%	54	35.5%	15	15.2%
Totals	197	100.0%	152	100.0%	99	100.0%

High School students: Practical Living/Vocational

	Teacher		Principal		Superintendent	
Strongly disagree	16	8.1%	10	6.8%	7	6.8%
Disagree	32	16.2%	26	17.7%	21	20.4%
Agree	48	24.2%	56	38.1%	53	51.5%
Strongly agree	6	3.0%	5	3.4%	12	11.7%
Not sure	96	48.5%	50	34.0%	10	9.7%
Totals	198	100.0%	147	100.0%	103	100.0%

FOR ALL OTHER STUDENTS: For each writing assessment and type of student with which you are familiar, please indicate whether you agree or disagree that it is a valid measure of students' writing. (T=Q13, P&S=Q10)

Elementary School students: Writing portfolios

	Teacher		Principal		Superintendent	
Strongly disagree	65	27.1%	40	20.3%	12	12.1%
Disagree	75	31.3%	69	35.0%	29	29.3%
Agree	38	15.8%	48	24.4%	37	37.4%
Strongly agree	3	1.3%	11	5.6%	8	8.1%
Not sure	59	24.6%	29	14.7%	13	13.1%
Totals	240	100.0%	197	100.0%	99	100.0%

Middle School students: Writing portfolios

	Teacher		Principal		Superintendent	
Strongly disagree	33	16.3%	19	12.7%	12	11.8%
Disagree	54	26.6%	45	30.0%	28	27.5%
Agree	38	18.7%	36	24.0%	41	40.2%
Strongly agree	4	2.0%	3	2.0%	7	6.9%
Not sure	74	36.5%	47	31.3%	14	13.7%
Totals	203	100.0%	150	100.0%	102	100.0%

High School students: Writing portfolios

	Teacher		Principal		Superintendent	
Strongly disagree	31	15.0%	14	9.7%	13	12.4%
Disagree	42	20.3%	44	30.3%	31	29.5%
Agree	50	24.2%	38	26.2%	42	40.0%
Strongly agree	8	3.9%	5	3.4%	7	6.7%
Not sure	76	36.7%	44	30.3%	12	11.4%
Totals	207	100.0%	145	100.0%	105	100.0%

Elementary School students: On-demand writing

	Teacher		Principal		Superintendent	
Strongly disagree	36	14.9%	18	9.2%	6	6.0%
Disagree	47	19.4%	54	27.6%	16	16.0%
Agree	92	38.0%	88	44.9%	56	56.0%
Strongly agree	6	2.5%	9	4.6%	11	11.0%
Not sure	61	25.2%	27	13.8%	11	11.0%
Totals	242	100.0%	196	100.0%	100	100.0%

Middle School students: On-demand writing

	Teacher		Principal		Superintendent	
Strongly disagree	16	7.9%	8	5.3%	7	6.9%
Disagree	35	17.2%	26	17.3%	12	11.9%
Agree	70	34.5%	68	45.3%	59	58.4%
Strongly agree	8	3.9%	3	2.0%	11	10.9%
Not sure	74	36.5%	45	30.0%	12	11.9%
Totals	203	100.0%	150	100.0%	101	100.0%

High School students: On-demand writing

	Teacher		Principal		Superintendent	
Strongly disagree	17	8.1%	8	5.6%	9	8.5%
Disagree	31	14.8%	24	16.7%	17	16.0%
Agree	71	34.0%	64	44.4%	57	53.8%
Strongly agree	10	4.8%	4	2.8%	11	10.4%
Not sure	80	38.3%	44	30.6%	12	11.3%
Totals	209	100.0%	144	100.0%	106	100.0%

For each of the following, indicate whether the item's weight in the accountability index for schools like yours is about right, whether its weight should be increased or decreased, or whether the item should be eliminated from the school accountability index. (T=Q14, P&S=Q11)

Reading

	Teacher		Principal		Superintendent	
Eliminate	5	1.7%	1	0.5%	0	0.0%
Decrease	28	9.3%	5	2.3%	3	2.9%
About right	182	60.3%	177	80.8%	74	71.8%
Increase	31	10.3%	28	12.8%	23	22.3%
Not sure	56	18.5%	8	3.7%	3	2.9%
Totals	302	100.0%	219	100.0%	103	100.0%

Math

	Teacher		Principal		Superintendent	
Eliminate	5	1.7%	1	0.5%	0	0.0%
Decrease	24	7.9%	5	2.3%	2	1.9%
About right	186	61.6%	177	81.6%	78	75.0%
Increase	30	9.9%	25	11.5%	21	20.2%
Not sure	57	18.9%	9	4.1%	3	2.9%
Totals	302	100.0%	217	100.0%	104	100.0%

