



Office Of
Education
Accountability

Legislative Research Commission

Test Familiarity And Performance: Comparing Scores On Kentucky Core Content Tests And Unfamiliar Tests

Research Report No. 402

Prepared By

Karen Timmel, Acting Director; Deborah Nelson, PhD; Albert Alexander;
Emily Spurlock; Jamie Giles; Jamie Houston; Brenda Landy; and Sabrina Olds

Test Familiarity And Performance: Comparing Scores On Kentucky Core Content Tests And Unfamiliar Tests

Project Staff

Karen Timmel, Acting Director
Deborah Nelson, PhD
Albert Alexander
Emily Spurlock
Jamie Giles
Jamie Houston
Brenda Landy
Sabrina Olds

Research Report No. 402

Legislative Research Commission
Frankfort, Kentucky
lrc.ky.gov

Accepted December 10, 2013, by the
Education Assessment and Accountability Review Subcommittee

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

Foreword

In December 2012, the Education Assessment and Accountability Review Subcommittee directed the Office of Education Accountability (OEA) to analyze changes in the rank of Kentucky schools in reading and mathematics from 2011, the last year in which students took the Kentucky Core Content Test, and 2012, the first year in which students took the Kentucky Performance Rating for Educational Progress tests. The committee also requested that OEA seek to explain unexpected changes in the rank of Kentucky schools relative to one another.

OEA would like to thank staff at the Kentucky Department of Education as well as the many district and school staff who assisted with this study. OEA would also like to thank Daniel Koretz and Ronald Hambleton for providing technical advice and feedback.

Marcia Ford Seiler
Acting Director

Legislative Research Commission
Frankfort, Kentucky
December 10, 2013

Contents

Summary.....	vii
Chapter 1: Introduction And Background	1
Major Conclusions.....	2
Description Of This Study	4
Organization Of The Report	5
Previous Research On Test Validity.....	5
Gaps Between What Is Measured On State Standardized Tests And What Students Should Learn.....	6
State Curriculum Goals.....	6
State Standards.....	7
State Tests	8
Use Of State Test Data For Accountability.	8
Factors Limiting The Ability Of Large-Scale Standardized Tests To Assess All State Standards.....	9
Test Question Formats	9
Test Questions Available In Test Item Banks.....	10
Sample Of Items Tested.....	10
Test Blueprints.....	10
Importance Of Untested Standards	13
<i>Recommendation 1.1</i>	14
<i>Recommendation 1.2</i>	14
Role Of Classroom Assessments In Supporting Validity Of State Test Scores	14
Advantages And Disadvantages Of Test-Focused Instructional Practices	17
State Support For Classroom Assessments That Complement Standardized Tests	18
Kentucky Department Of Education Training.....	18
Reading And Mathematics Intervention Grants.	18
Previous Research On Differences Between Trends On State Tests And Tests Of Similar Content	19
Chapter 2: Reading And Mathematics Performance On KCCT Compared To Other Reading And Mathematics Tests	21
Comparing School Performance On KCCT, ITBS, And K-PREP Reading And Math Tests	21
Methodology.....	21
Tests And Years Compared	21
Converting Scores On Different Tests To Standard Scores.....	23
Use Of Percentile Ranks	23
Differences In Format And Function Of KCCT, K-PREP, And ITBS Tests.....	24
Limitations	25
School Percentile Rank Change From KCCT To ITBS And K-PREP Tests	25

Standard Score Change Associated With Percentages Of Students Eligible For Free Or Reduced-Price Lunch	29
On Average, High-Performing High-Poverty Schools That Drop In Rank Still Relatively High Performing.	33
School Rank Change Based On Percentage Of Minority Students	34
Factors That Might Explain Change In Standard Scores.....	35
Test Design	36
Range Of Performance.....	36
Concerns About Possible Bias In Multiple-Choice Reading Tests.	36
District And School Curricula.	37
Out-Of-School Factors.....	37
Comparing Changes In KCCT Scores With Changes In National Assessment Of Educational Progress Scores.....	37
Methodology.....	38
Results.....	39
Chapter 3: Assessment Practices In Site Visit Schools	41
Methodology	41
Use Of The Terms <i>Data</i> And <i>Assessment</i> In Site Visit Interviews	42
Limitations	43
Summary Of Site Visit Findings.....	43
Assessments In State Test Formats.....	46
Practices Common In Site Visit Schools	46
Unit Tests In KCCT-Like Formats.	46
Interim Tests	47
Influence Of Unit and Interim Tests On Content Taught.	47
Ongoing Use Of State Test Formats.....	48
Practices More Common In Schools That Decreased In Rank From KCCT To Other Tests.....	48
Ongoing Use Of Multiple-Choice Questions.....	50
Use Of KCCT-Like Assessments To Identify Specific Skills Emphasized In Support For Struggling Students.....	50
Instructional Practices Specific To Standardized Tests.....	50
Administrators' Focus On Data In Standardized Test Formats.....	51
Assessments In Formats Other Than Those Used On State Tests	51
Practices Common In All Schools	52
Classroom Assessment In A Variety Of Formats.....	52
Practices More Common In Schools That Increased In Rank	53
Strong Value Of Assessments Not In State Test Formats	53
Observational/Diagnostic Assessments.	54
Transfer Tasks.....	54
Assessment Of Accelerated Learners	55
Advantages And Disadvantages Of Test Performance Pressure	55
Advantages Of Test Performance Pressure.....	56
Disadvantages Of Test Performance Pressure	57
Conclusion	59

Content And Format Of State Test Affects Local Tests	59
Local Assessments Vary And May Affect Validity Of State Test Data.....	60
Some Practices That Improve Scores On State Tests May Not Improve Student Mastery Of All Standards	60
<i>Recommendation 3.1</i>	60
Appendix A: Differences In Percentages Of Greatest Depth Of Knowledge Questions On KCCT Reading And Math Tests.....	61
Appendix B: Comparing School Performance On KCCT And ITBS Tests In 2011	63
Appendix C: Changes In School Performance By Standard Score And Performance Decile	67
Appendix D: Standard Score Changes Of Highest-Performing Schools.....	71
Appendix E: Standard Score Change Excluding Highest- And Lowest-Performing Students...	73
Appendix F: Changes In KCCT And NAEP Reading And Math Scores	77
Appendix G: Site Visit School Demographic And Test Data	81
Appendix H: Site Visit Interview Protocols And Survey.....	83

Tables

2.1 Tests, Years, And Students Compared In Rank Change Analysis	22
2.2 Years And Grades In Which Kentucky Students Took Both The KCCT And The NAEP In Reading Or Math, 2000-2011	38
3.1 Use/Value Of Data In State Test And Other Formats In Eight Site Visit Elementary Schools, By Schools That Increased Or Decreased In Rank	45
3.2 Use Of Standardized Test Formats For Formal Assessment In Eight Site Visit Elementary Schools That Decreased Or Increased In Rank	46
3.3 Use Of Assessments Not In State Test Formats In Eight Site Visit Elementary Schools, Schools That Decreased Or Increased In Rank.....	52

Figures

1.A Relationship Of State Academic Content Standards To State Standardized Tests.....	12
1.B Illustration Of Test-Focused Assessment Versus Standards-Focused Assessment	17
2.A School Percentile Rank Change, KCCT Reading Tests To K-PREP And ITBS Reading Tests, 2010-2012	27
2.B School Percentile Rank Change, KCCT Math Tests To K-PREP And ITBS Math Tests, 2010-2012.....	28
2.C Average School Standard Score Change, KCCT Reading Tests To K-PREP And ITBS Reading Tests, By Percentage Of Students Eligible For Free Or Reduced-Price Lunch, 2010-2012	30
2.D Average School Standard Score Change, KCCT Math Tests To K-PREP And ITBS Math Tests, By Percentage Of Students Eligible For Free Or Reduced-Price Lunch, 2010-2012	31
2.E Average Standard Score Change, KCCT Reading 2010 To KCCT Reading 2011, KCCT Reading 2010 To ITBS Reading 2010, Schools Performing In Top 20 Percent On KCCT Reading 2010, By Percentage Of Students Eligible For Free Or Reduced-Price Lunch.....	32

2.F	Average Standard Score Change, KCCT Math 2010 To KCCT Math 2011, KCCT Math 2010 To ITBS Math 2010, Schools Performing In Top 20 Percent On KCCT Math 2010, By Percentage Of Students Eligible For Free Or Reduced-Price Lunch.....	33
2.G	Average Standard Score Change, KCCT Reading 2010 To ITBS Reading 2010, KCCT Reading 2011 To K-PREP Reading 2012, By Percentage Of Minority Students.....	35
2.H	Standard Unit Changes, KCCT And NAEP 4 th - And 8 th -Grade Reading And Math Tests, 2007-2011.....	39
3.A	Percentage Of Classroom Teachers Reporting KCCT-Influenced Instructional Practices, Schools That Decreased Or Increased In Rank.....	49
3.B	Percentage Of Teachers Who Agreed With The Statement “Overall, the pressure to ensure that students performed well on the KCCT had a positive effect on the quality of teaching and learning in this school.”.....	57
3.C	Percentage Of Teachers Reporting That “The pressure to ensure that students performed well on the KCCT led me to make instructional decisions that were against my own standards of best practice.”.....	58

Summary

Education policies rely heavily on state standardized test data to gauge system quality and spur improvement. These data guide decisions about instruction, allocation of funds, and program evaluation. It is important that state test data be valid—that they mean what they are assumed to mean—for the variety of purposes for which they are used.

This study looks at one factor that may affect the validity of state test data: gaps between what is measured on standardized tests used for accountability and what students should know, as described in state academic content standards. Student learning of some state standards cannot be assessed, or cannot be assessed completely, with the types of questions typically used on large-scale standardized tests. State tests contain more questions on some standards than on others.

When state tests are used for accountability and become familiar over time, educators may feel pressure to focus more on frequently tested content than on other content that is also important to subject mastery. Educators may spend more time preparing students to answer the types of questions used on state tests than preparing them to apply knowledge in situations likely to be encountered in future education and adult life. When the focus is shifted in such ways, gains observed on state tests may differ from actual gains in student mastery of content standards. If the teaching of tested versus untested content varies among schools, the comparability of school test data could be reduced.

The Kentucky Core Content Test (KCCT) was used for more than a decade to hold educators accountable for student learning. This study evaluates the validity of KCCT data in reading and mathematics as they were used to monitor state performance over time and judge the academic performance of schools relative to each other. The study compares changes over time in state KCCT reading and math scores with changes in Kentucky students' scores on the National Assessment of Educational Progress (NAEP), a test administered every 2 years to a sample of Kentucky students. The study also compares schools' performance on the KCCT with their performance on two other standardized reading and math tests that had never been administered statewide—the Kentucky Performance Rating for Educational Progress (K-PREP) and the Iowa Test of Basic Skills (ITBS).

In most cases, changes over time in KCCT scores are validated by changes in NAEP scores. In fact, Kentucky students made greater gains on unfamiliar NAEP reading tests than they did on the familiar KCCT tests. KCCT and NAEP comparisons identify only one area of potential concern: 8th-grade math. Between 2007 and 2011, gains of Kentucky students were more than 2½ times as great on the KCCT as on the NAEP.

KCCT data also appear valid for judging the performance of most Kentucky schools relative to each other. Most schools that performed well on the familiar KCCT performed relatively well on other standardized tests. Most schools that were low performing on the KCCT were also low performing on these other tests. However, some schools (about 7 percent of all schools) performed much worse on the unfamiliar K-PREP and ITBS tests than on the KCCT, relative to their peers.

Discrepancies in KCCT and other test scores may be explained in part by differences in schools' focus on tested versus untested standards. Office of Education Accountability staff visited eight elementary schools, chosen from a larger sample of schools that performed either much better or much worse on unfamiliar K-PREP and ITBS tests than on the familiar KCCT. Schools that performed better on the KCCT than on unfamiliar tests had focused heavily on KCCT-tested standards and question formats. KCCT scores of these schools may have overstated students' mastery of reading and math standards relative to the mastery of other students in the state. Schools that performed even better on unfamiliar K-PREP and ITBS tests than on the familiar KCCT were more likely to have focused on a broader range of standards. KCCT scores in these schools may have understated students' mastery of reading and math standards relative to the mastery of other students in the state. Data collected from this small sample of schools provide preliminary evidence that the validity of state test data as used to judge the performance of schools may have been affected by differences among schools' focus on all standards versus those most likely to appear on state tests.

State test data alone may not always be valid for identifying practices that improve student learning of state academic standards. Some of the practices used in site visit schools that were high performing on KCCT reading and math tests did not lead to similarly high performance on other reading and math tests. Teachers in most site visit schools reported that the pressure to perform on state standardized tests caused them reduce emphasis on untested standards. In most site visit schools, for example, teaching of science and social studies was reduced or eliminated in grades in which those subjects were not tested. It is important that the data used to identify best practices in education reflect the full range of state standards and not primarily those aspects that are tested. If the data used to make instructional decisions are taken primarily from tests in state test formats, untested or less tested aspects of state standards may be overlooked.

The study makes three recommendations.

Recommendation 1.1

The Kentucky Department of Education should consider creating a document that clarifies which aspects of state standards are not fully assessed on state tests and should be assessed at the local level. The document should recommend methods for assessing aspects of state standards not assessed on state tests.

Recommendation 1.2

The Kentucky Department of Education should consider including in state, district, and school report cards a brief description of and reference to state documents that identify aspects of state standards that are not assessed on state tests and should be assessed at the local level.

Recommendation 3.1

When adopting strategies believed to increase state test scores in a particular subject area, districts and schools should look for evidence that the strategies improve student learning of all state standards in that subject. Evidence should include data capable of measuring aspects of standards not assessed in standardized formats.

Chapter 1

Introduction And Background

Education policies rely heavily on state standardized test data to gauge system quality and spur improvement. It is important that data from state tests be valid—that they mean what they are assumed to mean.

Education policies rely heavily on state standardized test data to gauge system quality and spur improvement. These data guide decisions about instruction, allocation of funds, and program evaluation. It is important that state test data be valid—that they mean what they are assumed to mean—for the variety of purposes for which they are used.

Policy makers and the public assume that state test data show whether students have learned what they are supposed to know and are prepared to use that knowledge in future education and adult life.

Policy makers and the public assume that state test data show whether students have learned what they are supposed to know and are prepared to use that knowledge in future education and adult life. Students in schools with high state test scores are assumed to know more about tested subjects than students in schools with lower scores. When state test scores increase over time, policy makers assume that students' mastery of the tested subjects has improved.

This study focuses on one factor that can affect the validity of these assumptions: gaps between what students should know and what is measured on state standardized tests.

This study looks at one factor that may affect the validity of these assumptions: gaps between what is measured on standardized tests used for accountability and what students should know, as described in state academic content standards. When particular state tests are used over time, educators become familiar with the content and types of questions most likely to appear on the test. Because they are accountable for students' performance on state tests, educators may feel more pressure to focus on highly tested content and skills than on those that are not tested or less tested but still critical for subject mastery.

The validity of test data may be influenced by what students are taught. State test data from a school that focuses on all standards may be more valid than state test data from a school that focuses on highly tested standards and question formats.

The validity of test data may be influenced by what students are taught. State test data are more likely to reflect students' subject mastery in schools that teach all standards and prepare students to demonstrate their knowledge in a variety of situations. Data from schools that focus heavily on highly tested standards and question formats may indicate less about students' broader subject mastery. If schools vary in their focus on tested versus untested content, their state test data may not be comparable.

This study evaluates the validity of data from the Kentucky Core Content Test (KCCT) as they were used to judge the academic performance of Kentucky schools relative to each other and to monitor state academic performance over time.

The study compares schools' performance on KCCT reading and math tests with their performance on two other reading and math tests that had never been given statewide. The study looks for potential links between discrepancies in schools' test scores and their focus on tested versus untested standards.

The study also compares changes over time in state performance on KCCT reading and math tests with changes in Kentucky students' performance on the National Assessment of Educational Progress (NAEP).

KCCT reading and math data appear valid as a way of judging the academic performance of most schools in reading and math relative to each other.

The Kentucky Core Content Test (KCCT) was used for more than a decade to monitor and hold educators accountable for student learning in the commonwealth. This study uses data from other standardized tests to evaluate the validity of KCCT reading and math data as they were used to judge the academic performance of Kentucky schools relative to each other and to monitor state academic performance over time.

If KCCT data reflect students' mastery of the subject tested, then students in schools that performed well on the KCCT should have performed relatively well on other standardized tests. The study compares schools' KCCT reading and math scores with their scores on two other tests: the Kentucky Performance Rating for Educational Progress (K-PREP) in 2012 and the Iowa Test of Basic Skills (ITBS) in 2010. K-PREP and ITBS data were taken from the first year in which the tests were given statewide. While most of the content on the tests should have been familiar to students, the particular tests were not. Staff also conducted site visits to eight elementary schools, chosen from a larger sample of schools that performed much better or worse on the unfamiliar tests than they had on the KCCT. Site visits explored potential links between the validity of state test data in these schools and the schools' focus on tested versus untested standards.

The study also compares changes over time in state performance on KCCT reading and math tests with changes in state performance on the National Assessment of Educational Progress (NAEP). Studies conducted in other states and previously in the commonwealth have shown discrepancies between changes in state test scores and NAEP scores.

Major Conclusions

- KCCT reading and math data appear valid as a way of judging the academic performance of most schools in reading and math, relative to each other. The KCCT rank of most Kentucky elementary and middle schools did not change dramatically when students took the K-PREP and ITBS reading and math tests. Most schools that performed well on the KCCT performed relatively well on the K-PREP and ITBS. Most schools that were low performing on the KCCT were also relatively low performing on the K-PREP and ITBS.

Students in about 7 percent of schools performed much worse on unfamiliar tests than they did on the KCCT.

State tests appear to exert a strong influence on which standards are taught most.

The validity of state test data in some schools may have been affected by the schools' focus on tested versus untested content.

In most cases, changes over time in the performance of Kentucky students on the KCCT are validated by the performance of Kentucky students on the National Assessment for Educational Progress. In 8th-grade math, however, student gains on the KCCT were more than 2½ times as great as gains on the NAEP.

- Some schools (about 7 percent of all schools) performed much worse relative to their peers when students took the unfamiliar K-PREP and ITBS tests than they had on the KCCT. Higher-poverty schools were more likely than other schools to lose rank when students took unfamiliar tests, especially those higher-poverty schools that were high performing on the KCCT.
- Data gathered from site visits to eight elementary schools suggest that state tests exert a strong influence on which standards are taught in most schools and on the methods used to teach those standards. State tests are more than measurement instruments; they are powerful drivers of instruction.
- The validity of state test data in some schools may have been affected by the schools' focus on tested versus untested content. Site visit schools that focused heavily on KCCT-tested content and question formats performed much worse relative to their peers when students took unfamiliar tests. KCCT scores for these schools may have overestimated students' mastery of state standards relative to their peers. Site visit schools that focused systematically on standards not tested on the KCCT performed even better on unfamiliar tests than they had on the KCCT. KCCT scores for these schools may have underestimated students' mastery of state standards relative to their peers.
- In most cases, changes over time in the performance of Kentucky students on the KCCT are validated by the National Assessment of Educational Progress (NAEP), which is administered every 2 years to a sample of Kentucky students. Between 2007 and 2011, Kentucky students made greater gains on the unfamiliar NAEP reading test than on the familiar KCCT reading test. The validity of gains on the KCCT 8th-grade math test is less certain; between 2007 and 2011, gains on the KCCT were more than 2½ times as great as gains on the NAEP.

Description Of This Study

In December 2012, the Education Assessment and Accountability Review Subcommittee directed the Office of Education Accountability to analyze changes in the rank of Kentucky schools in reading and math from the final year of the KCCT to the first year of the Kentucky Performance Rating for Educational Progress (K-PREP).

Staff also compared the performance of schools on KCCT and Iowa Test of Basic Skills (ITBS) and compared state performance over time on the KCCT and the NAEP.

Staff analyzed a variety of factors that might explain discrepancies between school performance on the KCCT reading and math tests and other tests.

In December 2012, the Education Assessment and Accountability Review Subcommittee directed the Office of Education Accountability (OEA) to analyze changes in the rank of Kentucky schools in reading and math between 2011—the final year in which students took the KCCT—and 2012, the first year in which students took the K-PREP. The committee also requested that OEA seek to explain differences between schools that maintained performance rank relative to their peers and schools whose rank changed unexpectedly.

In conducting the study, staff analyzed school rank change from the KCCT to K-PREP reading and math tests and school rank change from the KCCT to ITBS reading and math tests. Staff also compared changes over time in KCCT 4th- and 8th-grade reading and math scores to changes in NAEP 4th- and 8th-grade reading and math scores.

Staff analyzed a variety of factors that might explain discrepancies between school performance on the KCCT reading and math tests and other tests. These included student demographic characteristics, school assessment practices, and technical features of individual tests.

Data analyzed for this study consist primarily of student achievement and demographic data provided by the Kentucky Department of Education. These include

- student- and school-level KCCT reading and math data, 2000-2011;
- student- and school-level K-PREP reading and math data, 2012;
- student- and school-level ITBS reading and math data, 2010 and 2011; and
- student- and school-level data on student ethnicity and eligibility for free or reduced-price lunch programs.

Staff also analyzed NAEP data from the US Department of Education and site visit data that included interviews, surveys, and document collections in eight elementary schools.

Organization Of The Report

Chapter 1 describes the relationships among state curriculum goals, state content standards, and state tests. It also reviews research.

Chapter 2 compares scores on KCCT reading and math tests to scores on other standardized tests of reading and math.

Chapter 3 describes local assessment practices as identified in site visits to eight elementary schools.

Chapter 1 describes the relationships among state curriculum goals, state content standards, and state tests. It also reviews research documenting the effects of standardized tests on instructional practices and describes discrepancies found in the past between state test scores and other tests of similar content.

Chapter 2 compares school performance on KCCT reading and math tests and other standardized reading and math tests. It describes relationships between schools' demographic characteristics and their tendencies to perform relatively better or worse on unfamiliar tests. The chapter also compares changes over time in KCCT and NAEP reading and math scores.

Chapter 3 describes findings from OEA site visits to eight elementary schools, four of which had performed better on the KCCT than on unfamiliar tests, relative to their peers, and four of which had performed worse. Site visits looked at whether discrepancies in schools' scores might be explained by their focus on KCCT-tested standards versus untested or less-tested standards.

Previous Research On Test Validity

State statutes require that assessment validity be monitored.

Measurement experts have defined validity in educational testing as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.”¹ State statutes require that assessment validity be monitored regularly. Kentucky Revised Statute (KRS) 158.6453(17) requires that the Kentucky Department of Education (KDE) develop a biennial plan for validation studies that address many aspects of test validity. Duties of OEA as described in KRS 7.410(2)(c)(5) include conducting studies that “analyze, verify, and validate the state assessment program through other external indicators of academic progress” such as the National Assessment of Educational Progress.^a

The Kentucky Department of Education (KDE) has looked at many aspects of test validity and has found generally positive relationships between KCCT data and data from other sources.

Research conducted by the Human Resources Research Organization (HumRRO) in connection with KDE's validation research agenda has focused largely on issues of question development, test design, and scoring. In addition, several studies

^a In a 2011 report, OEA addressed security-related threats to test validity. That report addressed issues of test administration and identified unusual patterns in some state assessment data. The research described in this report does not address test security issues.

have used data from other sources to evaluate the validity of KCCT data. These studies have generally shown positive relationships among students' KCCT scores, other tests, and students' grades. The research has also identified positive relationships between school performance on state tests and indicators of school quality such as culture, leadership, and instructional practice.

Research conducted for KDE has also identified some discrepancies between KCCT data and data from other sources.

HumRRO's research has also identified some discrepancies between KCCT data and data from other sources. For example, higher-performing schools do not always exhibit stronger use of recommended practices than lower-performing schools.² A recent analysis of school-level differences between KCCT and K-PREP scores indicated that, on average, higher-poverty schools performed worse on the unfamiliar K-PREP test than they did on the familiar KCCT, relative to other schools.³

This study expands on validity research conducted in the past and analyzes a variety of factors that might explain school-level differences in performance on KCCT versus unfamiliar tests.

This study expands on validity research conducted in the past by comparing school-level scores on the KCCT with unfamiliar tests in 3 years and by comparing state gains on KCCT with state gains on NAEP. It also analyzes a variety of factors that might explain school-level differences in performance on KCCT versus unfamiliar tests.

Gaps Between What Is Measured On State Standardized Tests And What Students Should Learn

This section explains why there are gaps between what is measured on state standardized tests and what students should learn, as described in state academic content standards in reading and math. The discussion does not address standards and tests in other areas such as social studies, science, and writing. However, the limitations described in this section are also relevant to those subjects.

State Curriculum Goals

State statutes define broad curriculum goals but do not identify specific skills and concepts. The Kentucky Board of Education disseminates state curriculum standards to local schools and districts.

Kentucky's broad curriculum goals are described in KRS 158.6451. As relevant to reading and math, these goals include high levels of student achievement as well as the ability to use and apply basic skills and core concepts, think and solve problems, connect new knowledge with previous knowledge, and acquire new information. The statute specifies that students should have the capacity to apply these skills in the variety of situations and for the variety of purposes they will encounter throughout their lives.

The statute does not describe specific skills and concepts in particular subject areas but directs the Kentucky Board of Education to disseminate to local districts and schools a curriculum framework that is tied to these goals. This curriculum framework is commonly referred to as state standards.

State Standards

State standards are documents that identify the skills, concepts, and abilities expected of Kentucky students in particular subjects. The state recently adopted new standards in reading and math.

State standards are documents that identify the skills, concepts, and abilities expected of Kentucky students in particular subject areas. Prior to 2011, the Kentucky Program of Studies described the content standards in reading and math. In 2006, these standards were further clarified in the Core Content for Assessment that described the particular skills and concepts that would be assessed in each grade. Senate Bill 1, enacted by the General Assembly in 2009 through KRS 158.6453, directed KDE to revise state content standards. The standards that were adopted in reading and math, called the Kentucky Core Academic Standards, were developed jointly by the National Governors Association and the Council of Chief State School Officers. The standards that were adopted in reading are included in the Kentucky Core English language arts and literacy standards.

State standards documents include broad skills and goals and a list of specific skills and concepts.

State standards documents contain sections that describe broad goals and sections that include lists of specific skills and concepts. Broad goals are often found in introductory or explanatory texts. These sections describe ways in which students should understand and use skills, often providing examples of student behaviors that exhibit the goals. Here is an example of broad language contained in Kentucky's English language arts and literacy standards:

Students establish a base of knowledge across a wide range of subject matter by engaging with works of quality and substance. They become proficient in new areas through research and study. They read purposefully and listen attentively to gain both general knowledge and discipline-specific expertise.⁴

Here is an example of an individual skill described in the English language arts and literacy standards for grade 8 students:

“Compare and contrast the structure of two or more texts and analyze how the differing structure of each text contributes to its meaning and style.”⁵

State standards are sometimes equated with the specific concepts and skills described, but the documents describing state standards include many goals that are not captured fully in any particular skill.

While state standards are sometimes equated with the lists of specific concepts and skills, the documents include many goals that are not captured fully in any particular skill. Thus, the state standards are more than the sum of the individual skills and concepts described.^b

State Tests

State standardized tests are used to monitor student learning in academic subjects. State tests are designed to yield reliable data that are consistent across schools.

State standardized tests are the primary means by which policy makers and the public are informed about students' mastery of state standards in reading, math, and most academic subjects. For this purpose, standardized tests are valued over grades or teacher-made tests because they are known to yield consistent scores from students across classrooms and schools. Test questions are developed and tested to ensure that they are reliable, meaning students giving similar answers will receive similar scores regardless of where they take a test. Each question is also reviewed by test developers and a panel of other content experts, including Kentucky teachers, to validate its use in assessing a particular content standard.

From 1999 to 2011, state standards in reading and math were assessed primarily by the KCCT. In 2012, state tests in reading and math were changed to align with new standards through the K-PREP system of assessments.

State test data are used to hold districts and schools accountable for student learning. It is considered especially important that test data be reliable when they are used in state accountability systems.

Use Of State Test Data For Accountability. It is especially important that tests be reliable when they are used to assign rewards or consequences. State test data are used in state and federal policies that establish systems of rewards and consequences associated with test performance targets. The specific details of these accountability policies have changed in the last decade, but they have generally involved public recognition of schools and districts considered successful and a series of required plans, interventions, and state assistance for schools and districts that do not meet performance targets. The state's accountability system for schools and districts is established in KRS 158.6455.

^b This report is suggesting that standards that should be learned include the introductory or explanatory language surrounding specific skills and concepts. It is not suggesting that the standards that should be learned include exemplary texts or recommended practices.

Factors Limiting The Ability Of Large-Scale Standardized Tests To Assess All State Standards

Several factors affect the ability of large-scale standardized tests to fully assess state standards. These include limits in the types of test question formats that can be used when policy makers require high levels of reliability in scoring, the range of questions available to assess particular standards, and limits in the standards that can be tested in the time that students spend taking the state test.

Multiple-choice questions are favored on accountability tests because they yield reliable data.

Test Question Formats. Questions on tests used for accountability must constrain the data collected from students to the type of information that can be collected and scored with high reliability. These requirements are most commonly met through multiple-choice questions that restrict the answer selections available to students.

Questions that require students to write their responses allow for a wider variety of answers but must still constrain the questions asked and the scoring procedures to standardized formats.

Many standardized tests, including the KCCT and the K-PREP, also contain questions that require students to write their answers. These questions allow for a wider range of student responses than do multiple-choice questions and are used to assess learning of more complex standards. To be reliable, however, written questions must give students specific prompts and scoring directives and must be scored using rubrics that define the specific characteristics required of successful answers. Thus, written questions on standardized tests must also be constrained to a standardized format.

Some aspects of state standardized are difficult or impossible to measure using the types of questions typically used on large-scale standardized tests.

Some aspects of state standards are difficult or impossible to measure using the types of questions typically used on large-scale standardized tests. These questions do not assess whether students can perform a task, such as read with expression. They are not able to determine whether students can complete an extended and complex task (such as a research report) or identify which questions, problems, and skills are needed in a specific situation, without directives or prompts. These questions do not always reveal whether a student understands a concept well enough to apply it in a variety of situations. A student may answer a question correctly even without fully understanding the concept. For example, a teacher interviewed by OEA for this study explained that her students could answer questions about “main idea” on the KCCT but had difficulty when asked to answer questions about “central idea” on the K-PREP. The students had understood the concept of main idea in a particular context but did not understand it well enough to know that it describes a broader concept that could be described using a variety of terms.

Standardized tests could include a wider range of question formats, such as performance tasks. These assessments would likely be costly and require more time to administer. To date, performance assessments have not been as reliable as other formats.

The limitations of large-scale standardized tests in assessing all state standards are stated clearly in KDE technical manuals for both the KCCT and the K-PREP.

The ability of any state test to assess the full range of standards is linked in part to the size and quality of test items available.

Because of limitations in the amount of time that can be spent testing students, tests contain a small sample of the items that would be needed to assess any standard completely.

Test blueprints specify the proportion in which different sets of standards are tested and the percentages of questions believed to require more complex levels of understanding.

Standardized tests could include a wider range of question formats. For example, students could be observed performing a task such as reading a book, computing mentally, participating in a discussion, or conducting an experiment. However, these assessments would likely be costly to develop or to administer to large groups of students and would take more time to administer. To date, scoring of performance tasks that require students to use skills in situations that simulate real life have not been as reliable as multiple-choice or written questions with specific prompts. Portfolio assessments are no longer used in Kentucky, in part because of concerns about reliability of scoring.

The limitations of large-scale standardized tests in assessing all state standards are stated clearly in KDE technical manuals for both the KCCT and the K-PREP. For example, the K-PREP manual explains that state assessments do not measure all state content standards because some standards “are impractical or impossible to measure with a standardized assessment.”⁶ These manuals also explain that test scores cannot necessarily be assumed to represent student learning of all state standards.

Test Questions Available In Test Item Banks. The ability of any state test to assess the full range of standards is linked in part to the size and quality of test items available. Many different questions are necessary to test the full range of knowledge required for mastery of any individual standard and to provide questions that capture student performance at different levels of understanding. Test banks must also be large enough to provide questions for different test forms from year to year. High-quality test items are expensive to develop and to validate.

Sample Of Items Tested. Because of limitations in the amount of time that can be spent testing students, individual test forms never include the full range of items that would be necessary to test every standard, even when these items are available. Tests are always a small sample of the items that would be needed to assess students’ complete understanding of a particular standard.

Test Blueprints. Items that are sampled on individual test forms are chosen according to a test blueprint that specifies the proportion in which different sets of standards are tested and the percentages of questions believed to require more complex levels of understanding. As shown in Appendix A, the distribution of questions believed to assess the greatest depth of knowledge varied substantially among KCCT tests in reading and math and in

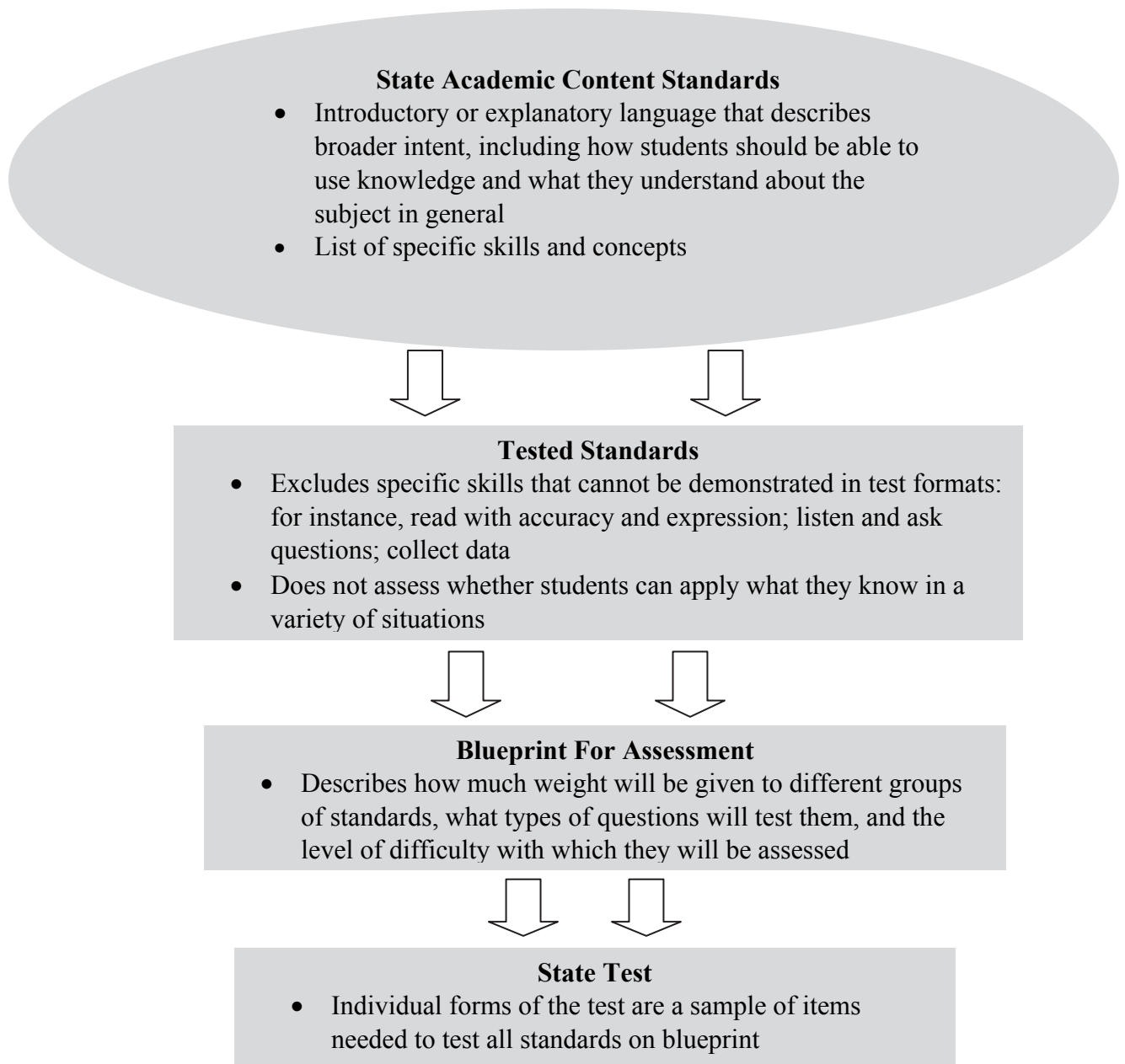
different years. In Kentucky, as in many states, test blueprints are available to the public.

Over time, educators become familiar with the content and the types of questions likely to appear on the test.

Over time, educators can become familiar with the content and the types of questions most likely to appear on the test. Some of the items that have been used on state tests are also released to the public.

Figure 1.A shows the relationship between state academic content standards and state standardized tests, as described above. The figure describes a sequence of steps in which the comprehensive set of skills and abilities expected of students in state content standards is narrowed to the set of skills that are actually tested on a particular state test. In each phase, standards that are not tested are those that cannot be tested at all or cannot be tested completely on any individual test.

Figure 1.A
Relationship Of State Academic Content Standards To State Standardized Tests



Source: Staff analysis of KDE technical assessment manuals and assessment literature.⁷

Some state tests assess a broader range of state standards than others. The subset of standards that are tested on any particular test is influenced by question formats, the range of test items available, and time spent testing.

Some state tests assess a broader range of state standards than do others. The subset of standards that are tested on any particular test is influenced by question formats, the range of test items available, and time spent testing. Some of the guidelines required of the state assessment system by KRS 158.6453 limit the range of standards that can be tested on any individual test form.⁸ The statute limits the amount of time that can be spent testing students. It also requires that the test contain some items that allow individual students to be compared to a national sample of students.

Researchers have not yet reached agreement on the extent to which large-scale standardized tests assess the full range of state standards.

Researchers have not yet reached agreement on the extent to which large-scale standardized tests assess the full range of state standards. Some feel that the vast majority of state standards can be tested on state tests even if they are not tested on an individual test form or in any single year. As long as educators know that most standards will be tested over the course of several years, they will feel accountable for teaching those standards.⁹ Others suggest that existing state tests do not contain the full range of items necessary to assess state standards and often contain items that are not related to the standards. Drawing from a national sample of state standards and tests, a 2011 study found that 50 percent or less of state standards were assessed on individual tests when the cognitive complexity of items was taken into consideration.¹⁰

Importance Of Untested Standards

The standards that are not tested are considered critical for ensuring mastery of tested content. They are not extra standards to be taught only if time permits.

The standards that are seldom or never tested can be critical for ensuring mastery of state standards. They are not extra standards to be taught only if time permits. Technical manuals for the KCCT provided examples from Kentucky's previous standards of "supporting standards" that were not tested, yet critical to the student's deep understanding of the overall content ... to be used by schools to build a foundation of knowledge, skills, and processes that will enable students to be successful on the *Kentucky Core Content Test*. **In order for students to reach proficiency and beyond on the KCCT, students need to master the supporting content as well as the state assessed.**¹¹ (emphasis added)

The state's current standards in math and English language arts and literacy contain some specific skills as well as thinking processes and performance tasks that cannot be assessed at all or cannot be assessed completely by large-scale standardized tests.

Assessment experts recommend explicitly identifying the gaps between what should be learned and what can be tested:

Most tests are unlikely to cover the full domain of content covered by content standards. Hence, it is important to make it clear which aspects of the content standards are left uncovered by the test, which are covered only lightly, and which receive the greatest emphasis.¹²

In identifying these gaps, it is important to consider the cognitive processes described in state standards, not just the particular skills and concepts that are listed.¹³ Introductory or explanatory language often addresses thinking processes (such as critical thinking or problem solving) that are not assessed fully in standardized formats.

Recommendation 1.1

Recommendation 1.1

The Kentucky Department of Education should consider creating a document that clarifies which aspects of state standards are not fully assessed on state tests and should be assessed at the local level. The document should recommend methods for assessing aspects of state standards not assessed on state tests.

Recommendation 1.2

Recommendation 1.2

The Kentucky Department of Education should consider including in state, district, and school report cards a brief description of and reference to state documents that identify aspects of state standards that are not assessed on state tests and should be assessed at the local level.

Role Of Classroom Assessments In Supporting Validity Of State Test Scores

If students are instructed disproportionately on tested content and taught to demonstrate knowledge in tested formats, then the data gained from the tested subset of standards will not necessarily reflect student learning of the entire set of standards.

If students are taught the entire set of learning standards outlined in state documents but are tested on only a subset of those standards, then the data gained from the subset sampled could be considered reasonably representative of the entire set. However, if students are instructed disproportionately on tested content and are taught to demonstrate knowledge in tested formats, then the data gained from the tested subset of standards will not necessarily reflect student learning of the entire set of standards.¹⁴

The validity of assumptions about student learning based on state test scores alone rests in part on what is taught and assessed at the school level.

Thus, the validity of assumptions about student learning based on state test scores alone rests in part on what is taught and assessed at the school level. In particular, it requires teachers to be teaching and assessing standards that the state test does not necessarily assess. This point is made clearly in the technical manual accompanying the K-PREP test:

It is through the classroom teacher that these (untested) standards and benchmarks are assessed. However, the Kentucky test is used for assessment of proficiency with respect to all standards. This is appropriate only if interpretations of the scores on the test can be validly extrapolated to apply to the larger domain of student achievement.¹⁵

Local assessments should complement state assessments, both in the frequency of the data they produce and in the nature of the data they produce.

As described in KRS 158.6453(8), the state’s system of balanced assessment may include district or school assessments that “provide data on how well their students are growing toward mastery of Kentucky academic core content.” These local systems are important, given limitations in the data provided by state tests. Data from state tests are limited because they are available only at the end of the year—too late to inform instruction—and, as discussed in this chapter, because they are not able to provide data on all aspects of state standards. Thus, local assessments are important not only for the frequency with which they generate data but for the nature of the data they are able to generate. Using observation of students, analysis of student work, or performance tasks, local assessments can provide information about students’ understanding of content and their application of knowledge that are not typically provided in standardized test formats.

Research suggests that state tests influence teachers to adopt classroom assessments in standardized test formats and to focus instruction on the particular qualities of state tests.

If schools are to assess student learning of all standards, they must sometimes use assessments in formats other than those used on standardized tests. However, research conducted in many states has demonstrated teachers’ tendencies to

- make frequent use of test formats used on state tests during classroom instruction,
- use items previously used on state tests to teach and assess students,
- conduct practice and review sessions aligned with tested content and formats, and
- teach test-taking strategies that increase students’ familiarity with test formats and scoring procedures.¹⁶

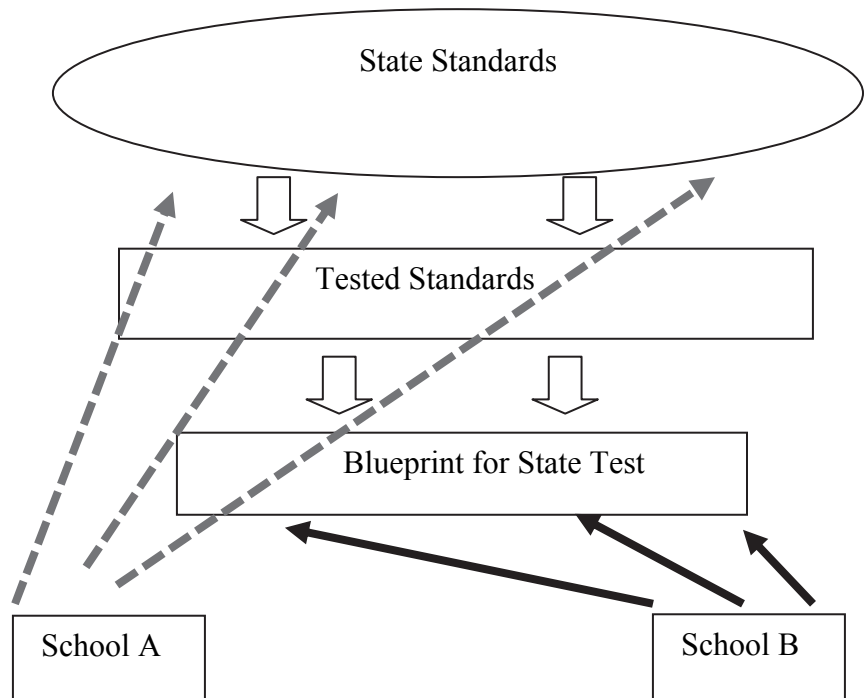
Research suggests, further, that state tests shape the allocation of resources and instructional time towards content tested on state tests and away from untested subjects.¹⁷

Given the central role of state standardized tests in the state's accountability system, educators may have greater incentives to focus on tested content and formats than on those that are not tested.

Given the central role of state standardized tests in the state's accountability system, educators may have greater incentives to focus on tested content and formats than on those that are not tested. As explained by a principal in one site visit school, state tests are “what we are graded on.” These incentives may be even greater for educators in schools that serve poor and minority children because these students tend to have lower performance on standardized tests. Research has documented greater use of test-focused instructional practices in schools under pressure to meet performance targets.¹⁸

Figure 1.B illustrates the effect that school assessment practices might have on the validity of state test data in two hypothetical schools. School A uses the full range of assessments necessary to determine whether students have mastered state standards. School B assesses students primarily in formats similar to those used on the state test. To the extent that assessment data drive instructional decisions, students in School B would be less likely to be taught the full range of state standards than students in School A. For this reason, assumptions about student learning of state standards based on state test data in School A would be more likely to be valid than assumptions based on state test data in School B.

Figure 1.B
Illustration Of Test-Focused Assessment Versus
Standards-Focused Assessment



Source: Staff analysis of assessment literature.¹⁹

Advantages And Disadvantages Of Test-Focused Instructional Practices

This report focuses on the possible negative consequences of test-focused instructional practices as they affect the validity of state test data.

Test-focused instructional practices can have positive as well as negative consequences.