Science

	Teacher		Principal		Superintendent	
Eliminate	7	2.3%	3	1.4%	0	0.0%
Decrease	31	10.3%	19	8.6%	6	5.8%
About right	185	61.5%	181	82.3%	84	80.8%
Increase	21	7.0%	9	4.1%	11	10.6%
Not sure	57	18.9%	8	3.6%	3	2.9%
Totals	301	100.0%	220	100.0%	104	100.0%

Social Studies

	Teacher		Principal		Superintendent	
Eliminate	7	2.3%	3	1.4%	0	0.0%
Decrease	33	11.0%	17	7.8%	7	6.9%
About right	185	61.7%	181	82.6%	81	79.4%
Increase	17	5.7%	8	3.7%	11	10.8%
Not sure	58	19.3%	10	4.6%	3	2.9%
Totals	300	100.0%	219	100.0%	102	100.0%

Arts and Humanities

	Teacher		Principal		Superintendent	
Eliminate	34	11.3%	36	16.4%	10	9.7%
Decrease	73	24.3%	52	23.7%	29	28.2%
About right	117	38.9%	114	52.1%	56	54.4%
Increase	20	6.6%	7	3.2%	3	2.9%
Not sure	57	18.9%	10	4.6%	5	4.9%
Totals	301	100.0%	219	100.0%	103	100.0%

Practical Living/Vocational

	Teacher		Principal		Superintendent	
Eliminate	25	8.4%	29	13.3%	13	12.5%
Decrease	57	19.3%	29	13.3%	27	26.0%
About right	121	40.9%	137	62.8%	55	52.9%
Increase	36	12.2%	13	6.0%	4	3.8%
Not sure	57	19.3%	10	4.6%	5	4.8%
Totals	296	100.0%	218	100.0%	104	100.0%

Writing portfolios

	Teacher		Principal		Superintendent	
Eliminate	109	36.0%	82	37.1%	34	32.7%
Decrease	79	26.1%	64	29.0%	41	39.4%
About right	61	20.1%	60	27.1%	23	22.1%
Increase	7	2.3%	8	3.6%	0	0.0%
Not sure	47	15.5%	7	3.2%	6	5.8%
Totals	303	100.0%	221	100.0%	104	100.0%

On-demand writing

	Teacher		Principal		Superintendent	
Eliminate	52	17.2%	38	17.2%	13	12.6%
Decrease	80	26.5%	62	28.1%	22	21.4%
About right	93	30.8%	80	36.2%	40	38.8%
Increase	25	8.3%	30	13.6%	22	21.4%
Not sure	52	17.2%	11	5.0%	6	5.8%
Totals	302	100.0%	221	100.0%	103	100.0%

CTBS/5

	Teacher		Principal		Superintendent	
Eliminate	15	5.3%	3	1.4%	0	0.0%
Decrease	22	7.8%	4	1.9%	5	4.9%
About right	114	40.3%	118	55.1%	57	55.3%
Increase	47	16.6%	74	34.6%	36	35.0%
Not sure	85	30.0%	15	7.0%	5	4.9%
Totals	283	100.0%	214	100.0%	103	100.0%

Are the following test formats appropriate for the grade levels indicated? (T=Q15, P&S=Q12)

Multiple choice: Grades 4 and 5

	Teacher		Principal		Superintendent	
Yes	166	93.8%	171	100.0%	96	99.0%
No	11	6.2%	0	0.0%	1	1.0%
Totals	177	100.0%	171	100.0%	97	100.0%

Multiple choice: Grades 7 and 8

	Teacher		Principal		Superintendent	
Yes	123	92.5%	105	100.0%	93	98.9%
No	10	5.6%	0	0.0%	1	1.0%
Totals	133	98.1%	105	100.0%	94	100.0%

Multiple choice: Grades 9 to 12

	Teacher		Principal		Superintendent	
Yes	143	91.1%	97	100.0%	96	97.0%
No	14	8.9%	0	0.0%	3	3.0%
Totals	157	100.0%	97	100.0%	99	100.0%

Open response: Grades 4 and 5

	Teacher		Principal		Superintendent	
Yes	113	63.8%	134	77.9%	80	83.3%
No	64	36.2%	38	22.1%	16	16.7%
Totals	177	100.0%	172	100.0%	96	100.0%

Open response: Grades 7 and 8

	Teacher		Principal		Superintendent	
Yes	108	82.4%	88	85.4%	83	88.3%
No	23	17.6%	15	14.6%	11	11.7%
Totals	131	100.0%	103	100.0%	94	100.0%

Open response: Grades 9 to 12

	Teacher		Principal		Superintendent	
Yes	134	85.9%	89	90.8%	87	88.8%
No	22	14.1%	9	9.2%	11	11.2%
Totals	156	100.0%	98	100.0%	98	100.0%

On-demand writing: Grades 4 and 5

	Teacher		Principal		Superintendent	
Yes	86	50.0%	94	55.3%	69	74.2%
No	86	50.0%	76	44.7%	24	25.8%
Totals	172	100.0%	170	100.0%	93	100.0%

On-demand writing: Grades 7 and 8

	Teacher		Principal		Superintendent	
Yes	91	72.2%	80	77.7%	76	80.9%
No	35	27.8%	23	22.3%	18	19.1%
Totals	126	100.0%	103	100.0%	94	100.0%

On-demand writing: Grades 9 to 12

	Teacher		Principal		Superintendent	
Yes	129	83.2%	83	83.8%	81	81.8%
No	26	16.8%	16	16.2%	18	18.2%
Totals	155	100.0%	99	100.0%	99	100.0%

Writing portfolio: Grades 4 and 5

	Teacher		Principal		Superintendent	
Yes	57	32.9%	62	37.1%	42	43.3%
No	116	67.1%	105	62.9%	55	56.7%
Totals	173	100.0%	167	100.0%	97	100.0%

Writing portfolio: Grades 7 and 8

	Teacher		Principal		Superintendent	
Yes	67	52.8%	58	55.8%	51	54.3%
No	60	47.2%	46	44.2%	43	45.7%
Totals	127	100.0%	104	100.0%	94	100.0%

Writing portfolio: Grades 9 to 12

	Teacher		Principal		Superintendent	
Yes	108	69.7%	67	67.7%	59	59.6%
No	47	30.3%	32	32.3%	40	40.4%
Totals	155	100.0%	99	100.0%	99	100.0%

In your view, what share of students try their best on CATS tests? (T=Q16, P&S=Q13)

	Teacher		Principal		Superintendent	
None or almost none	5	1.5%	1	0.4%	2	1.9%
Some	157	48.2%	51	22.7%	37	34.9%
Most	143	43.9%	112	49.8%	58	54.7%
All or almost all	17	5.2%	61	27.1%	8	7.5%
Not sure	4	1.2%	0	0.0%	1	0.9%
Totals	326	100.0%	225	100.0%	106	100.0%

Do you score or have you ever scored writing portfolios? (T=Q17)

Currently a scorer	158	48.3%
Scored in the past	86	26.3%
Never scored	83	25.4%
Total	327	100.0%

Have you received Writing Professional Development training? (T=Q18)

Yes	280	87.2%
No	41	12.8%
Total	321	100.0%

To what extent do you agree or disagree that the amount of time it takes to prepare writing portfolios is appropriate for the benefit received by students? (T=Q19, P&S=Q14)

	Teacher		Principal		Superintendent	
Strongly disagree	158	48.9%	124	54.9%	41	39.4%
Disagree	102	31.6%	63	27.9%	39	37.5%
Agree	40	12.4%	27	11.9%	20	19.2%
Strongly agree	5	1.5%	9	4.0%	2	1.9%
Not sure	18	5.6%	3	1.3%	2	1.9%
Totals	323	100.0%	226	100.0%	104	100.0%

Do teachers who score portfolios have biases that affect scores? (T=Q20, P&S=Q15)

	Teacher		Principal		Superintendent	
No	75	23.1%	64	28.3%	17	16.0%
Yes, a few	148	45.5%	110	48.7%	54	50.9%
Yes, many	46	14.2%	40	17.7%	19	17.9%
Not sure	56	17.2%	12	5.3%	16	15.1%
Totals	325	100.0%	226	100.0%	106	100.0%

Students' work on portfolios is mostly done as part of writing assignments in several classes OR in English or writing classes? (T=Q21, P&S=Q16)

	Teacher		Principal		Superintendent	
Several classes	177	56.2%	142	63.4%	58	55.2%
English or writing classes	138	43.8%	82	36.6%	47	44.8%
Totals	315	100.0%	224	100.0%	105	100.0%

Students' work on portfolios is mostly done throughout the school year OR during a block of time during the year set aside for emphasis on portfolios? (T=Q22, P&S=Q17)

	Teacher		Principal		Superintendent	
School year	285	89.6%	209	92.9%	81	77.1%
Block of time	33	10.4%	16	7.1%	24	22.9%
Totals	318	100.0%	225	100.0%	105	100.0%

A student working on a portfolio is more likely to work with: (T=Q23, P&S=Q18)

	Teacher		Principal		Superintendent	
Only one teacher?	108	33.9%	68	30.6%	51	49.0%
More than one teacher?	211	66.1%	154	69.4%	53	51.0%
Totals	319	100.0%	222	100.0%	104	100.0%

Survey of High School Students

1) What is your overall opinion of the CATS tests?