This report focuses on the possible negative consequences of test-focused instructional practices as they affect the validity of state test data. The report does not fully address the advantages and the disadvantages of test-focused instructional practices.

Test-focused instructional practices can have positive as well as negative consequences. Positive consequences include increased motivation and accountability, identification of specific skills that need greater emphasis in school curricula, and identification of content not yet mastered by students. Also, when the question formats of state tests model desired instructional behaviors, they can increase teachers' emphasis on these behaviors.²⁰ For example, KCCT open response questions required students to explain their thinking and provide specific examples from texts. Some have referred to tests that model desirable behaviors as "tests worth teaching to."²¹

State Support For Classroom Assessments That Complement Standardized Tests

Kentucky Department Of Education Training

KDE has provided training to all Kentucky districts on the range of assessments available and the strengths and weaknesses of each in assessing different types of learning.

Following direction of Senate Bill 1, KDE provided training to district and school staff in every Kentucky district on the range of school practices required in implementing the new reading and math standards. This training stressed assessment literacy for educators about the range of assessments available and the strengths and weaknesses of each in assessing different types of learning. The training included student performance and work products as important components of classroom assessments.²²

Training provided by KDE in connection with the statewide teacher Professional Growth and Evaluation System, to be fully implemented in 2015, also highlights the importance of classroom assessments that address learning goals not captured on standardized tests. For example, the Framework for Teaching, which was adopted in connection with the system, includes the following indicators of highly effective teaching: teachers understand the cognitive structures that ensure student understanding and plan lessons to address possible student misconceptions; teacher-designed assessments are authentic and have real-world applications; and teachers use a variety or series of questions or prompts to challenge students' thinking and promote higher-order thinking and understanding.²³

Reading And Mathematics Intervention Grants

The General Assembly funds grants for early intervention programs in reading and math. These programs include diagnostic assessments that provide data not available on state standardized tests.

The General Assembly has established reading and math intervention programs to assist students who are struggling in those subjects. Both train teachers to use assessments that provide types of data that are not available in state standardized test formats. The Reading Diagnostic and Intervention Grant Program, established in KRS 158.792 provides grants to schools for comprehensive programs to assist struggling readers. These programs include diagnostic assessment of early reading. The Mathematics Achievement Fund, established in KRS 158.844, provides grants to schools for comprehensive programs to assist students struggling in math. These programs include diagnostic assessments of foundational numeracy skills. Both programs focus primarily on students in the primary grades though assessments can also be used for older students who are struggling.

Previous Research On Differences Between Trends On State Tests And Tests Of Similar Content

Concerns about the validity of state test data alone are not new. They have been fueled in part by discrepancies between trends on state tests and trends on other tests of similar content.

Concerns about the validity of assumptions about student learning based on state test data alone are not new. They have been fueled in part by discrepancies between trends on state tests and trends on other tests of similar content. These discrepancies have been identified in a number of states and districts. In Texas, student state test score gains were three or more times greater in 4th-grade reading than on the NAEP 4th-grade reading test and four or more times greater on the state 8th-grade math test than on the NAEP 8th-grade math test.²⁴ Similar discrepancies were found in the early years of the Kentucky Instructional Results Information System (KIRIS) test. Between 1992 and 1994, average 4th-grade KIRIS reading scores increased substantially whereas NAEP average 4th-grade reading scores remained virtually unchanged.²⁵ Between 1992 and 1996, increases in average 4th-grade KIRIS math scores were 3.6 times greater than increases in average 4th-grade NAEP math scores, and increases in average 8th-grade KIRIS math scores were 4.1 times greater than increases in average 8th-grade NAEP math scores.²⁶ Discrepancies between increases on state tests and NAEP have also been found in North Carolina and Connecticut.²⁷ In Chicago, student increases on the ITBS reading and math tests during the late 1990s were not mirrored by increases in the state test in the same content areas. During this period, the ITBS was associated with high stakes for students and teachers in Chicago, but the state test was not.²⁸

Some researchers have hypothesized that these discrepancies are caused by disproportionate teaching of content and formats on accountability tests. Others caution that discrepancies in test data alone cannot be used to draw conclusions about instruction. Depending on the content tested on each test, some tests may be more tightly aligned with the curricula that teachers were expected to teach than others.

Chapter 2

Reading And Mathematics Performance On KCCT Compared To Other Reading And Mathematics Tests

This chapter evaluates the validity of KCCT reading and math data as they were used to rank Kentucky schools relative to each other and to monitor state academic progress over time. It compares KCCT data with data from reading and math tests that had never been administered statewide.

Overall, results do not raise serious concerns about the validity of KCCT data as they were used to judge the performance of Kentucky schools relative to each other. However, students in some schools—especially higher-poverty schools with high KCCT scores—perform much worse relative to their peers when they take unfamiliar tests.

For the most part, KCCT data also appear valid for monitoring the progress of Kentucky students over time.

This chapter evaluates the validity of KCCT reading and math data as they were used to judge the academic performance of Kentucky schools relative to each other and to monitor state academic progress over time. Because the KCCT was used for more than a decade, the content and question formats likely to appear on the test would have been familiar to educators and students. KCCT data are compared with data from other standardized tests that had never been administered statewide. While most of the content should have been familiar to students, the particular tests were not.

Overall, results do not raise serious concerns about the validity of KCCT data as they were used to judge the academic performance of schools relative to each other. Most of the schools that were high performing on the KCCT were relatively high performing on unfamiliar K-PREP and ITBS tests. However, some schools—especially higher-poverty schools with high KCCT scores—performed much worse on unfamiliar tests than on the KCCT, relative to their peers. Data presented in this chapter do not reveal the cause of this pattern. Chapter 3 presents some preliminary evidence that discrepancies in some schools' test scores may have been caused in part by the schools' heavy focus on content likely to appear on the KCCT.

For the most part, KCCT data also appear valid for monitoring the progress of Kentucky students over time. In reading, Kentucky students made greater gains on the unfamiliar NAEP test than on the familiar KCCT. However, Kentucky students' gains on KCCT 8th-grade math tests were more than 2½ as great as their gains on the NAEP.

Comparing School Performance On KCCT, ITBS, And K-PREP Reading And Math Tests

Methodology

Tests And Years Compared. Table 2.1 shows the tests used to compute changes in schools' performance from familiar to unfamiliar tests. The first two rows of data describe comparison of

school KCCT scores to school scores on unfamiliar tests of similar content:

- KCCT reading and math tests in 2010 to ITBS reading and math tests in 2010, and
- KCCT reading and math tests in 2011 to K-PREP reading and math tests in 2012.

The third row of data describes comparison of KCCT school scores in 2010 with KCCT school scores in 2011. This analysis was included to provide some sense of how much schools' performance changes from one year to the next, even on a familiar test.

Data used to compare school scores on two tests include only students who took each test in the same school. The analysis does not include high schools because data for comparison tests were not available in high school grades.

Table 2.1
Tests, Years, And Students Compared In Rank Change Analysis

Base Test for Comparison	Compared With	Student Scores Used to Calculate School Scores	Total Number of Student Scores Used to Compute School Scores	Number of Schools Included in Analysis*
Compares School Scores on KCCT Tests With Unfamiliar Tests of Similar Content				
KCCT reading and math grades 3-7, 2010**	ITBS reading and math grades 3-7, 2010**	Scale scores of students who took both the KCCT and the ITBS tests in 2010 in the same school	>227,000	963
KCCT reading and math grades 3-7, 2011	K-PREP reading and math grades 4-8, 2012	Scale scores of students who took the KCCT in 2011 and the K-PREP in 2012 in the same school	>166,000	948
Compares School Scores on KCCT with School Scores on KCCT Subsequent Year				
KCCT reading and math grades 3-7, 2010	KCCT reading and math grades 4-8, 2011	Scale scores of students who took the KCCT in 2010 and 2011 in the same school	>162,000	946

Notes: *Schools with fewer than 30 tested students on both tests were not included in the analysis.

**Staff also compared school scores on KCCT and ITBS in 2011. Results of the 2011 analysis are reported in Appendix B. They were similar to KCCT 2010 and ITBS 2010 results reported in this chapter.

Source: Staff compilation.

Converting Scores On Different Tests To Standard Scores. The school scores reported in this chapter are generated from scale scores of individual students. Scale scores are numerical scores that indicate the total number of points earned out of the total number possible on an individual test. Scale scores allow for comparison of student scores among different test forms and also provide for a more exact measure of achievement than do the broad performance designations of novice, apprentice, proficient, and distinguished that are typically used to report student scores.

This analysis compares schools' scores on different tests using a standard measure that adjusts for differences among tests.

To adjust for differences in the range and distribution of scale scores on the KCCT, K-PREP, and ITBS tests, staff converted each student's score on each test into a standard score. Standard scores provide a standard unit of measure by which scale scores on different tests can be compared.

A student's standard score is the difference between the student's score and the average, or mean, score of all students who took the test, divided by the standard deviation from the mean of all students who took the test. The standard deviation is a measure of how closely student scores are dispersed around the mean. After converting to standard scores, a student scoring at the state average on any test would have a standard score of 0. A standard score would be positive for a student scoring above the state mean and negative for a student scoring below the state mean. The standard score for a student scoring 10 points above the state mean would depend on how far other students' test scores were from the mean. The student's standard score would be higher if a small percentage of students tested scored 10 points or more above the mean and lower if a high percentage of students tested scored 10 points or more above the mean.

To compute school standard scores on individual tests, staff computed the average standard scores of students in each grade and then weighted each grade's standard score by the number of students in that grade to compute an average standard score for each school.

Each school was also assigned a percentile rank based on its standard score on each test.

Use Of Percentile Ranks. Each school was also assigned a percentile rank based on its standard score on each test. In some analyses, school rank change is reported as percentile rank change. This is done to make results easier to understand. School ranks are commonly reported in percentiles and rarely reported in standard scores.

An increase or decrease in a school's percentile rank does not indicate whether it increased or decreased in any absolute sense, only whether its rank increased or decreased relative to other schools. Also, a school can increase in rank and still be relatively low performing or decrease in rank but still be relatively high performing.

School percentile rank change was computed by subtracting a school's percentile rank on the KCCT from its percentile rank on a subsequent test. For example, if a school ranked in the 80th percentile on the KCCT reading test in 2010 and the 75th percentile on the KCCT reading test in 2011, the school would be assigned a percentile change of -5. The rank change does not indicate whether a school increased or decreased in any absolute sense, only whether its rank increased or decreased relative to other schools. Also, a school can increase in rank and still be relatively low performing or decrease in rank but still be relatively high performing. For example, if a school's score stayed the same from 2010 to 2011 but the scores of other schools increased, then the school's rank would decrease. Similarly, a school's rank could increase by 20 percentile points or more even if the school still scored below the state average. This could happen if the scores of other schools dropped—as many did on the K-PREP test in 2012—but the school's score stayed the same.

The format and content of tests used in this analysis differ in ways that could affect results reported in this chapter.

Differences In Format And Function Of KCCT, K-PREP, And ITBS Tests. The format and content of tests used in this analysis differ in ways that could affect results reported in this chapter. The KCCT and the K-PREP were criterion-referenced tests intended to determine whether students mastered material required to meet expectations set through a predetermined standard-setting procedure.^c In theory, all students could attain a proficient score if they met the predetermined standards. In contrast, the ITBS is a norm-referenced test. Norm-referenced tests are designed to determine how students compare to other students who took the test. While many of the test questions on the ITBS are similar in difficulty and content to the questions on KCCT and K-PREP, some were chosen specifically because they were unlikely to be answered correctly by most students and could therefore be used to rank students relative to each other.

The format of the test questions also differed. All of the tests contained multiple-choice questions, but the KCCT and K-PREP tests also included questions that required written responses, with 50 percent of the possible points on the test associated with these questions. The KCCT contained one form of written response, whereas the K-PREP contained two.

The content assessed on the tests was also different. The KCCT assessed student learning of Kentucky's state standards through

^c K-PREP also includes questions designed to allow comparison of Kentucky students' performance with national norms. Those questions were not used in this analysis.

2011. The K-PREP assessed student learning of Kentucky's new reading and math standards. The new standards suggest that many skills be taught in earlier grades than they had been taught previously. ITBS questions were chosen to be representative of content typically taught in the tested grade. However, it is likely that some of the questions on the ITBS did not align with Kentucky standards in the grade tested.

For different reasons, both the ITBS and the K-PREP captured a broader range of student ability than did the KCCT. In some grades, 10 percent or more of students attained the highest score possible on the KCCT. By design, the percentage of students attaining the maximum score on the ITBS is close to 1. In theory, a high percentage of students could attain the maximum score on the K-PREP tests. In fact, less than 1 percent of students attained the maximum score in 2012. Much of the content tested on the K-PREP was more difficult than the content that had been tested previously.

The analysis presented in this chapter does not address all concerns raised in Chapter 1 about the gap between what students should learn as described in state standards and what is measured on large-scale standardized tests.

This chapter looks for relationships between schools' demographic characteristics and changes in their performance on KCCT and unfamiliar tests. It does not identify the underlying causes of these relationships.

Limitations. The analysis presented in this chapter does not address all concerns raised in Chapter 1 about the gap between what students should learn as described in state standards and what is measured on state standardized tests. None of the data analyzed in this chapter directly captures learning goals that cannot be assessed in standardized formats. However, the analysis should indicate whether assumptions about the academic performance of Kentucky schools in reading and math based on KCCT scores are reflected by results on other tests.

This chapter looks for relationships between schools' demographic characteristics and changes in their performance on KCCT and unfamiliar tests. It does not indicate whether these trends are explained by school instructional practices, features of individual tests, differences in students' opportunities outside of school, or other factors.

School Percentile Rank Change From KCCT To ITBS And K-PREP Tests

Figures 2.A and 2.B show school percentile rank change from the familiar KCCT to the unfamiliar ITBS and K-PREP reading and math tests. Because schools change rank from year to year even on familiar tests, the figures also provide a baseline for comparison—schools' rank change from the KCCT reading and math tests in 2010 to the KCCT reading and math tests in 2011. Appendix C

contains data similar to data presented in Figures 2.A and 2.B but shows changes of schools by standard scores.

About half of Kentucky schools remained within 10 percentile points of their original rank on the KCCT when students took unfamiliar tests. More schools changed rank when students took unfamiliar tests than did from year to year on the KCCT.

About half of Kentucky schools remained within 10 percentile points of their original rank on the KCCT when students took unfamiliar tests, just as they did when students took the KCCT in a subsequent year. However, the percentile rank change of schools was slightly greater when students took unfamiliar tests than it was from year to year on KCCT tests.

- Comparing percentile rank of schools on the KCCT reading and math tests in 2010 to the KCCT reading and math tests in 2011, the percentage of schools that stayed within 10 percentile points of their original rank was 57 in reading and 56 in math.
- Comparing percentile rank of schools on the KCCT reading and math tests in 2011 with the K-PREP reading and math tests test in 2012, the percentage of schools that stayed within 10 percentile points of their original rank was 47 in reading and 49 in math.

About twice as many schools dropped more than 30 percentile points in rank when students took unfamiliar tests as did from year to year on the KCCT.

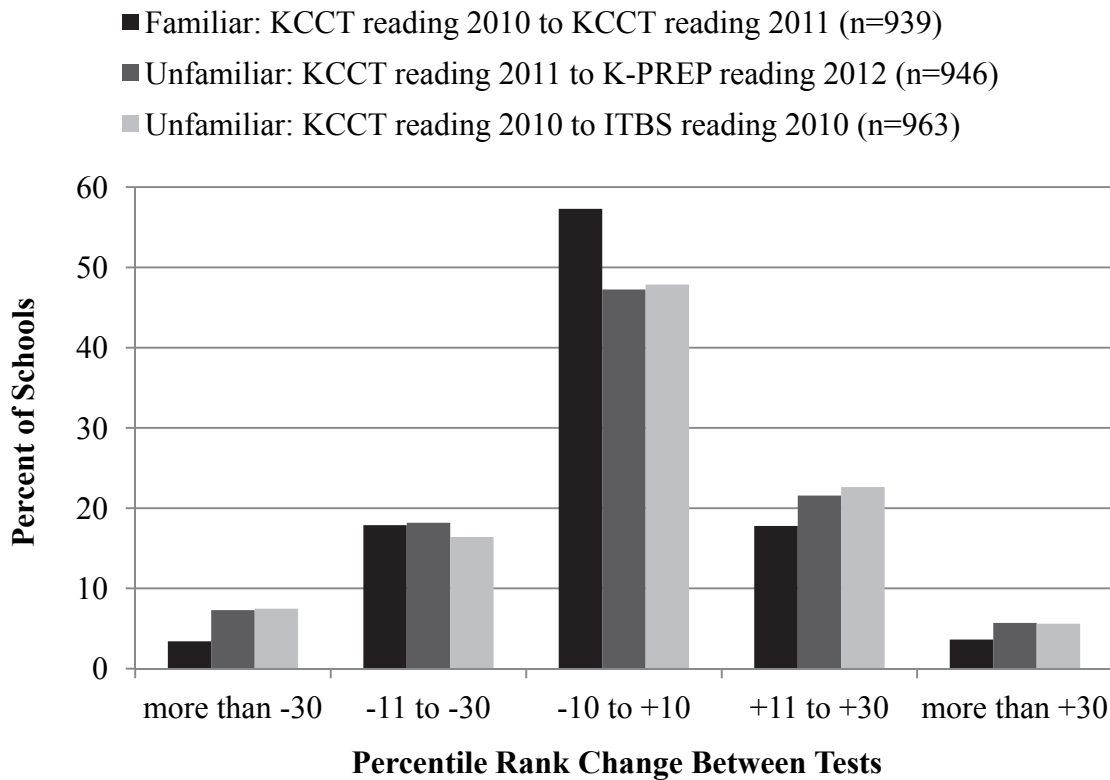
About twice as many schools dropped more than 30 percentile points in rank when students took unfamiliar tests as did from year to year on the KCCT:

- Comparing KCCT 2010 and KCCT 2011 reading and math tests, the percentage of schools that dropped by more than 30 percentile points was 3.4 in reading and 3.7 in math.
- Comparing KCCT 2011 and K-PREP 2012 reading and math tests, the percentage of schools that dropped by more than 30 percentile points was 7.3 in reading and 7.1 in math.

Figures 2.A and 2.B also show greater rank change from the familiar KCCT 2010 reading and math tests to the unfamiliar ITBS reading and math tests in 2010. Thus, schools changed rank more when compared on different tests in the same year than they did when compared on the same test in a subsequent year.

A greater percentage of schools changed rank on ITBS reading tests than did on ITBS math tests. When ranked on the ITBS reading test in 2010, 48 percent of schools stayed within 10 percentile points of their rank on the KCCT reading test in the same year. When ranked on the ITBS math test in 2010, 53 percent of schools stayed within 10 percentile points of their rank on the KCCT math test in the same year.

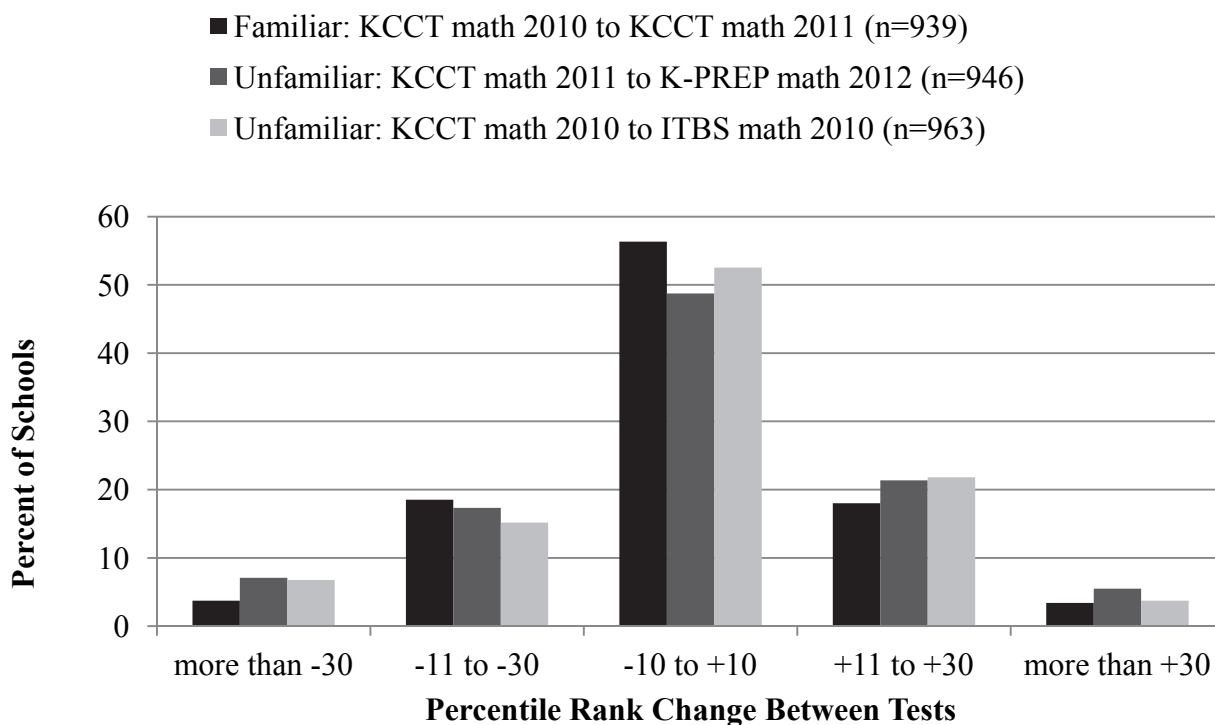
Figure 2.A
School Percentile Rank Change
KCCT Reading Tests To K-PREP And ITBS Reading Tests
2010-2012



Note: Rank change does not indicate absolute score, just change in rank relative to other schools. Many of the schools that increased in rank were still performing below state averages, and many of the schools that decreased in rank were still performing above state averages.

Source: Staff analysis of data from the Kentucky Department of Education.

Figure 2.B
School Percentile Rank Change
KCCT Math Tests To K-PREP And ITBS Math Tests
2010-2012



Note: Rank change does not indicate absolute score, just change in rank relative to other schools. Many of the schools that increased in rank were still performing below state averages, and many of the schools that decreased in rank were still performing above state averages.

Source: Staff analysis of data from the Kentucky Department of Education.

Most of the schools that were high performing on the KCCT were still high performing on the K-PREP. Most of the schools that were low performing on the KCCT were still low performing on the K-PREP.

The percentile rank change of schools that were highest and lowest performing on the KCCT was similar to the change reported above for all schools. As shown in Appendix C, at least half of the schools that were in the top or bottom 10 percent of schools on KCCT reading and math tests in 2011 remained in the top or bottom 10 percent of schools on K-PREP reading and math tests in 2012. The overwhelming majority of schools in the top 10 percent on the KCCT were still in the top 30 percent on the K-PREP, and the overwhelming majority of schools that were in the bottom 10 percent of schools on the KCCT were still in the bottom 30 percent on the K-PREP.

Standard Score Change Associated With Percentages Of Students Eligible For Free Or Reduced-Price Lunch

Changes in school performance from familiar to unfamiliar tests are associated with the percentage of students eligible for free or reduced-price lunch. This indicator is often used as a proxy for student poverty.

Data presented in the previous section showed that school percentile rank changed slightly more when students took unfamiliar tests than it did when students took a familiar test in 2 different years. This section shows how school standard scores change from one test to another based on the percentage of students eligible for the federal free or reduced-price lunch program. This measure is often used as a proxy for students considered to be living in poverty, although the percentage of students who qualify for the program is higher than the percentage of students who are living in poverty by federal definitions.^d Trends reported in standard scores are similar to what they would be as reported in percentile ranks. Standard scores are a more exact measure of difference between scores.

Figures 2.C and 2.D show the average standard score change on reading and math tests when schools are grouped according to the percentage of students eligible for free or reduced-price lunch. Both show greater variation among schools when students took unfamiliar tests than when students took the KCCT tests from one year to the next.

When students took unfamiliar tests, the scores of students in the lowest-poverty schools increased, on average, and the scores of students in the highest-poverty schools decreased, on average.

As shown in Figure 2.C, standard score changes from the KCCT reading test in 2010 to the KCCT reading test in 2011 did not differ substantially based on the percentage of students eligible for the program. Standard scores increased slightly, on average, in schools serving the lowest or highest percentages of students living in poverty. When students took unfamiliar reading tests, however, the standard score increased, on average, in schools serving lower percentages of students in poverty and decreased, on average, in the schools in which more than 80 percent of students were living in poverty. Increases in lower-poverty schools were greatest on the ITBS reading test, which includes only multiple-choice questions. Decreases were greatest in highest-poverty schools on the ITBS reading test.

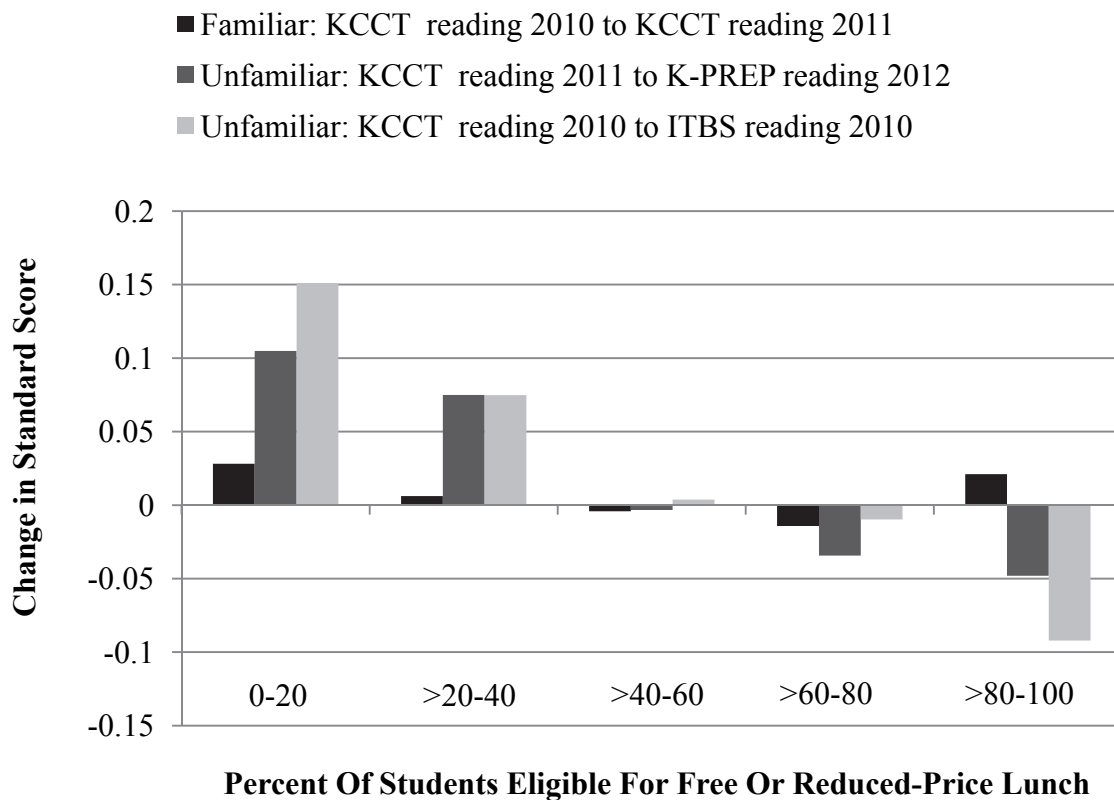
School rank change from familiar to unfamiliar math tests showed similar patterns associated with poverty.

Figure 2.D shows similar patterns on math tests. On average, standard unit scores of the highest-poverty schools increased from the KCCT math test in 2010 to the KCCT math test in 2011 but decreased from the familiar KCCT math test to the unfamiliar ITBS math test in 2010 and decreased even more from the familiar KCCT math test in 2011 to the unfamiliar K-PREP math test in

^d Students are eligible for free or reduced-price lunch if their family incomes are within 185 percent of the income that meets federal definitions of poverty.

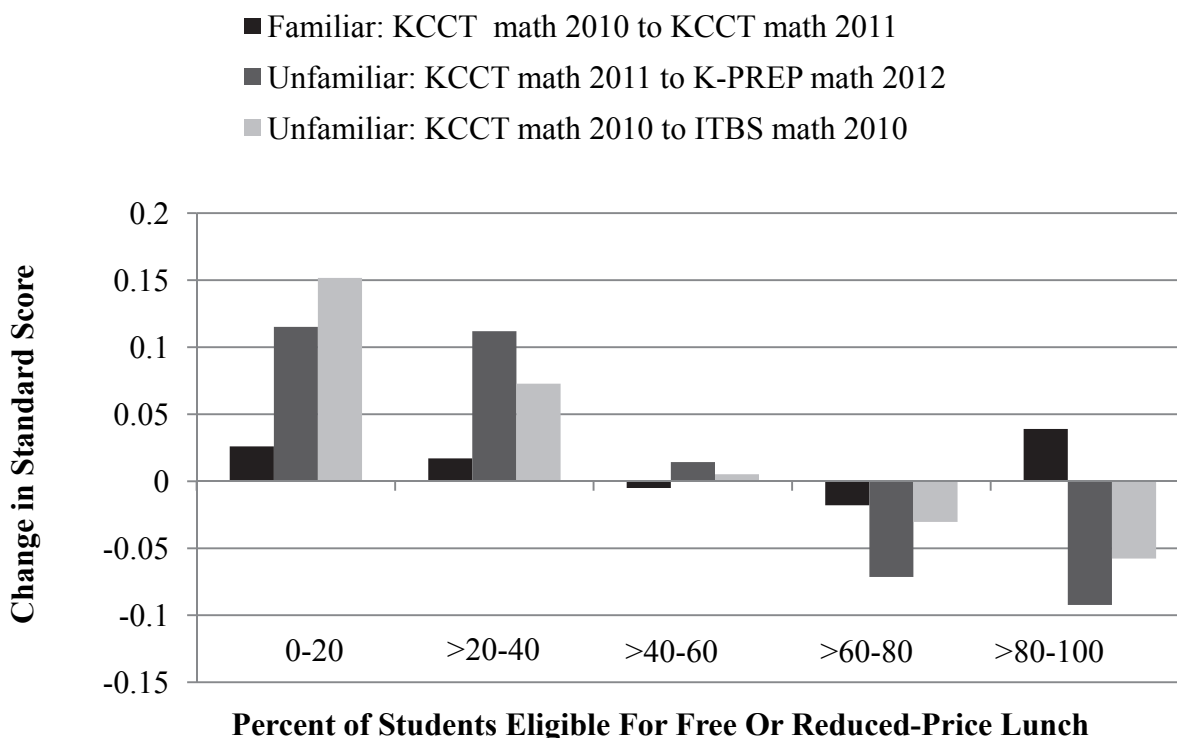
2012. In contrast, average standard scores of schools serving lower percentages of students in poverty increased from the 2010 KCCT test to the 2010 ITBS test and the 2011 KCCT test to the 2012 K-PREP test.

Figure 2.C
Average School Standard Score Change
KCCT Reading Tests To K-PREP And ITBS Reading Tests
By Percentage Of Students Eligible For Free Or Reduced-Price Lunch
2010-2012



Note: The figure reports average standard score change for all schools in designated school poverty ranges. Most ranges included some schools that increased in rank and some schools that decreased in rank.
Source: Staff analysis of data from the Kentucky Department of Education.

Figure 2.D
Average School Standard Score Change
KCCT Math Tests To K-PREP And ITBS Math Tests
By Percentage Of Students Eligible For Free Or Reduced-Price Lunch
2010-2012



Note: The figure reports average standard score change for all schools in designated school poverty ranges. Most ranges included some schools that increased in rank and some schools that decreased in rank.
Source: Staff analysis of data from the Kentucky Department of Education.

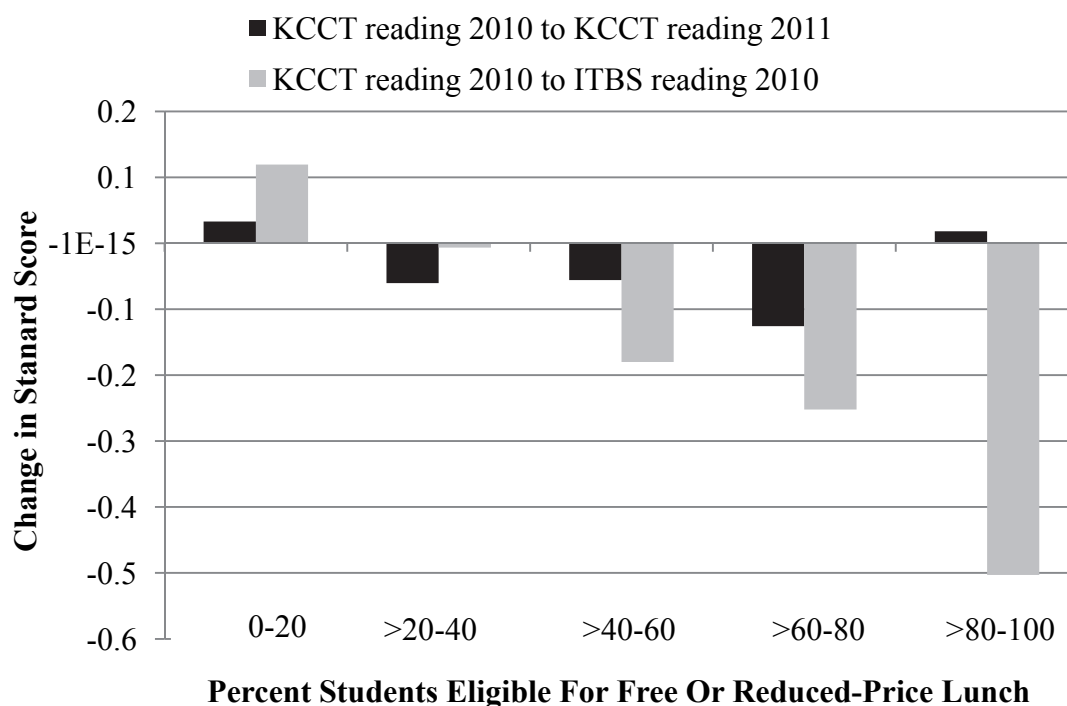
On average, the standard score of higher-poverty schools that were high performing on the KCCT dropped substantially when students took unfamiliar tests.

Changes in standard scores based on eligibility for free or reduced-price lunch were stronger among schools that were highest performing on the KCCT. Figures 2.E and 2.F show the average standard score change from the KCCT 2010 reading and math tests to the ITBS 2010 reading and math tests for only those schools that performed in the top 20 percent on the KCCT. The figures also show standard score changes for this group of schools from the KCCT 2010 to the KCCT 2011 reading and math tests.

It would be expected that this highest-performing group of schools would perform relatively less well on subsequent tests; this is the general tendency of high or low outliers in any type of data. As shown in the figures, the average standard scores of most schools that were in the top 20 percent on the KCCT in 2010 were lower on the KCCT in 2011. However, the average standard score decrease of the highest-poverty schools was much greater when the same

students took the unfamiliar ITBS test in the same year than it was when the same students took the KCCT test in a subsequent year. In reading, the standard score of highest-performing schools in which more than 80 percent of students were living in poverty (10 schools) decreased, on average, by nearly 0.5 standard units.^e In math, the standard score of highest-performing schools in which more than 80 percent of students were living in poverty (14 schools) decreased, on average, by 0.37 standard units.^f

Figure 2.E
Average Standard Score Change
KCCT Reading 2010 To KCCT Reading 2011
KCCT Reading 2010 To ITBS Reading 2010
Schools Performing In Top 20 Percent On KCCT Reading 2010
By Percentage Of Students Eligible For Free Or Reduced-Price Lunch



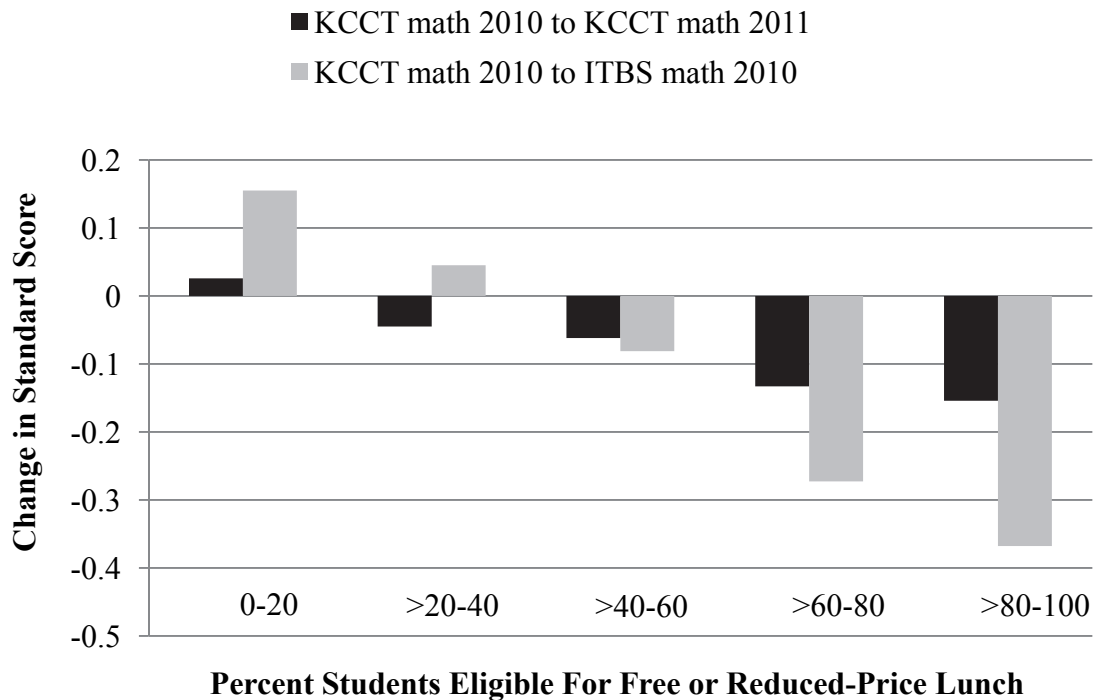
Note: The number of schools in lowest- and highest-poverty categories is small. For example, there were only 10 schools with more than 80 percent of students eligible for free or reduced-price lunch. Total numbers of schools in each category are shown in Appendix D

Source: Staff analysis of data from the Kentucky Department of Education.

^e This group of schools dropped an average of 44 percentile points from one test to another.

^f This group of schools dropped an average of 32 percentile points in math.

Figure 2.F
Average Standard Score Change
KCCT Math 2010 To KCCT Math 2011
KCCT Math 2010 To ITBS Math 2010
Schools Performing In Top 20 Percent On KCCT Math 2010
By Percentage Of Students Eligible For Free Or Reduced-Price Lunch



Note: The number of schools in lowest- and highest-poverty categories is small. For example, there were only 14 schools with more than 80 percent of students eligible for free or reduced-price lunch. Total numbers of schools in each category are shown in Appendix D.

Source: Staff analysis of data from the Kentucky Department of Education.

Appendix D shows the average KCCT and ITBS scores along with the total number of schools in each poverty category shown on the figure. It also shows similar patterns of rank change in higher-poverty higher-performing schools from the KCCT reading and math tests in 2011 to the K-PREP reading and math tests in 2012.

On average, high-performing higher-poverty schools that dropped in rank were still performing above their higher-poverty peers on unfamiliar tests.

On Average, High-Performing High-Poverty Schools That Drop In Rank Still Relatively High Performing. Results reported in Figures 2.E and 2.F show that, on average, high-poverty schools that were high performing on the KCCT had standard rank scores that were considerably lower on ITBS tests taken by the same students in the same year. While these schools dropped in rank relative to all schools, they were still performing, on average, far above their high-poverty peers on both the ITBS reading and the ITBS math tests. For example, high-performing

(top 20 percent on the KCCT) schools in which more than 80 percent of students were living in poverty performed, on average, slightly above the mean for all schools on the ITBS reading and math tests. The average standard score for all schools in the state in this poverty group was -0.39 in reading and -0.34 in math, well below average. These data are shown in Appendix D.

School Rank Change Based On Percentage Of Minority Students

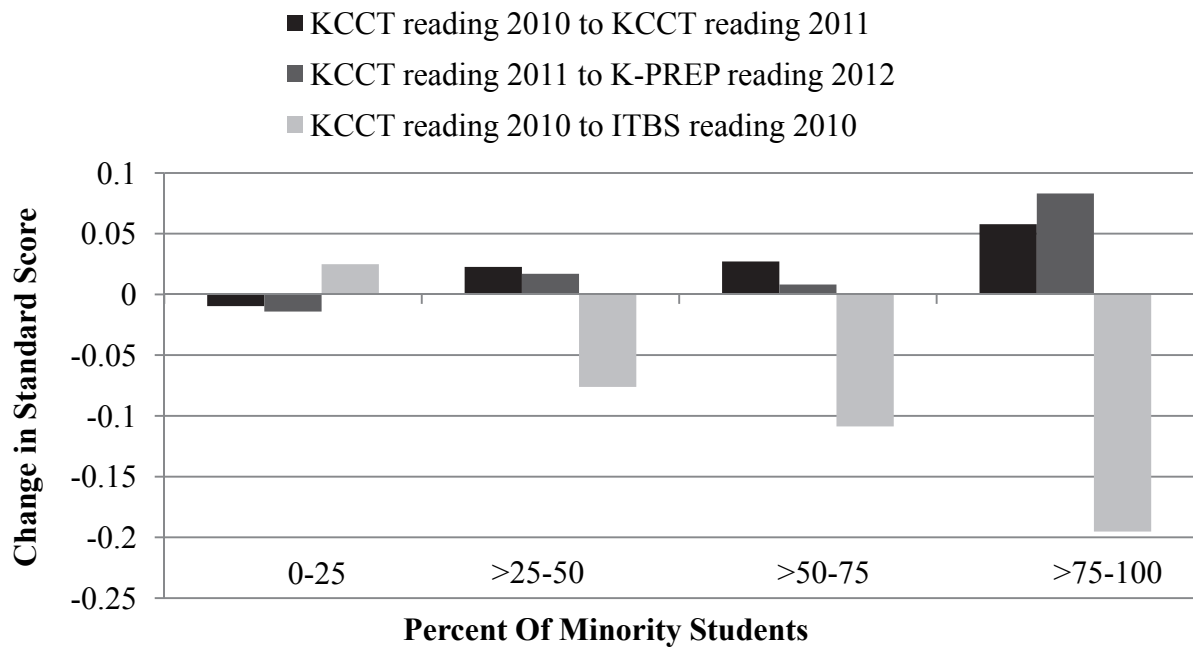
Changes in standard scores associated with the percentage of minority students in a school did not follow consistent patterns among different tests.

Changes in standard scores associated with the percentage of minority students in a school did not follow consistent patterns. The standard scores of schools with the highest percentages of minority students increased, on average, on the K-PREP reading and math tests and did not change substantially from the KCCT to ITBS math tests. However, the drop in standard scores from the KCCT reading test in 2010 to the ITBS reading test was substantial in schools with high percentages of minority students.

Standard scores of schools serving high percentages of minority students dropped substantially when students took an unfamiliar reading test composed of multiple-choice questions only.

Figure 2.G shows average standard score changes in reading from KCCT tests in 2010 to KCCT tests in 2011 and from KCCT tests to ITBS and K-PREP tests for schools based on percentages of minority students. Standard scores in schools with higher percentages of minority students increased, on average, from the KCCT 2010 to the KCCT 2011 reading tests and the KCCT 2011 to the K-PREP 2012 reading tests. However, standard scores decreased, on average, from the KCCT 2010 reading test to the ITBS 2010 reading test by nearly 0.2 in schools in which 75 percent of students or more were minorities. Unlike the KCCT and the K-PREP, the ITBS reading test contains only multiple-choice questions.

Figure 2.G
Average Standard Score Change
KCCT Reading 2010 To ITBS Reading 2010
KCCT Reading 2011 To K-PREP Reading 2012
By Percentage Of Minority Students



Source: Staff analysis of Kentucky Department of Education data.

Factors That Might Explain Change In Standard Scores

Chapter 3 explores potential links between instructional practices and score discrepancies. Many other factors could also explain these discrepancies.

Data presented in this chapter indicate some areas of potential concern associated with changes in school standard scores when students take unfamiliar tests, especially in higher-poverty schools. Research presented in Chapter 1 hypothesizes links between test-focused instructional practices and differences between state test scores and scores on unfamiliar tests. Researchers have also suggested that test-focused instructional practices are likely to be more evident in higher-poverty schools because these schools are more likely to be under pressure to improve test scores. Chapter 3 explores potential links between instructional practices and score discrepancies. Many other factors could also explain score discrepancies. Some of these factors are discussed below.

Test Design

The ITBS and K-PREP tests captured the performance of students whose full abilities were not captured on the KCCT.

Range Of Performance. The ITBS and K-PREP tests captured the performance of students whose full abilities were not captured on the KCCT. It is possible that the standard scores of schools with high percentages of very high-performing students increased on the ITBS and K-PREP tests because these tests were reflecting high performance that was not fully captured on the KCCT. It is also possible that the standard scores of schools with high percentages of very low-performing students decreased on the K-PREP and ITBS tests because these tests captured low performance that was not captured on the KCCT.

The KCCT did not capture the full range of student performance. The test ceiling did not extend beyond the skills deemed necessary to be considered proficient or distinguished. In some years, more than 10 percent of students attained the highest score possible on KCCT tests in some grades. Up to 3 percent of students in some grades also attained the lowest score possible. In contrast, only 1 percent of students typically attained the top or bottom score on the ITBS. The K-PREP also captured a broader range of performance because it included more difficult material. In 2012, less than 1 percent of students in any grade attained the highest possible score on the K-PREP. If the test ceiling affected score changes from the KCCT to other tests in some schools more than others, then it could have influenced some of the results reported in this chapter.