	Number of Students	% of Total
Very positive	19	12.0%
Somewhat positive	42	26.6%
Neutral	57	36.1%
Somewhat negative	28	17.7%
Very negative	12	7.6%
Total	158	100.0%

2) Overall, how does CATS testing affect what students learn?

Very positively	19	12.1%
Somewhat positively	64	40.8%
Neutral or No Effect	53	33.8%
Somewhat negatively	17	10.8%
Very negatively	4	2.5%
Total	157	100.0%

3) Do your teachers have enough instructional time to teach the subject matter to be covered on the CATS tests?

Yes	82	51.6%
No	47	29.6%
Not sure	30	18.9%
Total	159	100.0%

4) Please indicate whether you agree or disagree with each of these statements about CATS testing.

Getting ready for or taking the CATS tests takes too much time away from class time.

Agree	92	57.9%
Disagree	67	42.1%
Total	159	100.0%

CATS testing is too stressful.

Agree	97	61.4%
Disagree	61	38.6%
Total	158	100.0%

CATS testing reduces enjoyment of learning.

Agree	102	65.0%
Disagree	55	35.0%
Total	157	100.0%

CATS testing is a good way to measure how schools are doing.

Agree	85	53.5%
Disagree	74	46.5%
Total	159	100.0%

Too often, what was covered on a CATS test was inappropriate for my grade level.

Agree	52	33.3%
Disagree	104	66.7%
Total	156	100.0%

CATS testing helps teachers and students focus on what's most important.

Agree	64	40.5%
Disagree	94	59.5%
Total	158	100.0%

I would do better if the test time was shorter.

Agree	41	25.9%
Disagree	117	74.1%
Total	158	100.0%

I would do better if the test time was longer.

Agree	89	56.7%
Disagree	68	43.3%
Total	157	100.0%

Too often, how questions are asked makes them more difficult to answer than necessary.

Agree	128	80.5%
Disagree	31	19.5%
Total	159	100.0%

Being taught only what is to be covered on the test is too limiting.

Agree	120	75.9%
Disagree	38	24.1%
Total	158	100.0%

CATS test results provide useful information about how well I'm doing in school.

Agree	78	49.1%
Disagree	81	50.9%
Total	159	100.0%

Teachers and students are forced to cover material too quickly because of CATS testing.

Agree	117	74.1%
Disagree	41	25.9%
Total	158	100.0%

5) About what percentage of class time throughout the school year do you spend preparing for CATS testing, including test-taking techniques and practice tests? Please do not include time spent learning the subject matter covered by the tests or time spent on writing portfolios.

[Open-ended question, responses coded by staff]

% of Time		
0-10	49	31.4%
11-20	24	15.4%
21-30	18	11.5%
31-40	15	9.6%
41-50	16	10.3%
51-60	6	3.8%
61-70	7	4.5%
71-80	11	7.1%
81-90	6	3.8%
More than 90	4	2.6%
Total	156	100.0%

Average=32.5, Median=25

6) Please rate the extent to which you agree or disagree whether each of the CATS tests listed below is a good measure of your knowledge of that subject.

Reading		
Strongly disagree	7	4.4%
Disagree	27	17.0%
Agree	72	45.3%
Strongly agree	39	24.5%
Not sure	14	8.8%
Total	159	100.0%

Math		
Strongly disagree	10	6.3%
Disagree	27	17.1%
Agree	68	43.0%
Strongly agree	43	27.2%
Not sure	10	6.3%
Total	158	100.0%

Science		
Strongly disagree	17	10.7%
Disagree	34	21.4%
Agree	69	43.4%
Strongly agree	31	19.5%
Not sure	8	5.0%
Total	159	100.0%

Social Studies		
Strongly disagree	7	4.4%
Disagree	32	20.3%
Agree	76	48.1%
Strongly agree	32	20.3%
Not sure	11	7.0%
Total	158	100.0%

Arts and Humanities		
Strongly disagree	15	9.5%
Disagree	43	27.2%
Agree	63	39.9%
Strongly agree	20	12.7%
Not sure	17	10.8%
Total	158	100.0%

Practical Living/Vocational		
Strongly disagree	13	8.2%
Disagree	39	24.5%
Agree	59	37.1%
Strongly agree	27	17.0%
Not sure	21	13.2%
Total	159	100.0%

7) Based on your own experience, to what extent do you agree that writing portfolios and on-demand writing have been good measures of how well you wrote at that time?