Staff recalculated analyses reported in this chapter using data sets that excluded students who had attained the highest and lowest scores possible on the KCCT. Appendix E provides data showing performance patterns similar to those reported in this chapter, even when very high- and low-performing students were removed from the data.

Some researchers have raised concerns that multiple-choice reading questions use words that are less familiar to minority students than to white students.

Concerns About Possible Bias In Multiple-Choice Reading Tests. Data reported in this chapter showed that the standard score of schools with high percentages of minority students dropped by nearly 0.2 from the KCCT to ITBS reading tests but did not drop from the KCCT to the K-PREP tests.

The lower performance of minority students on ITBS versus KCCT and K-PREP tests might be explained by differences in the types of questions used on the tests. Unlike the KCCT and the K-PREP, the ITBS comprises entirely multiple-choice questions. Some researchers have raised concerns that multiple-choice

reading questions use words in ways that are less familiar to minority students than to white students.²⁹ This concern has not been studied widely, and researchers have yet to reach consensus on questions of bias in multiple choice questions.

District And School Curricula

Changes in school scores from the KCCT to K-PREP tests may have been affected by resources available to schools to support teaching of the new standards.

Changes in school scores from the KCCT to K-PREP tests could have been affected by differences among schools in the speed with which curricula and instructional resources were adopted to teach the new standards assessed in 2012. It is possible that some schools began this process earlier than others. Districts with greater resources may have been better able to purchase instructional materials and provide teachers with professional development and support to teach the new standards assessed by the K-PREP.

Out-Of-School Factors

It is also possible that changes in test scores were affected by students' opportunities to learn outside of school.

It is also possible that patterns in standard score change reported in this chapter are explained in part by out-of-school factors. Students in lower-poverty schools may have greater access to out-of-school learning opportunities than do most students. Some of the content assessed on unfamiliar tests could be content that students have learned outside of school. Students in highest-poverty schools may have fewer opportunities than other students to learn after school or in the summer. These students would then be at relative disadvantage when taking unfamiliar tests.

Comparing Changes In KCCT Scores With Changes In National Assessment Of Educational Progress Scores

The National Assessment of Educational Progress is administered to a sample of Kentucky students every 2 years.

The National Assessment of Educational Progress is administered by the US Department of Education to gauge academic performance of students across the nation over time. Since 2003, NAEP has been administered biennially in reading and math to a small sample of 4th- and 8th-grade students in all states. The test is likely unfamiliar to students and educators and is not included in states' accountability systems. For this reason, NAEP scores are less likely to be influenced by test-focused instructional practices than are state test scores.

Kentucky students have made steady progress on NAEP reading and math tests.

As reported in OEA's Compendium of State Rankings, NAEP data indicate steady progress of Kentucky students in both reading and math. Kentucky 4th-graders outscore the nation in both reading and

math. Kentucky 8th-graders outscore the nation in reading and are close to the national average in math.³⁰

For this report, NAEP data are useful for comparing the degree of progress on state KCCT scores with the degree of progress on an unfamiliar test of similar content.

For the purposes of this report, NAEP data are useful as a means of comparing the degree of progress as indicated by state KCCT scores with the degree of progress indicated by an unfamiliar test of similar content.

KCCT trends should not necessarily be expected to align exactly with NAEP trends. NAEP does not assess all of Kentucky's standards. Also, differences in the sample of students who were tested may affect scores. KCCT assessed all students, whereas NAEP assessed a representative sample.

Table 2.2 shows years in which the NAEP and the KCCT reading and math tests were administered in Kentucky in the same grades. Both KCCT and NAEP were administered in 2007, 2009, and 2011. Comparisons between NAEP and KCCT can be made as far back as 2000 in 8th-grade math and 2002 in 4th-grade reading. Changes to the KCCT in 2006, including changes to the minimum and maximum possible scores and the "cut" scores used to designate performance labels of proficient and distinguished on the KCCT, preclude comparisons of KCCT and NAEP score changes before and after 2006. This chapter compares reading and math scores since 2007. KCCT and NAEP scores prior to 2006 are compared in Appendix F.

Table 2.2
Years And Grades In Which Kentucky Students Took Both
The KCCT And The NAEP In Reading Or Math
2000-2011

	Grade	2000	2002	2003	2005	Changes to range of minimum and maximum scores on KCCT	2007	2009	2011
Math	4						X	X	X
	8	X		X	X	X	X	X	
Reading	4		X	X	X	X	X	X	
	8					X	X	X	

Source: Staff analysis of data from the Kentucky Department of Education and US Department of Education.

Methodology

Staff compared changes in KCCT and NAEP scores by converting changes in each into standard units that could be compared. This was done on each test by dividing the change in scale score from 2007 to 2011 by the standard deviation from the mean on that test in 2007. Scale scores and standard deviations for each year are shown in Appendix F.

Results

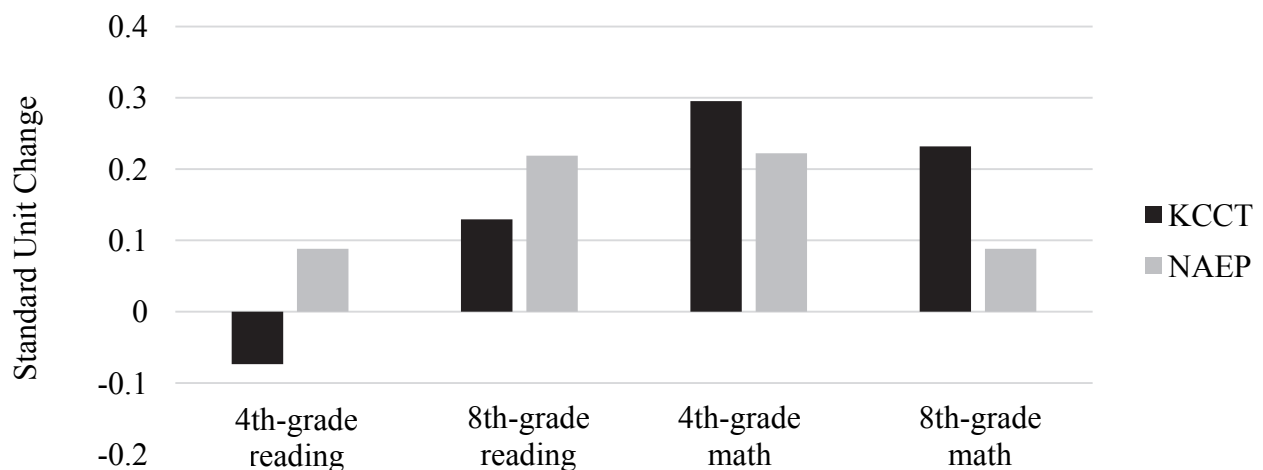
The scores of Kentucky students increased **more** on the unfamiliar NAEP reading test than they did on the familiar KCCT reading test.

Figure 2.H summarizes standard unit changes on KCCT and NAEP 4th- and 8th-grade reading and math tests between 2007 and 2011. NAEP reading gains exceeded KCCT gains between 2007 and 2011. While KCCT 4th-grade reading scores dropped slightly by 0.07 standard units, they increased by 0.09 standard units on NAEP. While KCCT scores increased by 0.13 standard units in 8th-grade reading, NAEP scores increased by 0.22 standard units.

Kentucky students have made gains on both KCCT and NAEP math tests, but their gains were greater on KCCT tests than on NAEP tests. Student KCCT gains in 8th-grade math were more than 2½ times as great as on NAEP.

Figure 2.H shows Kentucky students making math gains in the 4th and 8th grade on both the KCCT and the NAEP between 2007 and 2011. However, 4th-grade KCCT math gains of 0.3 were slightly greater than NAEP gains of 0.22 and 8th-grade KCCT math gains of 0.23 were more than 2½ times as great as NAEP gains of 0.09.

Figure 2.H
Standard Unit Changes
KCCT And NAEP 4th- And 8th-Grade Reading And Math Tests
2007-2011



Note: Standard unit changes were computed by dividing the difference in scale scores between 2007 and 2011 by the standard deviation of each test in 2007. Appendix F shows the KCCT and NAEP scale scores in 2007, 2009, and 2011.

Source: Staff analysis of data from the US Department of Education and the Kentucky Department of Education.

One factor that might explain differences between KCCT and NAEP trends in reading and math is differences in the difficulty of KCCT reading and math tests. Kentucky reading tests contained a higher percentage of questions considered to require the greatest depth of knowledge than did math tests.

One factor that might explain differences between KCCT and NAEP trends in reading and math is differences in the difficulty of KCCT reading and math tests. As shown in Appendix A, a higher percentage of test questions were taken from questions considered to measure the greatest depth of knowledge on reading tests than on math tests. It is possible that KCCT gains reflected increases in students' mastery of some standards but not necessarily those requiring the greatest depth of knowledge.

Chapter 3

Assessment Practices In Site Visit Schools

This chapter provides some preliminary evidence that discrepancies in schools' scores on state versus other tests might be explained in part by schools' focus on content likely to be tested versus all content necessary for subject mastery.

If schools are to monitor student learning of all state standards—as opposed to those that are likely to be tested—they must sometimes use assessment formats other than those used on state standardized tests.

OEA visited eight elementary schools to determine what kinds of assessments were being used to monitor student learning in tested subjects. Schools that performed relatively worse on unfamiliar tests than they had on the familiar KCCT assessed students primarily in KCCT-like formats. Schools that performed relatively better on unfamiliar tests used a broader range of assessments.

As shown in Chapter 2, some schools perform much better relative to their peers on familiar state tests than they do on other standardized tests in the same subject areas. Education researchers have suggested that these kinds of discrepancies might be explained by schools' focus on content likely to appear on the state test versus all content necessary for subject mastery. This chapter provides some preliminary evidence that this could be true.

Assessments used by schools and teachers to monitor student learning throughout the year provide some indication of what students are likely to be taught. If schools are to monitor student learning of all state standards—as opposed to those that are likely to be tested—they must sometimes use assessment formats other than those used on state standardized tests. Some standards are difficult or impossible to assess using standardized test formats.

OEA visited eight elementary schools to determine what kinds of assessments were being used to monitor student learning in tested subjects. Site visit schools were chosen from a larger sample of schools that performed much better or worse on the K-PREP and ITBS tests than they had on the KCCT. Schools that decreased in rank (performed relatively worse on unfamiliar tests than they had on the familiar KCCT) assessed students primarily using formats similar to the KCCT. Schools that increased in rank used a broader range of assessments.

Methodology

Staff chose site visit schools based on differences between how they performed relative to other elementary schools in reading and math on the familiar KCCT versus the unfamiliar ITBS and K-PREP. Of the eight site visit schools, four had performed relatively worse on unfamiliar tests in each year from 2010 to 2012 and four had performed relatively better on unfamiliar tests in each of those years.[§] Site visit schools were chosen to represent both rural and urban locations and a variety of student poverty levels.

[§] At the time that site visit schools were chosen, schools' performance was calculated using standard scores derived from the percent of students proficient and distinguished on the KCCT and K-PREP and the average national percentile rank on the ITBS.

Appendix G provides standard scores and demographic data for each site visit school.

Data collected from each site visit school include interviews, surveys, and examples of locally given assessments. Staff interviewed 13 administrators, 48 classroom teachers, and 37 exceptional child or special area teachers.

Data collected from each site visit school included interviews, surveys, and examples of locally given assessments. Each site visit consisted of three separate 1-hour interviews with the following groups of educators: administrators; classroom teachers in grades 3 through 5 who had been teaching in the building since at least 2010; and exceptional child or special area teachers who had been teaching in the building since at least 2010. Staff interviewed 13 administrators, 48 classroom teachers, and 37 exceptional child or special area teachers. Of the teachers interviewed, 34 worked in schools whose rank had decreased and 51 in schools whose rank had increased.

Site visit data instruments were designed to determine the range of assessments used at each school and to elicit administrators' and teachers' views on prioritizing of tested versus untested or less tested standards. Staff conducted semi-structured interviews with teachers and administrators and also asked teachers to complete a short survey. The survey asked teachers to report frequency of using KCCT-like formats and test preparation practices and to assess the advantages and disadvantages of test performance pressure. Finally, staff requested that each school submit at least one example of each type of locally given assessment believed to be important in guiding teaching and learning at the school.

Because site visit data sought to understand school rank change from the KCCT to other tests, teachers were asked to describe instruction and assessment as it existed in the school in 2010 and 2011. While standards and specific assessments changed in 2012, teachers in most schools did not report major shifts in school instructional practices. Appendix H includes interview protocols and the teacher survey.

Use Of The Terms *Data And Assessment* In Site Visit Interviews

Teachers were asked about use of data from a variety of sources, not just assessments in standardized formats.

The term *data* is often used in education to refer to numbers from standardized tests. When asking interview questions about data and assessments, staff encouraged educators to consider many forms of data, including data taken from informal assessments such as student work or observations of students during class.

Limitations

Because of the small sample of schools visited, the fact that the sample did not include middle and high schools, and the limited amount of data collected in each school, findings should be interpreted with caution.

Because of the small sample of schools visited, the fact that the sample did not include middle and high schools, and the limited amount of data collected in each school, findings should be interpreted with caution. Practices documented in site visit schools are not necessarily representative of practices in schools across the state. Site visits were exploratory in nature and can only provide preliminary data about variation in assessments used among Kentucky schools.

Summary Of Site Visit Findings

Schools that performed relatively better on KCCT than on unfamiliar tests were more likely to make frequent use of state test formats. Schools that performed even better on unfamiliar tests than they had on the KCCT were more likely to use and value a broad range of assessments.

Site visit findings are summarized in Table 3.1. The table shows how site visit schools used and valued data from assessments in state test formats and data from other sources. Schools that performed relatively better on KCCT than on unfamiliar tests were more likely to make frequent use of state test formats. Schools that performed even better on unfamiliar tests than they had on the KCCT were more likely to use and value a broad range of assessments.

The KCCT exerted a strong influence on the format of assessments used in most schools.

The table shows the strong influence of the KCCT on assessments used in site visit schools. In most schools, unit tests and interim tests—commercial tests designed to track student progress several times throughout the year—were conducted primarily or exclusively in the two question formats used on the KCCT: multiple choice and open response. Multiple-choice questions require students to choose among four predetermined answers. Open-response questions require students to write, justify, and explain their answers.

KCCT question formats were also used once a week or more during instruction.

KCCT question formats were also used once a week or more during instruction. In schools that decreased in rank, students were more likely to answer questions in multiple-choice formats daily, participate in test-practice activities, and receive additional instruction on skills assessed on the state test.

Systematic assessment in formats other than those used on state tests was reported more often in schools that increased in rank. Teachers and administrators in these schools were also more likely to value data from these other assessments.

Teachers in all schools valued many forms of data; however, systematic assessment in formats other than those used on state tests was reported more often in schools that increased in rank. For example, diagnostic assessments that required observation of individual students were used to identify priority needs of struggling students in most schools that increased in rank. Teachers and administrators in the schools that increased in rank were more

likely to consider data from assessments other than those in state test formats critical in determining whether students had truly mastered reading and math standards. Also, ongoing learning of students who had already mastered grade-level standards was assessed more systematically in schools that increased in rank.

**Table 3.1
Use/Value Of Data In State Test And Other Formats
In Eight Site Visit Elementary Schools
By Schools That Increased Or Decreased In Rank**

Use/Value of Data in State Test Formats	Schools
Unit tests are primarily or exclusively in KCCT-like formats	Most
Interim tests in multiple-choice formats given several times a year	Most
Students answer questions in state test formats once a week or more	Most
Students answer multiple-choice questions daily	↓
Students receive additional instruction in KCCT-tested skills not yet mastered*	↓
Students take timed KCCT-like tests*	↓
Students learn test-taking strategies*	↓
Students receive awards/rewards for meeting test targets*	↓
Principals value data in state test formats most in determining whether students have mastered standards	↓
Use/Value of Data Other Than State-Tested Formats/Content	
Teachers use and value assessment formats other than those used on KCCT (student work, short answer, fill in the blank, etc.)	All
Assessments that capture skills not tested on KCCT used systematically (for example, observational assessments; transfer tasks)	↑
Teachers attend to KCCT-like data but value other types of data most in determining whether students have truly mastered standards**	↑
Principals value and encourage use of assessments that capture skills not tested on KCCT	↑
Accelerated learning of students who have mastered state-tested skills monitored as frequently as learning of students who have not**	↑
Key	
↓	More common in schools that decreased in rank.
↑	More common in schools that increased in rank.

Notes: *These practices were also reported in most schools that increased in rank, but they were more common in schools that decreased in rank.

**These practices were also reported in one school that decreased in rank. This school still performed far above state averages, even after its rank decreased.

Source: Staff analysis of site visit data.

Assessments In State Test Formats

This section reports site visit schools' use of assessments that were similar in format to those used on state tests. Practices common in most site visit schools are reported first, followed by practices more common in schools that scored relatively higher on the KCCT than they did on ITBS and K-PREP tests (schools that decreased in rank).

Practices Common In Site Visit Schools

Seven out of eight schools designed tests given to assess student learning at the end of instructional units to resemble the KCCT and six out of eight schools used a commercial multiple-choice test several times a year to track student progress in reading and math.

Practices described in Table 3.2 illustrate the strong influence of KCCT formats on formal assessment in most site visit schools. Seven out of eight schools designed tests given to assess student learning at the end of instructional units to resemble the KCCT. Six out of eight schools used a commercial multiple-choice test several times a year to track student progress in reading and math.

Table 3.2
Use Of Standardized Test Formats For Formal Assessment
In Eight Site Visit Elementary Schools
That Decreased Or Increased In Rank

Assessment Practice	Number of Schools	
	Decreased in Rank	Increased in Rank
Unit tests composed primarily or exclusively of questions in KCCT question formats	4	3
Interim test in multiple-choice format used to track student progress throughout the year	3	3

Source: Staff analysis of site visit interview and document data in eight elementary schools.

Unit Tests In KCCT-Like Formats. KDE recommends that standards-based instructional units begin with the state learning standards and end with unit tests that assess the primary goals of these units.³¹ In all but one site visit school, the unit tests that assessed these instructional goals comprised primarily or exclusively multiple-choice or open-response questions. In constructing the tests, teachers either wrote questions themselves or took questions from other sources including data banks of items that had previously been used on the KCCT. Many unit tests were constructed to look like the KCCT (or now K-PREP), with the test logo at the top of the paper and instructions relating to use of number 2 pencils, writing between the lines, or avoiding stray pencil marks. In at least three schools, unit tests were designed to

have the same ratio of multiple-choice to open-response questions found on the KCCT.

Interim Tests. All but two site visit schools monitored student progress several times a year using a commercial test that tracked student learning of state standards through multiple-choice items. These tests, which are commonly called “interim” tests, were prized largely for their ability to determine whether students were making progress or required intervention, to identify KCCT-tested skills not yet mastered, and to predict student performance on the KCCT. Interim tests were used to group students for instruction, identify students for additional support, and identify specific skills that should be emphasized in the classroom or with individual students.

Two schools tracked student learning throughout the year using a test that was not in KCCT formats. The interim test used in these schools included a fluency test that required students to read passages out loud and answer questions. The math portion required students to write their own answers. Both of the schools were in the same district. One had increased in rank, and the other had decreased in rank but was still performing far above state averages.

Because unit tests are intended to assess primary instructional goals, it seems likely that these goals were more likely to be those that could be tested in KCCT-like formats.

Influence Of Unit And Interim Test On Content Taught. The extent to which teachers taught KCCT-tested versus untested or less tested standards in reading and math cannot be determined from the data collected for this study. However, unit and interim tests provide a strong indication of content most likely to be taught. Unit tests reflect primary instructional goals and interim tests identify specific students and skills for extra attention. In most site visit schools, unit and interim tests were primarily or exclusively in KCCT-like formats. In these schools, content that is not easily assessed with KCCT-like formats would be less likely to be included in primary instructional goals. In the absence of other forms of systematic assessment, student learning of content not assessed on the KCCT might be monitored and retaught less consistently than KCCT-tested content.

Teachers and administrators in seven of eight site visit schools reported that instruction in science and social studies was reduced or eliminated in untested grades.

While science and social studies were not a focus of site visit data collection, the allocation of teaching time in these subjects indicates the clear influence of the KCCT test itself on the state curriculum content that is actually taught. The Program of Studies describes learning goals for science and social studies in all elementary grades but these subjects were only tested by the KCCT (and are only tested by the K-PREP) in two elementary grades: science in the 4th grade and social studies in the 5th grade.

Teachers and administrators in seven of eight site visit schools reported that instruction in science and social studies was reduced or eliminated in untested grades.

The overwhelming majority of classroom teachers interviewed for this study reported that students answered multiple-choice questions and open-response questions at least once a week. More than half of these teachers reported that students rewrote open-response questions using scoring rubrics at least once a week.

Ongoing Use Of State Test Formats. Survey data indicate regular use of KCCT question formats in most schools. The overwhelming majority of classroom teachers—about 80 percent—reported that students answered multiple-choice and open-response questions once a week or more. About half of classroom teachers (52 percent) reported that students rewrote open-response questions using scoring rubrics once a week or more. Scoring rubrics are documents that describe specific characteristics associated with low (0) and high (4) scores on open-response questions. While Kentucky’s general scoring rubric focuses on broad skills, such as providing complete answers that demonstrate in-depth understanding, some of the rubrics used in schools required students to answer questions in prescribed formats that specified the sequence in which different parts of the question were addressed and the specific number of sentences or examples associated with each answer component.

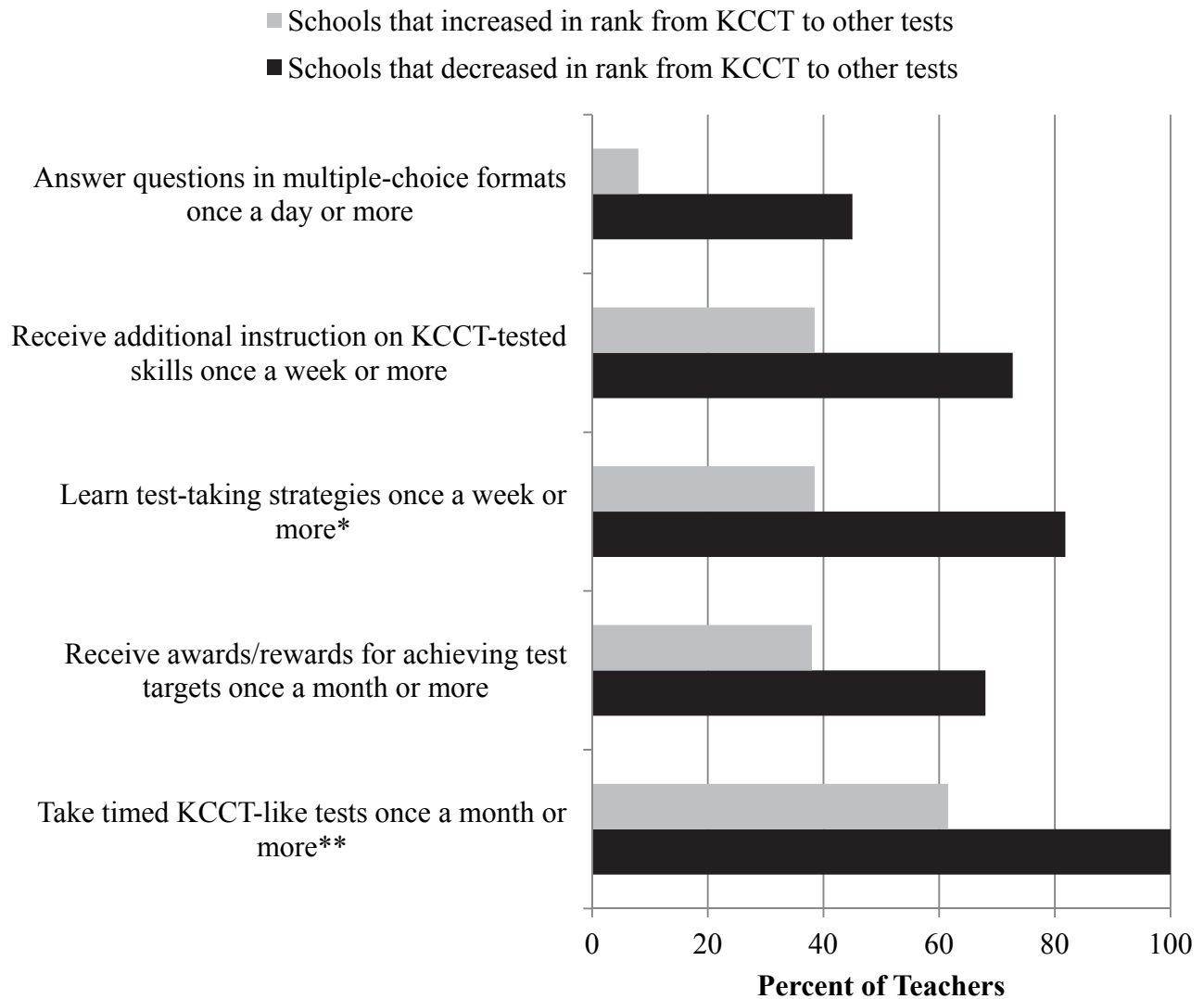
Practices More Common In Schools That Decreased In Rank From KCCT To Other Tests

While assessments in formats similar to the state test played a prominent role in most site visit schools, the frequency with which these formats were used and the variety of school practices influenced by standardized test data were greater in schools that decreased in rank from the KCCT to other tests.

While assessments in formats similar to the state test played a prominent role in most site visit schools, the frequency with which these formats were used and the variety of school practices influenced by standardized test data were greater in schools that decreased in rank from the KCCT to other tests.

Figure 3.A shows the frequency of KCCT-influenced instructional practices in schools as reported by classroom teachers. Teachers in schools that decreased in rank were more likely to report that students answered questions in multiple-choice formats, received additional instruction on KCCT-tested skills, learned test-taking strategies, and took timed KCCT-like tests than were teachers in schools that increased in rank. These practices are described in greater detail following the figure.

Figure 3.A
Percentage Of Classroom Teachers Reporting
KCCT-Influenced Instructional Practices
Schools That Decreased Or Increased In Rank



Notes: Included in percentages of teachers reporting practices as described in the figure were those that reported these practices even more often.

*The percentage of teachers reporting that students learn test-taking strategies once a day or more was 36 in schools that decreased in rank and 8 in schools that increased in rank.

**The percentage of teachers reporting that students take timed KCCT-like tests once a week or more was 45 percent in schools that decreased in rank and 19 percent in schools that increased in rank.

Source: Staff analysis of surveys administered to 48 classroom teachers in eight site visit schools.

Forty-five percent of classroom teachers in schools that decreased in rank reported that students answered questions in multiple-choice formats at least once a day compared to only 8 percent of classroom teachers in schools that increased in rank.

Ongoing Use Of Multiple-Choice Questions. As reported earlier, most classroom teachers used multiple-choice questions at least once a week. Teachers in schools that decreased in rank from the KCCT to subsequent were more likely to report use of these questions daily. Forty-five percent of teachers in schools that decreased in rank reported that students answered multiple-choice questions once a day or more compared to only 8 percent of classroom teachers in schools that increased in rank.

Use Of KCCT-Like Assessments To Identify Specific Skills Emphasized In Support For Struggling Students. All site visit schools reported ways in which struggling students were given extra support in reading and math. This support was most commonly provided in small group settings during regular instruction. Several schools also made additional support available during specifically designated support periods or by withdrawing children from other activities. The content of the additional support provided to students varied among schools.

All schools reported additional support for struggling students. Schools that decreased in rank focused more on KCCT-tested skills.

In schools that decreased in rank from KCCT to unfamiliar tests, extra support was more likely to be focused on specific KCCT-tested skills not yet mastered. One school had a checklist of KCCT-tested skills that should be mastered by all students at the end of the year. Another school kept spreadsheets of KCCT-tested skills mastered by each student in each instructional unit and used the spreadsheet to identify skills for reteaching. Seventy-three percent of teachers in schools that decreased in rank from KCCT to other tests reported that students received additional instruction in KCCT-tested skills once a week or more, compared to 38 percent in schools that performed relatively better on unfamiliar tests.

Teachers in schools that decreased in rank were more likely to report teaching of test-taking strategies, timed KCCT-like tests, and awards or rewards given to students for meeting test targets.

Instructional Practices Specific To Standardized Tests. Teachers in most schools reported use of practices designed specifically to prepare students for standardized tests, but teachers in schools that decreased in rank were more likely to report frequent use. Eighty-three percent of classroom teachers interviewed in schools that decreased in rank reported weekly instruction in test-taking strategies, compared to 39 percent of teachers in schools that increased in rank. In schools that decreased in rank, 100 percent of teachers reported that students took timed KCCT-like tests once a month or more, compared to 61 percent of teachers in schools that increased in rank. In schools that decreased in rank, 68 percent of teachers reported that students received awards or rewards for meeting test targets once a month or more, compared to 38 percent of teachers in schools that increased in rank.

Researchers and educators have raised concerns about practices such as teaching of test-taking strategies or frequent administration of practice tests because these practices may be more likely to raise scores on standardized tests than to improve student mastery of content standards.

Administrators in all four schools that decreased in rank valued data in standardized formats more than any other kind of data; that was the case with only one administrator in a school that increased in rank.

Some principals acknowledged that they focused more on tested than untested content. One explained that state tests are “what we are graded on.”

The majority of teachers interviewed for this study agreed with the statement “In this district student scores on standardized tests are the indicator most valued by administrators in judging the quality of my teaching.”

Researchers and educators have raised concerns about practices such as teaching of test-taking strategies or frequent administration of practice tests because these practices may be more likely to raise scores on standardized tests than to improve student mastery of content standards. The Administration Code for Kentucky’s Educational Assessment Program, which is incorporated by reference in 703 KAR 5:080, does not consider these practices inappropriate, as long as they can be shown to include content or identify content areas that should be addressed during instruction.

Administrators’ Focus On Data In Standardized Test Formats.

It is possible that principals’ beliefs about standardized test data influenced the degree to which teachers in site visit schools used standardized test formats and engaged in test preparation activities. When asked which kinds of data they found most valuable in determining whether the school’s students had mastered reading and math standards, administrators in all four schools that decreased in rank cited data in standardized test formats, compared to only one administrator in the schools that increased in rank. Some administrators expressed active suspicion of data not in standardized formats, believing that these data were not objective and could not reveal whether students were really learning.

Principals in schools that performed relatively better on the KCCT than on unfamiliar tests felt a strong sense of accountability to ensure that their schools were successful on the KCCT. Some acknowledged that this required schools to deemphasize untested learning goals. As one principal explained, state tests are “what we are graded on.”

The majority of teachers interviewed for this study agreed with the statement “In this district student scores on standardized tests are the indicator most valued by administrators in judging the quality of my teaching.” Teachers in schools that decreased in rank were more likely to agree with this statement (63 percent) than were teachers in schools that increased in rank (53 percent).

Assessments In Formats Other Than Those Used On State Tests

This section reports site visit schools’ use of assessments in formats other than those used on state tests. Practices common in most site visit schools are reported first, followed by practices more common in schools that scored relatively higher on K-PREP and ITBS tests than they did on the KCCT (schools that increased in rank).

Teachers and administrators in schools that increased in rank from the KCCT to other tests were more likely to value and systematically use data taken from assessments other than those in state test formats.

As shown in Table 3.3, teachers in all schools used data from assessments that were not in state test formats, but teachers and administrators in schools that increased in rank were more likely to value and systematically use these data.

Table 3.3
Use Of Assessments Not In State Test Formats
In Eight Site Visit Elementary Schools
Schools That Decreased Or Increased In Rank

Assessment Beliefs or Practices	Number of Schools	
	Decreased in rank	Increased in rank
Teachers value and use a variety of assessment formats	4	4
Teachers attend to KCCT-like data but value other types of data most in determining whether students have truly mastered standards	1*	3
Administrators value and encourage use of assessments that capture skills not tested on KCCT	0	3
Struggling students supported through specifically designed program that included diagnostic, observational assessments	1*	3
Projects, transfer tasks, or other performance tasks a regular feature of assessment	0	1
Accelerated learning of students who had already mastered grade-level content monitored as frequently as learning of students who had not mastered grade-level content	1*	3

*This was the only site visit school that decreased in rank from the KCCT to other tests but still performed well above state averages.

Source: Staff analysis of data from site visit interviews.

Practices Common in All Schools

Teachers in all schools valued data from assessments not in standardized formats.

Classroom Assessment In A Variety Of Formats. When asked what kinds of data they found most valuable in determining whether students had mastered reading and math standards, teachers in all schools answered that they valued data from a variety of assessments. In addition to data from questions in state test formats, teachers mentioned short-answer formats, fill-in-the-blank formats, and class discussions as important sources of information about student learning. Though not asked to do so by the interview protocol, teachers in all schools cited state standards that were important in their school though not assessed in state test formats. Examples included reading with expression and without error, mental computation, ability to identify the broader purpose and use of skills tested, and ability to choose and read books for pleasure.

Teachers in all schools expressed a strong professional commitment to fully understand learning goals described in state standards and to provide students with the range of learning experiences necessary to achieve those goals.

Teachers in all schools expressed a strong professional commitment to fully understand learning goals described in state standards and to provide students with the range of learning experiences necessary to achieve those goals. When asked what they needed most to teach Kentucky’s new content standards, teachers mostly requested time and instructional materials. Teachers wanted more time to meet with each other or to do their own research and planning. Teachers also wanted instructional materials aligned with new learning goals. Teachers described intensive efforts individually and with colleagues to understand and prepare to teach the state’s new content standards. They rarely requested test preparation materials or examples of test questions.

Practices More Common In Schools That Increased In Rank

Strong Value Of Assessments Not In State Test Formats.

Teachers and administrators in schools that increased in rank on unfamiliar tests were more likely to value and systematically use data from assessments in formats other than those used on state tests.

Teachers in four schools—three that had increased in rank and one that had decreased in rank—explained that, although they valued and paid attention to data from KCCT-like assessments, they did not believe that a student had truly mastered the standards until they observed that the student could demonstrate knowledge in a variety of ways.

Teachers in four schools—three that had increased in rank and one that had decreased—explained that, although they valued and paid attention to data from KCCT-like assessments, they did not believe that a student had truly mastered the standards until they observed that the student could demonstrate knowledge in a variety of ways such as classroom discussions, class work, or the way that the student read texts or explained answers. In addition to student work and class discussions, teachers in these schools used a variety of classroom assessments to collect data about student learning. Examples of these assessments are observational checklists of student skills or learning processes, “learning inventory” assessments that described what kinds of instructional strategies or content motivated different students to learn, and classroom assessments that asked students to demonstrate their knowledge by making concept “maps” showing how important ideas relate to each other.

Teachers in several schools that had increased in rank stated that students who transferred into their school from other schools often did not meet the school’s standards of mastery even when those students had received a proficient score on the KCCT at a previous school. They explained that the standards for proficiency in their school went beyond what was required to be proficient on state tests and what could be determined in standardized formats alone.

Administrators in most schools that increased in rank valued and encouraged use of assessments that could capture student learning of standards not tested on the KCCT.

Administrators in three schools that increased in rank stated explicitly that some skills are important to teach even though these skills are not directly tested. Examples of these skills included reading fluency, mental computation skills in math, and students' ability to verbally explain key concepts they were learning. These administrators also valued and encouraged use of assessments that could capture student learning of standards not tested on the KCCT. Several principals valued diagnostic assessments of struggling students. One principal valued student writing journals and encouraged teachers to share and analyze student work in these journals.

Four schools (three that increased in rank and one that decreased in rank) made regular use of observational, diagnostic assessments to determine which underlying skills or learning processes to emphasize in additional support for struggling students.

Observational/Diagnostic Assessments. Four schools (three that increased in rank and one that decreased in rank) made regular use of observational, diagnostic assessments to determine which underlying skills or learning processes to emphasize in additional support for struggling students. These diagnostic assessments were generally administered by a teacher sitting with a single student or a small group of students and watching how the child approached tasks or problems. In math, three schools used a diagnostic assessment obtained through the Kentucky Center for Mathematics, a state-funded program that trains teachers in diagnostic assessment and intervention for struggling math students. The diagnostic assessment focused on such areas as counting forward and backward, and decomposing numbers into groups of 5 and 10. The principal in one of these schools explained that while a multiple-choice question could tell you that a student had not mastered a particular skill, it could not tell you **why**. The fourth school used a program that helped struggling students to visualize and represent number quantities. The diagnostic reading assessment used in one school required teachers to observe a child read and discuss a book using skills such as predicting, using prior knowledge, and using picture or context clues. Diagnostic reading assessments used in another school helped struggling students to connect sounds as they are heard with how they are spelled and to develop comprehension skills by listening to as well as reading stories.

Only one school included extended tasks or projects as a regular feature of the assessment system.

Transfer Tasks. Only one school included extended tasks or projects as a regular feature of its assessment system. As a part of the unit test in this school, students completed transfer tasks that required them to use skills in situations that simulated real life. For example, students learning how to calculate elapsed time were asked to design a field day schedule and present it to the class. Students were given a list of events, duration of events, and characteristics of the location in which the event would take place.

Students would have to determine which skills to use in completing the task. Teachers and administrators in this school attributed their students' success in reading and math primarily to the school's focus on learning processes and student thinking. Students were given daily open-ended tasks that required them to make connections and reach conclusions in unfamiliar situations, without direct prompting but with assistance or coaching from teachers.

While all eight site visit schools provided ongoing assessment of and support for struggling students, only four provided systematic opportunities for accelerated students to progress beyond content assessed on grade-level state tests.

Assessment Of Accelerated Learners. While all eight site visit schools provided ongoing assessment of and support for struggling students, only four provided systematic opportunities for accelerated students to progress beyond content assessed on grade-level state tests.^h Two of eight site visit schools—both of which increased in rank—provided daily teacher-led instruction that addressed learning goals beyond those tested in a student's assigned grade. In one school, students identified as gifted were instructed in reading and math in a separate class that moved beyond grade level standards. In another school, accelerated learning opportunities, such as assignments that went beyond grade level content or challenging projects, were incorporated into planning of every instructional unit. Students likely to master grade-level content quickly were grouped in classrooms that were taught or assisted by gifted teachers.

In two other schools—one that had increased in rank and one that had decreased in rank (but was still well above the state average) — students spent 30 minutes daily on self-paced computer software that provided instruction and assessment in reading and math. Using this software, students who had mastered grade-level content had daily opportunities to learn additional skills.

Advantages And Disadvantages Of Test Performance Pressure

This chapter focuses on unintended, negative consequences of test performance pressure on the validity of state test data. Test performance pressure also has intended, positive consequences.

Data presented thus far in this chapter focus on the unintended, negative consequences of test performance pressure. The chapter suggests that the validity of state test data in some schools may be affected when educators focus heavily on tested versus untested content. However, accountability associated with state tests also has intended, positive consequences. This section describes

^h All schools provided some additional instruction for accelerated students either in pull-out classes or in extra work assigned during the regular day. However, this additional instruction was more likely to be teacher led, daily, and systematic in some schools than in others.

advantages and disadvantages of test performance pressure as reported by educators in site visit schools.

Advantages Of Test Performance Pressure

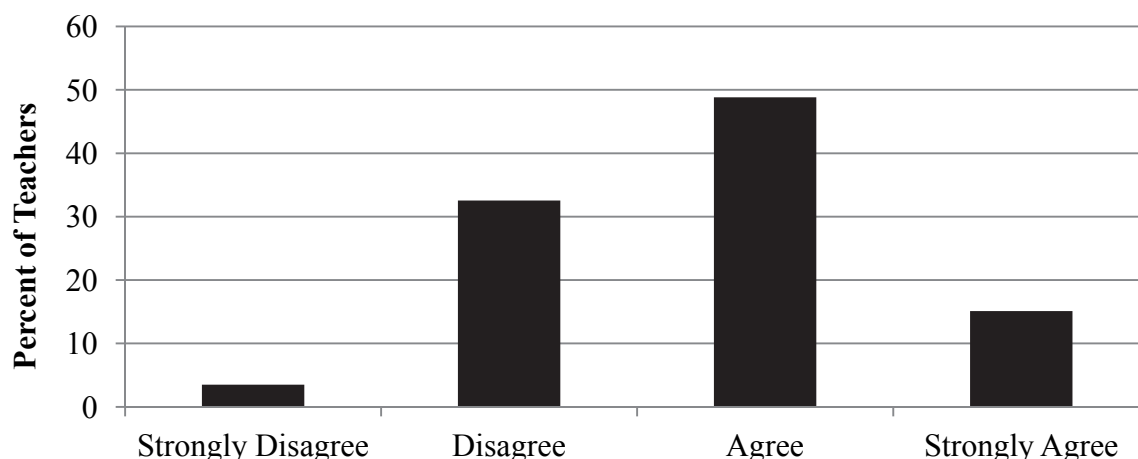
The majority of teachers interviewed for this study agreed that, overall, test performance pressure had improved the quality of teaching and learning in their schools.

As shown by Figure 3.B, the majority of teachers (64 percent) interviewed for this study agreed that, overall, test performance pressure had improved the quality of teaching and learning in their schools, with 15 percent strongly agreeing. When asked to explain their school's success in helping students to master reading and math standards, many teachers mentioned intentional planning of instruction and assessment to align with standards, citing an individual and collective sense of responsibility to ensure that students were actually learning rather than just being taught.

Some teachers and administrators specifically credited test-based accountability for motivating students and teachers to work harder.

Teachers and administrators in two schools specifically credited accountability for performance on state tests for motivating teachers and students to work harder. Teachers in these schools wanted to demonstrate to students and to the community at large that students from low poverty-backgrounds could be successful in school. One teacher explained that what students "have been hearing about themselves all their life is a lie," meaning that students had been given messages that they were inferior to children from wealthier backgrounds while they could in fact achieve at high levels. These teachers also credited test performance pressure for combating the sense of inertia that can exist when teachers become overwhelmed by students' academic needs and personal challenges.

Figure 3.B
Percentage Of Teachers Who Agreed With The Statement
“Overall, the pressure to ensure that students performed well on the KCCT had a positive effect on the quality of teaching and learning in this school.”



Source: Staff analysis of surveys administered to a total of 85 teachers in eight site visit schools.

Disadvantages Of Test Performance Pressure

Some teachers had strong concerns about the negative effects of test-based accountability on teaching and learning. Teachers in one school felt that they no longer had time to teach some content that would make learning meaningful or relevant to their students.

As shown in Figure 3.B, 36 percent of teachers interviewed during site visits did not agree that test performance pressure had a positive effect on teaching and learning in their school. During interviews, teachers typically cited concerns about the sacrifice of important, untested content. For example, teachers in one school felt that they no longer had time to teach some content that would make learning meaningful or relevant to their students—such as writing of personal memoirs—because this content was not believed to align with the core content that was assessed. In this school, which had decreased in rank from the KCCT to other tests, teachers explained that they did not feel that they could really teach until after students had taken the state test.

Teachers were most likely to report negative consequences of test performance pressure on students receiving special education services and other students who were struggling to master grade-level content.

Teachers were most likely to report negative consequences of test performance pressure on students receiving special education services and other students who were struggling to master grade-level content. Teachers explained that often these students needed to master prerequisite skills in order to understand grade-level content but instead spent the majority of the day working on content that they were not able to understand thoroughly. Several teachers expressed concerns that struggling students might be taught to answer questions correctly but did not understand the material well enough to remember it. One teacher wondered whether struggling students forgot most of what they had learned

Teachers in four schools reported extreme stress associated with test performance pressure.

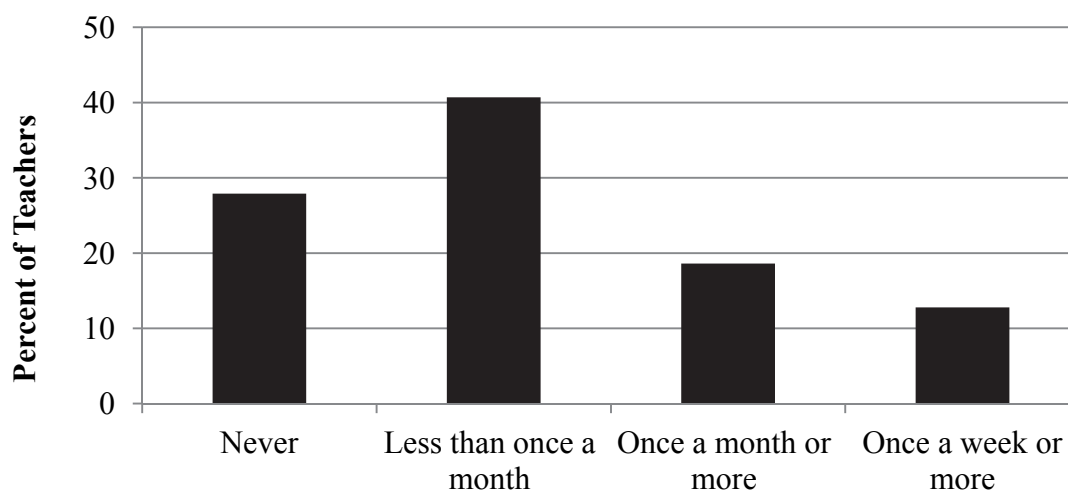
Most teachers reported that test performance pressure sometimes caused them to make decisions that were against their own standards of best practice.

within weeks after they took the state test. These concerns were expressed more often in schools that decreased in rank.

Although not specifically asked about personal stress, teachers in four schools—three that had decreased in rank and one that had increased—reported extreme stress associated with test performance pressure. Teachers in three schools reported that students cried because of test performance pressure. An administrator in one school reported that teachers cried because of performance pressure. Teachers in two schools that had increased in rank explained that they had previously worked in schools in which they felt there was a stressful and unproductive focus on test performance and that they had left the schools for that reason.

Teachers were asked how often, as a result of test performance pressure, they made an instructional decision that they did not feel was the best decision for a particular student. As shown by Figure 3.C, only 28 percent of teachers reported never making such decisions. Nineteen percent of teachers reported making these decisions once a month or more, and 13 percent reported once a week or more. Percentages of teachers reporting these decisions once a week or more were higher in schools that decreased in rank (20 percent) than in schools that increased in rank (8 percent).