Writing portfolios		
Strongly disagree	22	14.0%
Disagree	26	16.6%
Agree	67	42.7%
Strongly agree	36	22.9%
Not sure	6	3.8%
Total	157	100.0%

On-demand writing		
Strongly disagree	27	17.3%
Disagree	31	19.9%
Agree	60	38.5%
Strongly agree	22	14.1%
Not sure	16	10.3%
Total	156	100.0%

8) In your view, what percent of students try their best on CATS tests?

None or almost none	14	8.9%
Some	83	52.5%
Most	48	30.4%
All or almost all	10	6.3%
Not sure	3	1.9%
Total	158	100.0%

9) How would you describe your effort on CATS tests?

I rarely or never gave my best effort	2	1.3%
I usually did not try my best	14	8.9%
I usually tried my best	47	29.7%
I tried my best for all the tests	95	60.1%
Total	158	100.0%

10) Have your parents or guardians received reports about your performance on the CATS subject tests?

Yes	108	68.4%
No	20	12.7%
Not sure	30	19.0%
Total	158	100.0%

11) Did your parents or guardians find that the information was provided in a way that was easy to understand?

Yes	66	61.1%
No	26	24.1%
Not sure	16	14.8%
Total	108	100.0%

12) Did your parents or guardians find the results useful?

Yes	53	49.1%
No	28	25.9%
Not sure	27	25.0%
Total	108	100.0%

13) Did you find the results useful?

Yes	79	51.0%
No	55	35.5%
Not sure	21	13.5%
Total	155	100.0%

14) To what extent do you agree that the amount of time it takes to prepare a writing portfolio is appropriate for the benefit you receive from it?

Strongly agree	21	13.3%
Agree	63	39.9%
Disagree	34	21.5%
Strongly disagree	27	17.1%
Not sure	13	8.2%
Total	158	100.0%

15) Is your work on portfolios mostly done as part of writing assignments in several classes OR in English or writing classes only?

Several classes	125	79.6%
English or writing classes	32	20.4%
Total	157	100.0%

16) Is your work on portfolios mostly done throughout the school year OR during a block of time during the year set aside for emphasis on portfolios?

School year	126	80.3%
Block of time	31	19.7%
Total	157	100.0%

17) When working on the portfolio, do you mostly work with:

One teacher?	80	51.3%
More than one teacher?	76	48.7%
Total	156	100.0%

18) Overall, how satisfied or dissatisfied are you with the writing portfolio?

Very satisfied	21	13.3%
Somewhat satisfied	73	46.2%
Somewhat dissatisfied	30	19.0%
Very dissatisfied	26	16.5%
Not sure	8	5.1%
Total	158	100.0%

19) To what extent do you agree that the performance categories (novice, apprentice, proficient, and distinguished) are appropriate categories to assign to students' writing portfolios?

Strongly agree	27	17.1%
Agree	76	48.1%
Disagree	23	14.6%
Strongly disagree	14	8.9%
Not sure	18	11.4%
Total	158	100.0%

20) Overall, how satisfied or dissatisfied are you with on-demand writing?

Very satisfied	11	7.0%
Somewhat satisfied	65	41.1%
Somewhat dissatisfied	34	21.5%
Very dissatisfied	31	19.6%
Not sure	17	10.8%
Total	158	100.0%

21) To what extent do you agree that the performance categories (novice, apprentice, proficient, and distinguished) are appropriate categories to assign to students' on-demand writing?

Strongly agree	21	13.3%
Agree	67	42.4%
Disagree	30	19.0%
Strongly disagree	19	12.0%
Not sure	21	13.3%
Total	158	100.0%

22) Do teachers who score portfolios have biases that affect scores?

No	24	15.2%
Yes, a few do	53	33.5%
Yes, many do	25	15.8%
Not sure	56	35.4%
Total	158	100.0%

Survey of Parents

1) Do you have a child in a public school in Kentucky in the 4th grade or higher? [Based on Question 1, only the results for parents who answered "yes" are included.]

2) What grade is your child in?

Grade	Number of Parents	% of Total
4	12	11.3%
5	12	11.3%
6	12	11.3%
7	8	7.5%
8	14	13.2%
9	9	8.5%
10	18	17.0%
11	10	9.4%
12	11	10.4%
Total	106	100.0%

3) Are the CATS tests fair measures of your child’s knowledge of school subjects?