Figure 3.C
Percentage Of Teachers Reporting That “The pressure to ensure that students performed well on the KCCT led me to make instructional decisions that were against my own standards of best practice.”*



*Prior to completing the survey, teachers were told that this question was asking them how often they made an instructional decision based on the need to prepare a student to perform well on state tests that they would not have otherwise felt was the best instructional decision for that student. Teachers were specifically told that this question was not asking whether they ever made decisions that they felt were unethical.

Source: Staff analysis of surveys administered to a total of 85 teachers in eight site visit schools.

Overall, relatively few teachers expressed serious concerns that they were sacrificing untested learning goals in order to prepare students for state tests.

It should be noted, however, that teachers and administrators were specifically asked during interviews whether they felt that they had to make sacrifices in teaching of untested learning goals to ensure that students performed well on state tests, and relatively few expressed serious concerns about these trade-offs. In most schools, teachers and administrators expressed commitment to ensuring that the arts, extracurricular activities, and character education remained prominent in the school's curriculum. In all but one site visit school, students received instruction in music, art, and physical education at least once a week. While teachers in particular schools were not always in complete consensus with each other, there was only one site visit school in which most teachers expressed strong concerns about the sacrifices that had been made to the quality of instruction in order to prepare students for state tests.

Conclusion

It is important that, to the extent possible, local assessments address the full range of skills and processes outlined in state standards.

KRS 158.6453 outlines a system of balanced assessments in which data about student learning that are not available on state tests are supplemented at the local level by interim or formative tests. Just as state tests affect local tests and, by extension, instructional practices, it is likely that these local tests will also affect instructional practices. It is important that, to the extent possible, these local assessments address the full range of skills and processes outlined in state standards.

Content And Format Of State Test Affects Local Tests

The content and format of state tests are likely to have a strong impact on the content and format of local assessments. It is important that instructional effects of assessments are considered when evaluating the advantages and disadvantages of different tests.

Data presented in this chapter suggest that the content and format of state tests used for accountability are likely to have a strong impact on the content and format of local assessments in most schools and, by extension, the content that is taught. By law, state curriculum goals are expressed in state standards, and state tests are adopted to assess those standards. In practice, the content that is assessed on state tests—including the proportion in which the content is sampled on tests and the format in which it is assessed—may influence what is taught in some schools as much as the standards themselves. For this reason, it is important that the instructional effects of assessments be taken into account when evaluating the advantages and disadvantages of different tests.

Local Assessments Vary And May Affect Validity Of State Test Data

Schools vary in their use of assessments capable of measuring student learning of all standards. This variation may affect the validity of state test data as it is used to judge schools' academic performance relative to each other.

Schools vary in their use of local assessments capable of measuring student learning of all standards. This variation may affect the validity of state test data as it is used to judge schools' academic performance relative to each other. Site visit schools that performed relatively better on KCCT than on other tests were more likely to rely on KCCT test formats to monitor student learning and to focus systematically on preparing students to be successful on the KCCT. It is possible that KCCT scores of students in these schools reflected students' knowledge of standards likely to be tested but not necessarily their knowledge of all standards or their ability to demonstrate knowledge in a variety of situations.

Site visit schools that performed relatively better on unfamiliar tests were more likely to have monitored student learning using a broader range of assessments. It is possible that KCCT data underestimated the academic strengths of students in these schools, relative to their peers.

Some Practices That Improve Scores On State Tests May Not Improve Student Mastery Of All Standards

Some of the practices used to promote high student achievement on the KCCT did not appear to promote similarly high achievement on the ITBS and K-PREP tests.

State test data are commonly used to identify schools and practices believed to be effective at promoting student learning of state standards. Data presented in this chapter suggest that some of the practices that promoted high reading and math achievement on the KCCT did not promote similarly high achievement on the ITBS and K-PREP tests. Schools that performed relatively better on the KCCT than on other tests were more likely to assess students in KCCT-like formats, teach test-taking skills, reward students for meeting test targets, and reteach KCCT-tested skills not yet mastered. These schools were less likely to spend time identifying and addressing weaknesses in students' foundational skills or systematically monitoring learning of accelerated students. While data collected for this study are not sufficient to indicate the relative advantages and disadvantages of particular practices, they do raise concerns about the use of state test data alone to identify best practices.

Recommendation 3.1

Recommendation 3.1

When adopting strategies believed to increase state test scores in a particular subject area, districts and schools should look for evidence that the strategies improve student learning of all state standards in that subject. Evidence should include data capable of measuring aspects of standards not assessed in standardized formats.

Appendix A

Differences In Percentages Of Greatest Depth Of Knowledge Questions On KCCT Reading And Math Tests

The KCCT comprised two types of questions: multiple-choice questions that required students to choose among predetermined answers, and open-response questions that required students to write their answers and justify their thinking. Standards determined to require the greatest depth of understanding were tested exclusively in open-response formats.

KCCT questions were assigned ratings of 1 to 3 to reflect the depth of knowledge (DOK) required to answer the question. Questions requiring the greatest depth of knowledge were assigned a 3. DOK 3 questions required open-response formats. Beginning in 2009, KDE set a goal that at least 50 percent of open-response questions should be DOK 3.³²

Table A.1 shows the percentage of DOK 3 open-response questions on 4th- and 8th-grade KCCT reading and math tests in 2009, 2010, and 2011. With the exception of 4th-grade reading in 2010, the percentage of open-response questions at DOK 3 on reading tests was close to or exceeded KDE's goal of 50 percent. In contrast, the goal was not met on any of the KCCT math tests. Thus, it appears that KCCT reading tests required students to show greater depth of knowledge than did math tests during these years.

**Table A.1
Distribution Of KCCT Open-Response Items Considered To Require
The Greatest Depth Of Knowledge (3)**

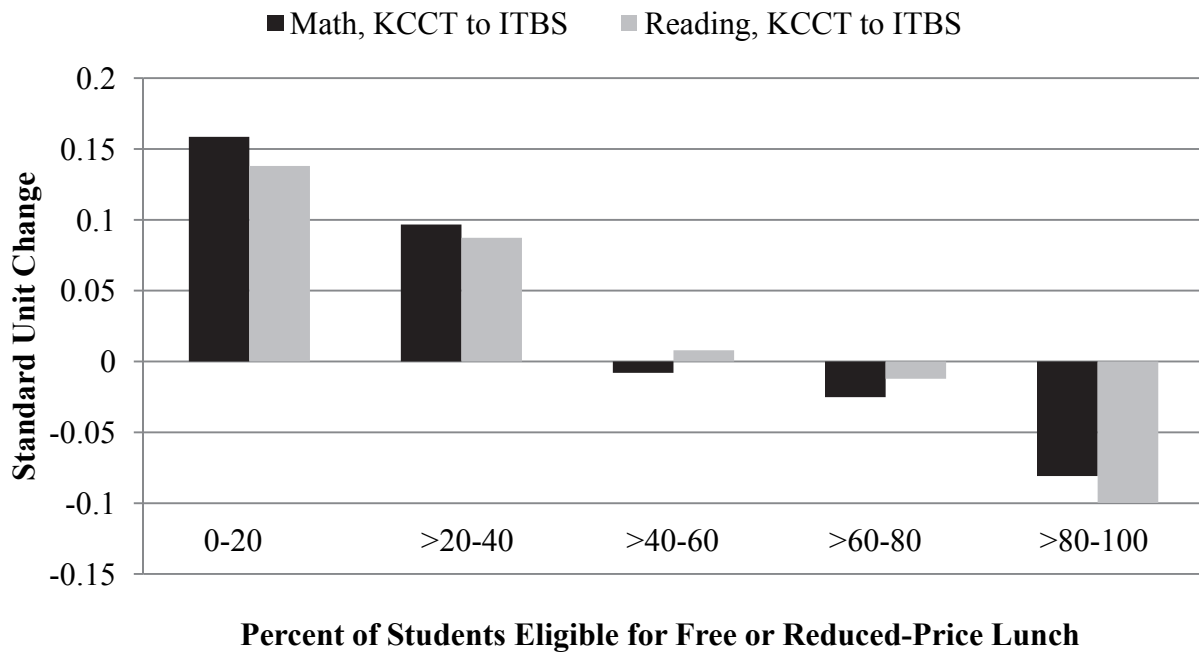
Year	Grade	Percent of Open-Response Items	
		Reading	Math
2009	4	44	25
	8	94	25
2010	4	6	22
	8	67	0
2011	4	44	25
	8	94	38

Source: Staff compilation based on data reported in Commonwealth Accountability Testing System technical manuals for 2009 (pp. 19-24), 2010 (pp. 18-23), and 2010-2011 (pp. 6-10).

Appendix B

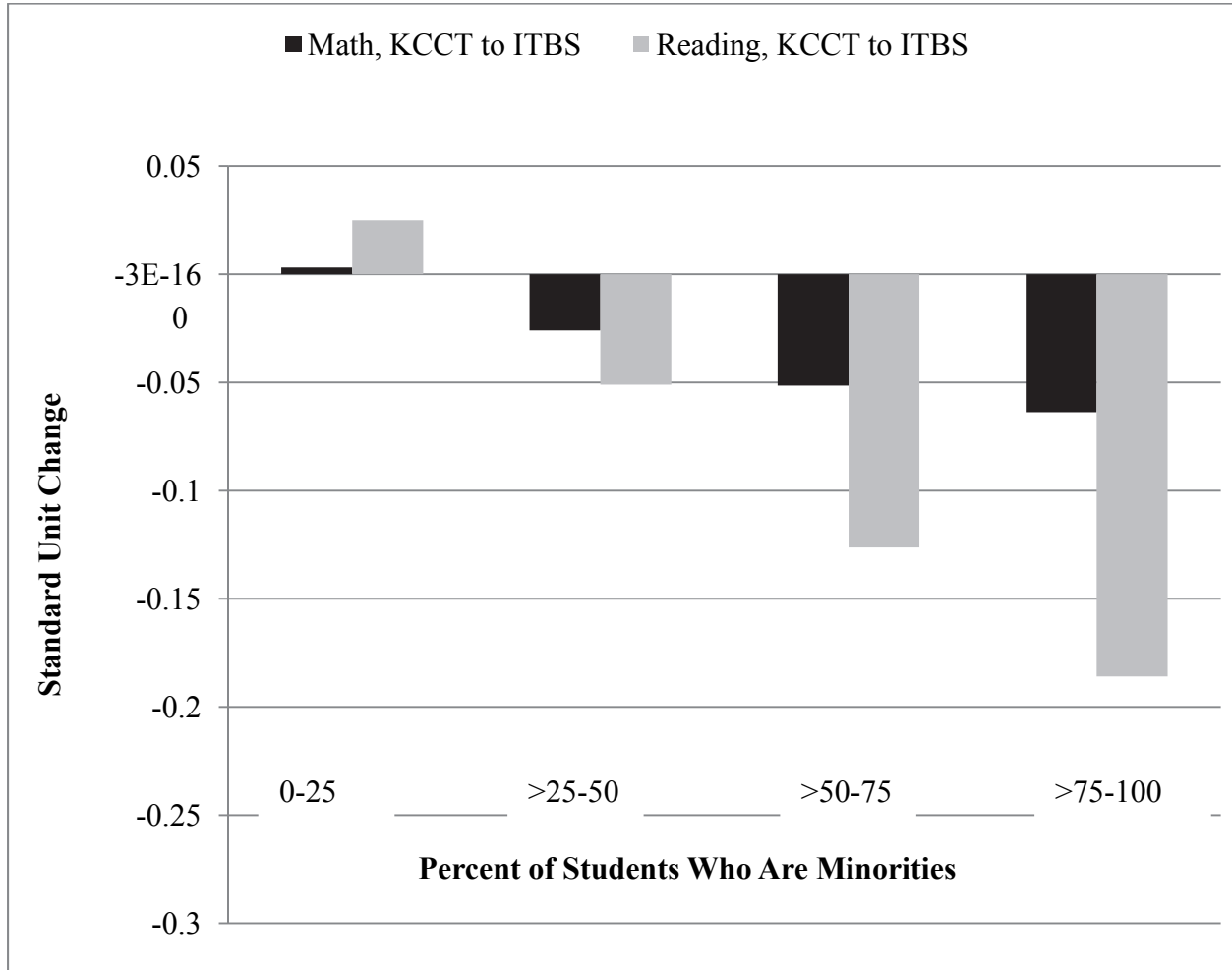
Comparing School Performance On KCCT And ITBS Tests In 2011

Figure B.A
Average School Standard Unit Change
KCCT Reading And Math Tests To ITBS Reading And Math Tests
By Percent Of Students Eligible For Free Or Reduced-Price Lunch
2011



Source: Staff analysis of data from the Kentucky Department of Education.

Figure B.B
Average School Standard Unit Change
KCCT Reading And Math Tests To ITBS Reading And Math Tests
By Percent Of Students Who Are Minorities
2011



Source: Staff analysis of data from the Kentucky Department of Education.

Table B.1
Average Standard Score Change
KCCT Reading 2011 To ITBS Reading 2011
Schools Performing In Top 20 Percent On KCCT Reading 2011
By Percent Of Students Eligible For Free Or Reduced-Price Lunch
And Average ITBS Reading Scores For All Schools
By Percent Of Students Eligible For Free Or Reduced-Price Lunch

	Only Schools in Top 20 Percent on KCCT Reading 2011				All Schools
Percent Eligible for Free or Reduced- Price Lunch	Number of Schools	KCCT Reading 2011	ITBS Reading 2011	Difference Between ITBS and KCCT Reading 2011	ITBS Reading 2011
0-20	27	0.49	0.60	0.11	0.54
>20-40	49	0.42	0.43	0.01	0.31
>40-60	67	0.39	0.23	-0.17	0.05
>60-80	39	0.40	0.22	-0.18	-0.09
>80-100	9	0.45	0.24	-0.22	-0.42

Note: Differences between ITBS and KCCT scores reflect rounding.

Source: Staff analysis of data from the Kentucky Department of Education.

Table B.2
Average Standard Score Change
KCCT Math 2011 To ITBS Math 2011
Schools Performing In Top 20 Percent On KCCT Math 2011
By Percent Of Students Eligible For Free Or Reduced-Price Lunch
And Average ITBS Math Scores For All Schools
By Percent Of Students Eligible For Free Or Reduced-Price Lunch

Percent Eligible for Free or Reduced-Price Lunch	Only Schools in Top 20 Percent on KCCT Math				All Schools
	Number of Schools	KCCT Math 2011	ITBS Math 2011	Difference Between ITBS and KCCT Math 2011	ITBS Math 2011
0-20	31	0.48	0.63	0.15	0.59
>20-40	49	0.37	0.45	0.07	0.31
>40-60	57	0.38	0.27	-0.11	0.03
>60-80	40	0.44	0.22	-0.22	-0.10
>80-100	14	0.44	0.09	-0.36	-0.37

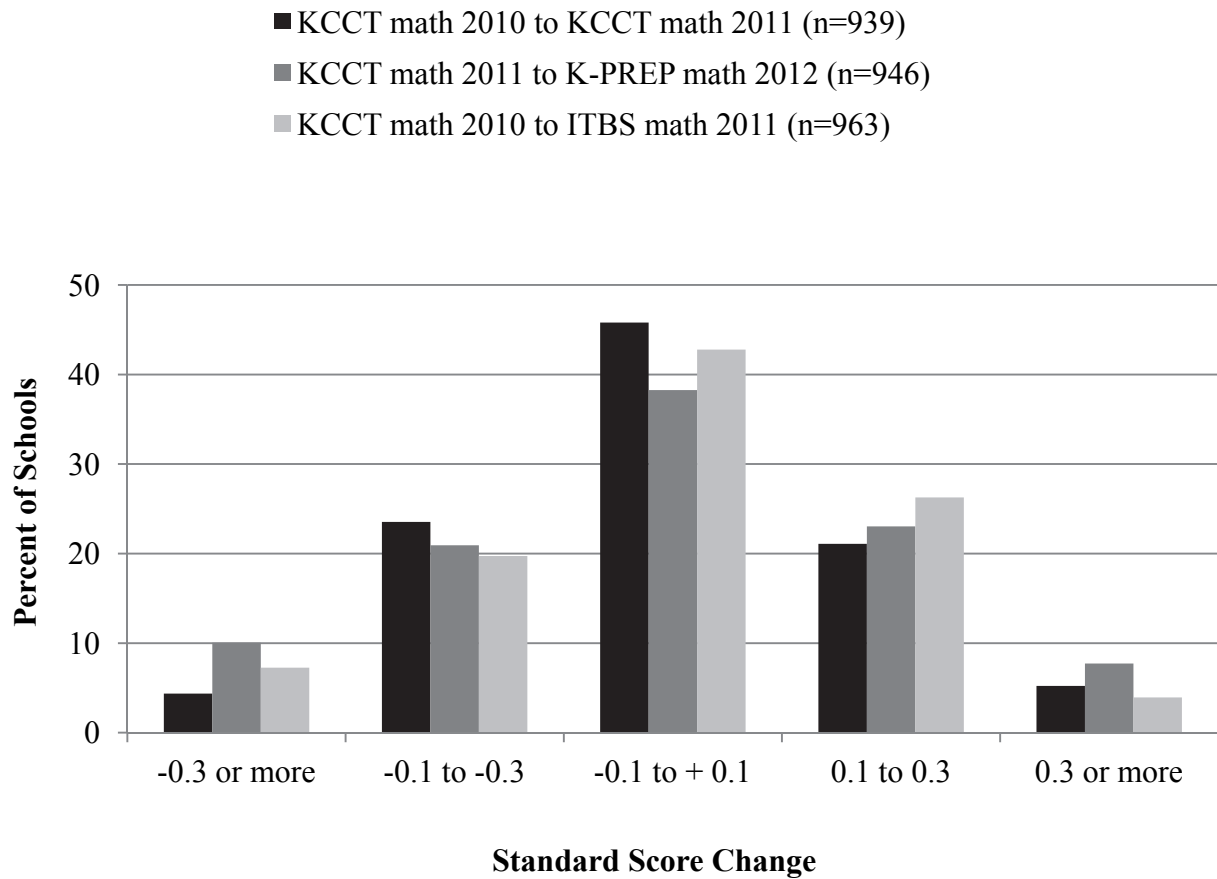
Note: Differences between ITBS and KCCT scores reflect rounding.

Source: Staff analysis of data from the Kentucky Department of Education.

Appendix C

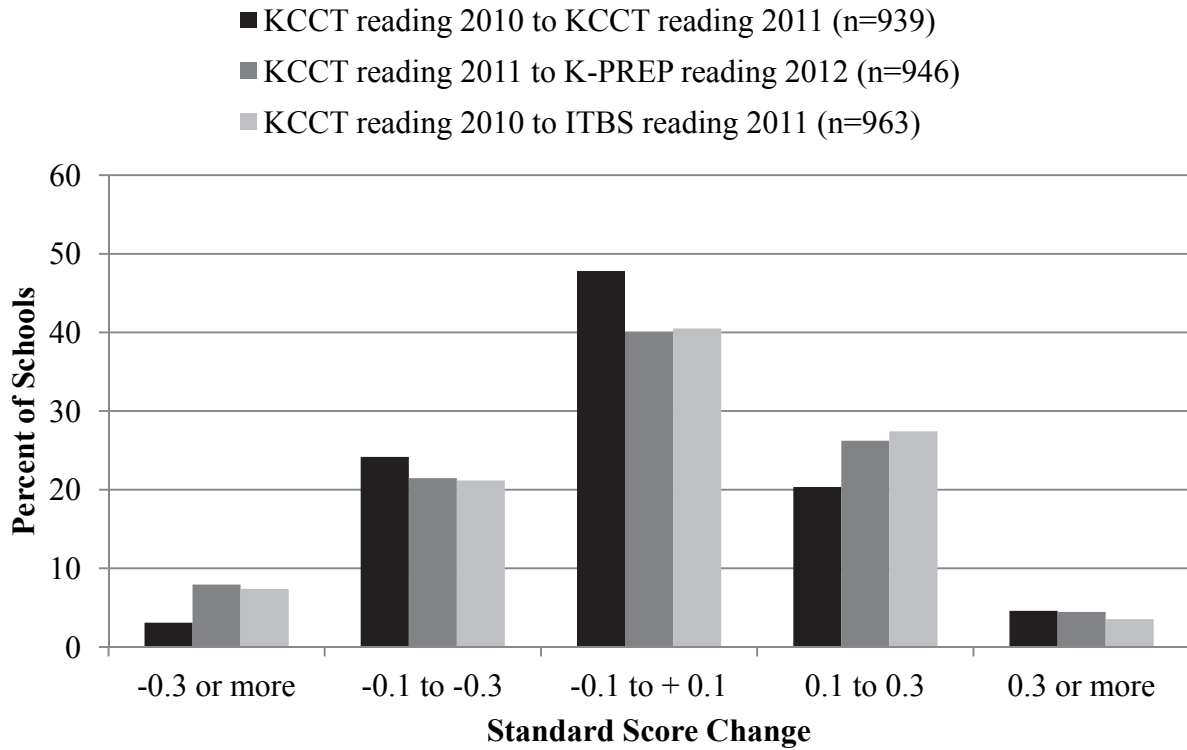
Changes In School Performance By Standard Score And Performance Decile

Figure C.A
Standard Score Change Of Elementary And Middle Schools
KCCT Math Tests To K-PREP And ITBS Math Tests
2010-2012



Source: Staff analysis of data from the Kentucky Department of Education.

Figure C.B
Standard Score Change Of Elementary And Middle Schools
KCCT Reading Tests To K-PREP And ITBS Reading Tests
2010-2012



Source: Staff analysis of data from the Kentucky Department of Education.

Table C.1
K-PREP 2012 Math Performance Decile Of Schools
In Top Decile (Top 10 Percent) Of Schools
On KCCT Math In 2011

K-PREP Math 2012 Decile	Number of Schools Out of 95	Percent of Schools	Average Percent of Students Eligible for Free or Reduced-Price Lunch
1	52	55	30
2	15	16	53
3	11	12	64
4	4	4	64
5	1	1	64
6	3	3	70
7	2	2	57
8	4	4	72
9	2	2	90
10	1	1	74

Source: Staff analysis of data from the Kentucky Department of Education.

Table C.2
K-PREP 2012 Reading Performance Decile Of Schools
In Top Decile (Top 10 Percent) Of Schools
On KCCT Reading In 2011

K-PREP Reading 2012 Decile	Number of Schools Out of 95	Percent of Schools	Average Percent of Students Eligible for Free or Reduced-Price Lunch
1	50	53	27
2	13	14	55
3	9	9	50
4	7	7	71
5	5	5	72
6	5	5	63
7	3	3	84
8	2	2	79
9	1	1	74

Source: Staff analysis of data from the Kentucky Department of Education.

Table C.3
K-PREP 2012 Math Performance Decile Of Schools
In 10th Decile (Bottom 10 Percent) Of Schools On
KCCT Math In 2011

K-PREP Math 2012 Decile	Number of Schools Out of 95	Percent of Schools	Average Percent of Students Eligible for Free or Reduced-Price Lunch
10	52	55	83
9	27	28	81
8	10	11	74
7	4	4	62
5	1	1	74
4	1	1	77

Source: Staff analysis of data from the Kentucky Department of Education.