Yes	44	42.3%
No	28	26.9%
Not sure	32	30.8%
Total	104	100.0%

4) Indicate whether you agree or disagree with these statements about CATS testing.

Getting ready for or taking the CATS tests takes too much time away from class time.

Agree	60	61.2%
Disagree	38	38.8%
Total	98	100.0%

The CATS tests are too stressful and reduce enjoyment of learning.

Agree	50	50.5%
Disagree	49	49.5%
Total	99	100.0%

The CATS tests are a good way to measure how schools are doing.

Agree	51	51.5%
Disagree	48	48.5%
Total	99	100.0%

Too often, what is covered on the CATS tests is inappropriate for my child.

Agree	27	28.1%
Disagree	69	71.9%
Total	96	100.0%

The CATS tests help teachers and students focus on what's most important.

Agree	40	41.7%
Disagree	56	58.3%
Total	96	100.0%

Teaching only what is to be covered on the CATS tests is too limiting.

Agree	86	86.9%
Disagree	13	13.1%
Total	99	100.0%

The CATS tests are a good way to measure how my child is doing in school.

Agree	45	45.9%
Disagree	53	54.1%
Total	98	100.0%

The CATS tests cover too many subjects.

Agree	21	22.3%
Disagree	73	77.7%
Total	94	100.0%

Teachers and students are forced to cover material too quickly in preparation for the CATS tests.

Agree	64	66.0%
Disagree	33	34.0%
Total	97	100.0%

6) Do you usually receive reports about your child's performance on the CATS subject tests?

Yes	79	75.2%
No	16	15.2%
Not sure	10	9.5%
Total	105	100.0%

7) Is the information in the CATS tests report presented in a way that makes it easy to understand?

Yes	59	74.7%
No	17	21.5%
Not sure	3	3.8%
Total	79	100.0%

8) Does someone from your child's school usually discuss the CATS tests report with you?

Yes	11	13.9%
No	66	83.5%
Not sure	2	2.5%
Total	79	100.0%

9) Does the CATS tests report help you understand how your child is doing in school?

Yes	36	45.6%
No	39	49.4%
Not sure	4	5.1%
Total	79	100.0%

10) Do you usually receive the School Report Card for your child's school?

Yes	94	90.4%
No	8	7.7%
Not sure	2	1.9%
Total	104	100.0%

11) Is the information presented in a way that makes it easy to understand?

Yes	77	81.9%
No	13	13.8%
Not sure	4	4.3%
Total	94	100.0%

12) Does the School Report Card help you understand how your child's school is doing?

Yes	54	57.4%
No	34	36.2%
Not sure	6	6.4%
Total	94	100.0%

13) To what extent do you agree or disagree that the benefit your child receives is worth the amount of time it takes to prepare writing portfolios?

Strongly agree	12	11.7%
Agree	19	18.4%
Disagree	22	21.4%
Strongly disagree	35	34.0%
Not sure	15	14.6%
Total	103	100.0%

14) To what extent are you satisfied or dissatisfied that your child's writing portfolio fits in with work done in different classes?

Very satisfied	20	19.0%
Somewhat satisfied	20	19.0%
Somewhat dissatisfied	15	14.3%
Very dissatisfied	34	32.4%
Not sure	16	15.2%
Total	105	100.0%

15) Do you think that your child's score on the writing portfolio is a good measure of how well your child writes?

Yes	46	43.8%
No	40	38.1%
Not sure	19	18.1%
Total	105	100.0%

Survey of School Board Members

1) Overall, how satisfied are you with CATS testing?

	Number of Respondents	% of Total
Very satisfied	19	12.8%
Somewhat satisfied	76	51.4%
Neutral	14	9.5%
Somewhat dissatisfied	27	18.2%
Very dissatisfied	12	8.1%
Total	148	100.0%

2) Given your satisfaction with CATS testing, what is your judgment of the amount of time spent by educators and students preparing for and taking the CATS tests? (Do not include time spent teaching or learning the subject matter to be covered on the tests.)

About the right amount	53	36.3%
Too much time	79	54.1%
Not enough time	6	4.1%
Not sure	8	5.5%
Total	146	100.0%

3) Should other subject areas or nonacademic factors be included in the assessment and accountability system?

No	67	45.3%
Yes	48	32.4%
Not sure	33	22.3%
Total	148	100.0%

4a) Please indicate whether you agree or disagree with each of these statements about CATS testing.

CATS testing assures that goals are the same for everyone

Agree	91	62.8%
Disagree	37	25.5%
Not sure	17	11.7%
Total	145	100.0%

CATS testing is too stressful.

Agree	72	48.3%
Disagree	56	37.6%
Not sure	21	14.1%
Total	149	100.0%

CATS testing reduces enjoyment of teaching and learning.

Agree	86	57.7%
Disagree	48	32.2%
Not sure	15	10.1%
Total	149	100.0%

CATS tests provide useful information about how well teachers are doing.

Agree	76	51.0%
Disagree	55	36.9%
Not sure	18	12.1%
Total	149	100.0%

Too often, test content is inappropriate for the students taking a test.

Agree	45	30.4%
Disagree	66	44.6%
Not sure	37	25.0%
Total	148	100.0%

CATS testing provides needed focus and organization.

Agree	93	62.4%
Disagree	38	25.5%
Not sure	18	12.1%
Total	149	100.0%

Question format is too often inappropriate for the students taking a test.

Agree	43	29.7%
Disagree	61	42.1%
Not sure	41	28.3%
Total	145	100.0%

CATS testing helps align the curriculum.

Agree	99	66.9%
Disagree	32	21.6%
Not sure	17	11.5%
Total	148	100.0%

Teaching what is to be covered on the test is too limiting.

Agree	92	63.0%
Disagree	37	25.3%
Not sure	17	11.6%
Total	146	100.0%

CATS test results provide useful information about how well students are doing.

Agree	87	58.4%
Disagree	46	30.9%
Not sure	16	10.7%
Total	149	100.0%

Teachers and students are forced to cover material too quickly because of CATS testing.

Agree	82	55.0%
Disagree	43	28.9%
Not sure	24	16.1%
Total	149	100.0%

CATS test results provide useful information about how well schools are doing.

Agree	88	59.1%
Disagree	44	29.5%
Not sure	17	11.4%
Total	149	100.0%

4b) For each of the following, indicate whether the item's weight in the accountability index for schools is about right, whether its weight should be increased or decreased, or whether the item should be eliminated from the school accountability index.

Reading

Eliminate	1	0.7%
Decrease	0	0.0%
About right	91	61.9%
Increase	47	32.0%
Not sure	8	5.4%
Totals	147	100.0%

Math

Eliminate	1	0.7%
Decrease	1	0.7%
About right	87	59.6%
Increase	50	34.2%
Not sure	7	4.8%
Total	146	100.0%

Science		
Eliminate	1	0.7%
Decrease	9	6.1%
About right	94	63.9%
Increase	35	23.8%
Not sure	8	5.4%
Total	147	100.0%
Social Studies		
Eliminate	1	0.7%
Decrease	13	8.9%
About right	103	70.5%
Increase	19	13.0%
Not sure	10	6.8%
Total	146	100.0%
Arts and Humanities		
Eliminate	6	4.1%
Decrease	47	32.2%
About right	55	37.7%
Increase	26	17.8%
Not sure	12	8.2%
Total	146	100.0%
Practical Living/Vocational		
Eliminate	9	6.2%
Decrease	33	22.6%
About right	64	43.8%
Increase	25	17.1%
Not sure	15	10.3%
Total	146	100.0%
Writing portfolios		
Eliminate	34	23.0%
Decrease	46	31.1%
About right	52	35.1%
Increase	5	3.4%
Not sure	11	7.4%
Total	148	100.0%

On-demand writing		
Eliminate	11	7.5%
Decrease	29	19.7%
About right	65	44.2%
Increase	28	19.0%
Not sure	14	9.5%
Total	147	100.0%
CTBS/5		
Eliminate	5	3.4%
Decrease	4	2.8%
About right	66	45.5%
Increase	31	21.4%
Not sure	39	26.9%
Total	145	100.0%

Appendix C

Description of the Analysis of Writing Portfolio Audits

Each year, a contractor for the Kentucky Department of Education audits a sample of writing portfolios to "verify or adjust" the portfolio scores provided by schools (CTB McGraw-Hill 1). Comparing the scores assigned by schools and the auditors provides an opportunity to evaluate how differences among raters affect the reliability of portfolios.

The auditing process is described in a report submitted by CTB McGraw-Hill to the Kentucky Department of Education, "2003-2004 Writing Portfolio Audit Rational and Procedures." According to the report, schools are selected for the audit through two samples. The first is a purposeful sample in which a school is selected because its writing portfolio index is significantly different than expected based on its other CATS scores. The expectations are set based on the relationship between schools' writing portfolio indices and schools' academic indices excluding the writing portfolio. The second way a school may be selected to be audited is through a random sample. For randomly selected schools, all portfolios except those marked incomplete or those "scored at an alternative school" were submitted to CTB McGraw-Hill (2). In addition, the portfolios of certain students were not audited. These portfolios were from students who were

- part of a foreign exchange program,
- exempt for medical reasons,
- limited English proficient,
- enrolled in public school for less than 100 days,
- participants in the alternate portfolio per KDE, or
- tested at non-A1 schools.

Certain programs such as those that focus entirely on vocational education, preschool, or special education are considered non-A1 schools. Table C.1 shows the number of schools and portfolios that were included in both the purposeful and random samples.

Only the random sample was used in the analysis. While the purposeful sample is useful for targeting an audit, it is not generalizable to all schools. In its 2002 Technical Report, KDE stated: "It is ... hoped that the random sample of schools will help provide a snapshot of statewide portfolio development and scoring" (9-8). Because the schools in this sample were randomly selected, their audit results should be reasonably representative.

TABLE C.1
Number of Schools and Writing Portfolios Included in the Purposeful and Random Audit Samples

	PURPOSEFUL				RANDOM				COMBINED
	4th	7th	12th	Total	4th	7th	12th	Total	
	Year 1999								
Schools	34	15	1	50	33	13	10	56	106
Portfolios	2,021	1,035	54	3110	2,063	2,424	1,664	6151	9261
	Year 2000								
Schools	39	8	3	50	32	14	10	56	106
Portfolios	1,957	1,634	425	4016	1,949	2,336	1,564	5849	9865
	Year 2001								
Schools	36	6	8	50	36	17	12	65	115
Portfolios	1,960	238	920	3118	2,076	2,490	2,071	6637	9755
	Year 2002								
Schools	27	22	1	50	28	11	12	51	101
Portfolios	1,279	3,236	81	4596	1,798	1,940	1,458	5196	9792
	Year 2003								
Schools	8	10	5	23	52	26	13	91	114
Portfolios	655	688	477	1820	2,824	3,013	1,938	7775	9595
	Year 2004								
Schools	17	5	4	26	42	24	9	75	101
Portfolios	674	427	437	1538	2,678	3,288	1,788	7754	9292

Source: Commonwealth. Kentucky Dept. of Ed.

According to KDE's 2002 Technical Report, audit scorers come from a variety of backgrounds including professional writers, college graduate students, and retired business people. Writing portfolio consultants from KDE train the auditors using the same procedures and materials used in training Kentucky teachers.

Before a school assigns a portfolio score, more than one rater may read the portfolio. Likewise, the auditor may have more than one rater read a portfolio before assigning it a final score. The data only include the final score assigned by the school and the final score assigned by the auditors. Therefore, the reliability measures only address these final scores and do not consider the process by which the schools and auditor arrive at their final scores.

Several measures of the consistency of raters were calculated. In calculating each of these measures, values were assigned to the four primary score classifications: a value of 13 was assigned for novice, a value of 60 was assigned for apprentice, a value of 100 was assigned for proficient, and a value of 140 was assigned for distinguished.

The interrater agreement rates represent the percentage of portfolios for which the auditor's scores and the schools' scores were the same. Correlations, which indicate the

degree to which the auditors and schools consistently rated portfolios, were based on a Pearson correlation.

A final measure considered in this report was the generalizability-study, or g-study. The g-study, developed by Cronbach et al (1972), measures the amount of variation or error that is attributable to various sources of error, such as the number of raters evaluating an assessment or the number of tasks that students must complete. These measures allow for an analysis of how each of the various sources of error affects the reliability of an assessment such as writing portfolios.

There are two types of g-studies that are typically performed: relative and absolute. Shavelson and Webb describe *relative* g-study as evaluating the reliability of an assessment that is used when making decisions about the relative rankings or ordering of students or schools. They describe *absolute* g-study as evaluating reliability used when making decisions about assigning or categorizing students or schools based on an absolute level of performance. For example, if a performance assessment intends to compare schools based on the scores of all the other schools, then a *relative* G-coefficient should be the appropriate measure. If the assessment intends to determine if schools have obtained some level of competency, regardless of the scores of other schools, then an *absolute* G-coefficient is appropriate. Because CATS assesses schools based on specific standards and not based on ranking relative to other schools, the *absolute* G-coefficient should be used to evaluate interrater reliability (HumRRO. "The Accuracy of School Classifications.")

The generalizability analysis presented in Chapter 4 measures the amount of error attributable to the schools, the raters, and interaction of schools with raters. The student-level audit data was collapsed into school-level data, with the two scores representing the writing portfolio index based on the school's original scores and the auditor's scores. Again, as with the previous reliability measures, only randomly selected schools were included in the analysis. The specific calculations for estimating the variance are described in Brennan (2003) and Swartz et al (1999).