Table C.4
K-PREP 2012 Reading Performance Decile Of Schools
In 10th Decile (Bottom 10 Percent) Of Schools
On KCCT Reading In 2011

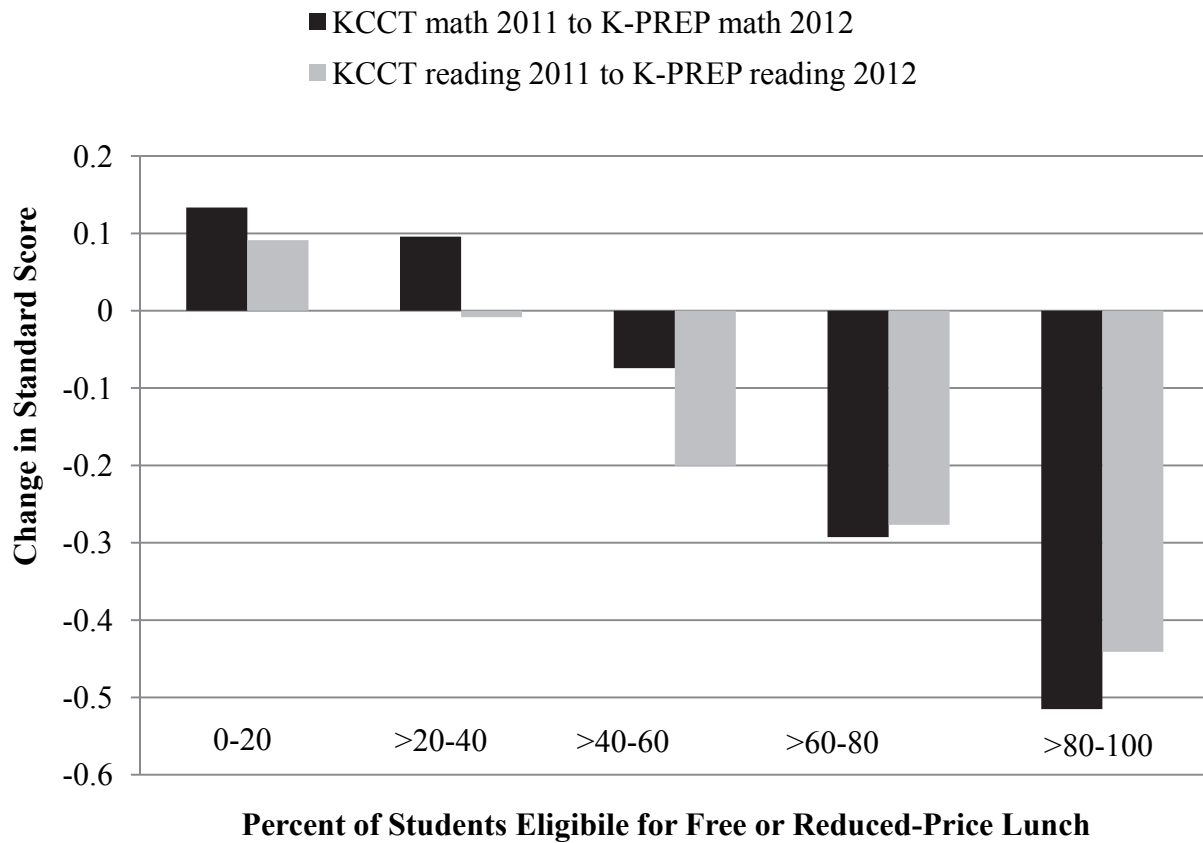
K-PREP Reading 2012 Decile	Number of Schools Out of 95	Percent of Schools	Average Percent of Students Eligible for Free or Reduced-Price Lunch
10	60	63	86
9	22	23	76
8	7	7	76
7	3	3	57
5	1	1	59
4	2	2	65

Source: Staff analysis of data from the Kentucky Department of Education.

Appendix D

Standard Score Changes Of Highest-Performing Schools

Figure D.A
Average Standard Score Change
KCCT Math 2011 To K-PREP Math 2012
KCCT Reading 2011 To K-PREP Reading 2012
Schools Performing In Top 20 Percent
On KCCT Math And KCCT Reading 2011
By Percent Of Students Eligible For Free Or Reduced-Price Lunch



Source: Staff Analysis of data from the Kentucky Department of Education

Table D.1
Average Standard Score Change
KCCT Reading 2010 To ITBS Reading 2010
Schools Performing In Top 20 Percent On KCCT Reading 2010
By Percent Of Students Eligible For Free Or Reduced-Price Lunch
And Average ITBS Reading Scores For All Schools
By Students Eligible For Free Or Reduced-Price Lunch

Percent of Students Eligible for Free or Reduced-Price Lunch	Only Schools in Top 20 Percent on KCCT Reading in 2010				All Schools
	Number of Schools	KCCT Reading	ITBS Reading	Difference Between ITBS and KCCT Reading	ITBS Reading
0-20	23	0.47	0.59	0.12	0.54
>20-40	50	0.43	0.42	-0.01	0.29
>40-60	79	0.41	0.23	-0.18	0.07
>60-80	31	0.41	0.16	-0.25	-0.10
>80-100	10	0.49	-0.02	-0.50	-0.39

Source: Staff analysis of Kentucky Department of Education data.

Table D.2
Average Standard Score Change
KCCT Math 2010 To ITBS Math 2010
Schools Performing In Top 20 Percent On KCCT Math 2010
By Percent Of Students Eligible For Free Or Reduced-Price Lunch
And Average ITBS Math Scores For All Schools
By Students Eligible For Free Or Reduced-Price Lunch

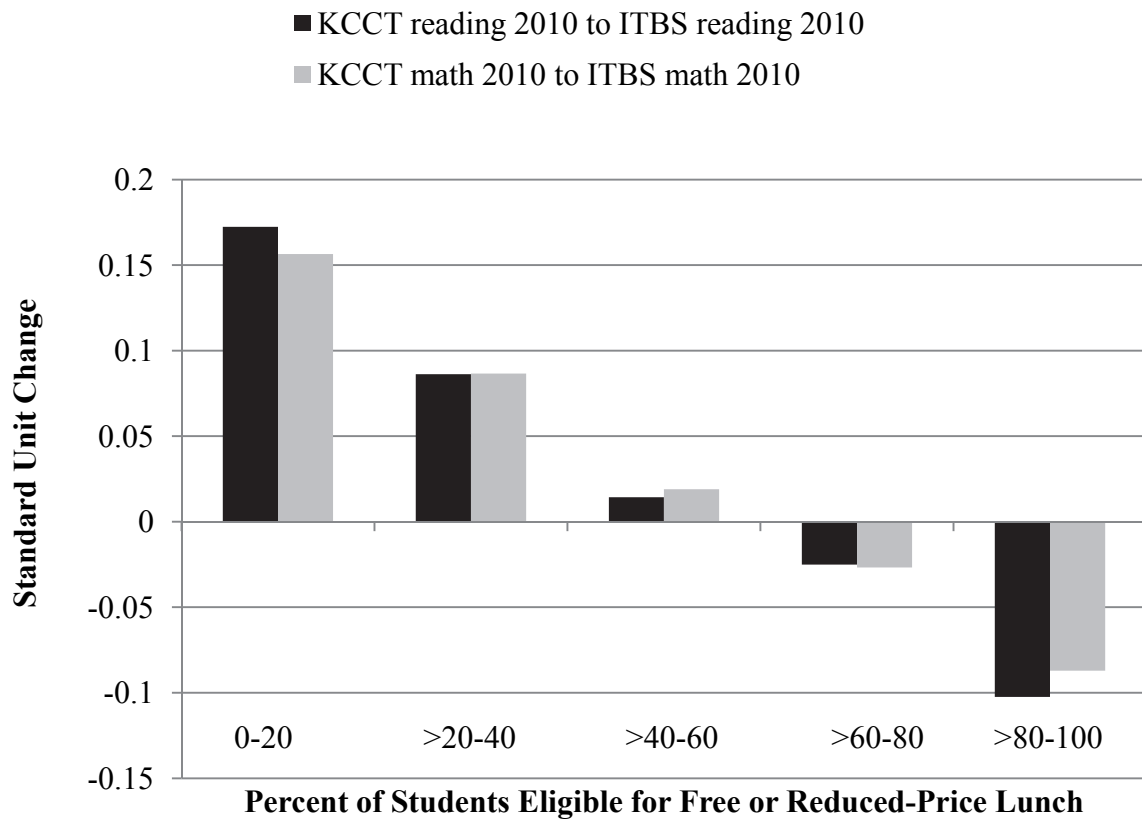
Percent of Students Eligible for Free or Reduced-Price Lunch	Only Schools in Top 20 Percent on KCCT Math in 2010				All Schools
	Number of Schools	KCCT Math	ITBS Math	Difference Between ITBS and KCCT Math	ITBS Math
0-20	26	0.49	0.64	0.16	0.59
>20-40	48	0.41	0.45	0.05	0.28
>40-60	58	0.40	0.32	-0.08	0.06
>60-80	47	0.44	0.17	-0.27	-0.11
>80-100	14	0.43	0.07	-0.37	-0.34

Source: Staff analysis of Kentucky Department of Education data.

Appendix E

Standard Score Change Excluding Highest- And Lowest-Performing Students

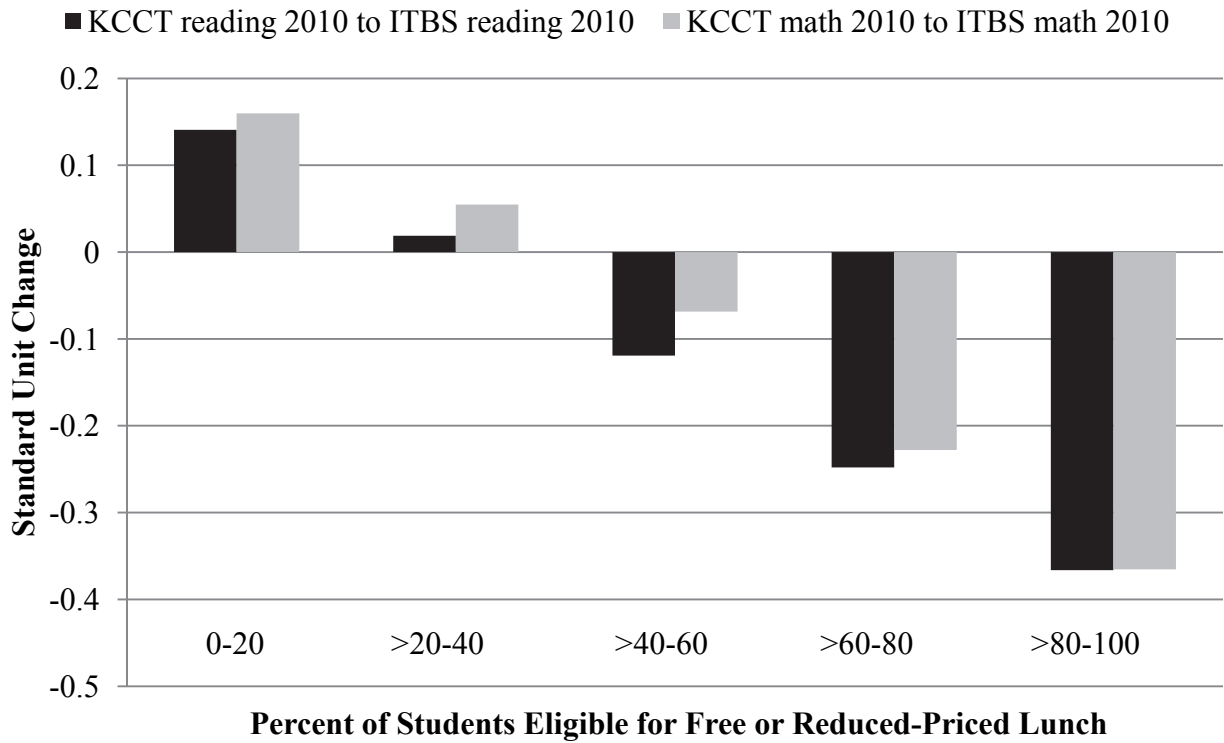
Figure E.A
Average School Standard Unit Change
KCCT Reading And Math Tests To ITBS Reading And Math Tests
Excluding Students With Maximum Or Minimum KCCT Scores*
By Percentage Of Students Eligible For Free Or Reduced-Price Lunch
2010



*This analysis excludes KCCT and ITBS scores for all students who attained a maximum score of 80 and a minimum score of 0 on the KCCT.

Source: Staff analysis of data from the Kentucky Department of Education.

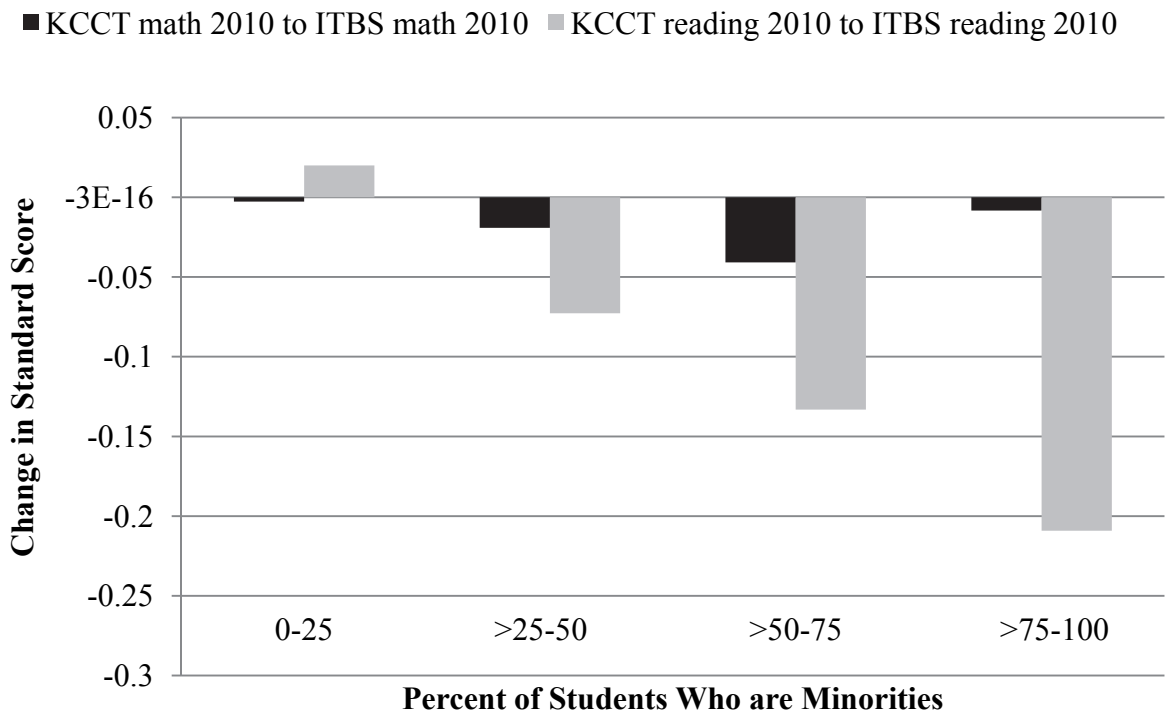
Figure E.B
Average School Standard Unit Change
KCCT Reading And Math Tests to ITBS Reading And Math Tests
Of Schools Performing In Top 20 Percent On KCCT Tests
Excluding Students With Maximum Or Minimum KCCT Scores*
By Percentage of Students Eligible For Free or Reduced-Price Lunch
2010



* This analysis excludes KCCT and ITBS scores for all students who attained a maximum score of 80 and a minimum score of 0 on the KCCT.

Source: Staff analysis of data from the Kentucky Department of Education.

Figure E.C
Average School Standard Unit Change
KCCT Reading And Math Tests To ITBS Reading And Math Tests
Excluding Students With Maximum Or Minimum KCCT Scores*
By Percentage Of Students Who Are Minorities
2010



* This analysis excludes KCCT and ITBS scores for all students who attained a maximum score of 80 and a minimum score of 0 on the KCCT.

Source: Staff analysis of data from the Kentucky Department of Education.

Appendix F

Changes In KCCT And NAEP Reading And Math Scores

Table F.1 shows KCCT and NAEP reading scale scores at the 4th- and 8th-grade levels in 2007, 2009, and 2011. The second-to-last column shows the change in scale scores from 2007 to 2011. The final column adjusts for the different scale scores used on each test by dividing each scale score gain by a unit considered common to each test—the standard deviation from the mean. This common unit is shown for 2007 in the first column of the table. The final column shows the change in scale scores from 2007 to 2011 as measured in this common unit, derived by dividing each scale score change 2007 to 2011 by the standard deviation in 2007.

The table shows 8th-grade reading gains on both the KCCT and the NAEP, with NAEP gains of 0.22 standard units outpacing gains of 0.13 standard units on the KCCT. At the 4th-grade level, KCCT scores dropped by 0.07, but scores increased by 0.09 on the NAEP. Thus, at both the 4th-and 8th- grade levels, NAEP gains exceeded KCCT gains in reading.

Table F.1
Change In Mean Scale Score And Standard Units
KCCT And NAEP 4th- And 8th-Grade Reading
2007-2011

Grade	Test	Standard Deviation	Mean Scale Score				Standard Unit Change
		2007	2007	2009	2011	2007-2011	2007-2011
8	KCCT	17	46.6	47.4	48.8	2.2	0.13
	NAEP	32	262.0	267.0	269.0	7.0	0.22
4	KCCT	19	52.1	52.2	50.7	-1.4	-0.07
	NAEP	34	222.0	226.0	225.0	3.0	0.09

Source: Staff analysis of data from the US Department of Education and the Kentucky Department of Education.

Table F.2 compares 4th-grade gains on the KCCT and the NAEP reading tests between 2000 and 2005. During these years, there were small gains on both the KCCT and the NAEP, with KCCT gains twice as large as NAEP gains.

Table F.2
Change In Mean Scale Score And Standard Units
KCCT And NAEP 4th-Grade Reading
2002-2005

Test	Standard Deviation	Mean Scale Score				Standard Unit Change
		2002	2003	2005	2002-2005	
KCCT	26.4	49.8	50.2	51.5	1.7	0.06
NAEP	33.0	219.0	219.0	220.0	1.0	0.03

Source: Staff analysis of data from the US Department of Education and the Kentucky Department of Education.

Table F.3 compares KCCT and NAEP gains in 4th- and 8th-grade math between 2007 and 2011. KCCT gains were greater than NAEP gains during these years. KCCT 4th-grade gains of 0.30 exceeded NAEP gains of 0.22. KCCT 8th-grade gains of 0.23 were more than 2½ times as great as NAEP 8th-grade gains of 0.09.

Table F.3
Change In Mean Scale Score And Standard Units
KCCT And NAEP 4th- And 8th-Grade Math
2007-2011

Grade	Test	Standard Deviation	Mean Scale Score				Standard Unit Change
			2007	2009	2011	2007-2011	
8	KCCT	22	39.4	42.3	44.5	5.1	0.23
	NAEP	33	279.0	27.0	282.0	3.0	0.09
4	KCCT	22	47.4	51.7	53.9	6.5	0.30
	NAEP	27	235.0	239.0	241.0	6.0	0.22

Source: Staff analysis of data from the US Department of Education and the Kentucky Department of Education.

Table F.4 shows 8th-grade KCCT and NAEP gains in math between 2000 and 2005. Kentucky students posted gains on both tests during these years, with NAEP gains of 0.11 greater than KCCT gains of 0.06.

Table F.4
Change In Mean Scale Score And Standard Units
KCCT And NAEP 8th-Grade Math
2000-2005

Test	Standard Deviation	Mean Scale Score				Standard Unit Change
	2000	2000	2003	2005	2000-2005	2000-2005
KCCT	27	46.9	47.9	48.5	1.6	0.06
NAEP	35	270.0	274.0	274.0	4.0	0.11

Source: Staff analysis of data from the US Department of Education and the Kentucky Department of Education.

Appendix G

Site Visit School Demographic And Test Data

Table G.1
Standard Scores On KCCT, ITBS And K-PREP Reading And Math Tests
And Demographic Data
OEA Site Visit Schools

	Schools That Decreased in Rank				Schools That Increased in Rank			
	A	B	C	D*	E	F	G	H
Average Percentage Students Eligible for Free or Reduced-Price Lunch 2010-2012	58	85	71	71	72	57	16	55
Average Percentage Minority Students 2010-2012	39	63	2	2	3	11	16	49
Average Standard Scores of Students Who Took the KCCT and ITBS Tests in the Same School in the Same Year								
KCCT Math 2010	0.16	0.00	0.32	0.41	0.14	0.07	0.29	-0.05
ITBS Math 2010	-0.20	-0.24	-0.22	0.28	0.27	0.27	0.52	0.29
KCCT Reading 2010	0.27	0.36	0.02	0.14	-0.10	-0.05	-0.06	-0.02
ITBS Reading 2010	-0.07	-0.34	-0.38	0.18	0.08	0.29	0.41	0.17
KCCT Math 2011	0.20	0.00	0.09	0.51	0.13	-0.13	0.24	-0.21
ITBS Math 2011	-0.14	-0.26	-0.14	0.43	0.18	0.08	0.57	0.16
KCCT Reading 2011	0.38	0.22	-0.03	0.29	-0.15	-0.19	0.10	-0.12
ITBS Reading 2011	-0.03	-0.26	-0.30	0.30	0.10	0.29	0.47	0.13
Average Standard Scores of Students Who Took the KCCT 2011 and K-PREP 2012 in the Same School								
KCCT Reading 2011	0.37	0.16	0.04	0.15	-0.16	-0.07	0.12	-0.07
K-PREP Reading 2012	0.24	-0.07	-0.11	0.24	0.02	0.14	0.36	0.12
KCCT Math 2011	0.15	-0.05	0.08	0.42	0.06	-0.16	0.19	-0.10
K-PREP Math 2012	0.15	-0.13	-0.03	0.65	0.50	-0.04	0.31	0.20

Note: Site visits schools were chosen early in this study based on differences on KCCT ranks calculated based on the percentage of students who were proficient and distinguished each year. Data in this table show standard scores based on school average scale scores (rather than proficiency rates) for only those students who took the KCCT and comparison tests in the same school.

*The rank change patterns for School D are not consistent when they are calculated based on standard scale scores and include only those students who took the KCCT and comparison test in the same school. Unlike the other schools that decreased in rank, this school decreased in rank on some tests but not others and performed above average on ITBS and K-PREP tests.

Source: Staff analysis of Kentucky Department of Education data.

Appendix H

Site Visit Interview Protocols And Survey

Office of Education Accountability 2013 Assessment Study

Note: OEA site visit data do not report names of individuals, districts, or schools. Previous OEA reports: www.lrc.ky.gov/oea/material.htm.

Classroom Teacher Questions

1. What accounts for your school's success in helping students to master reading and math standards?
2. What curriculum documents did you use to plan instruction in 2010? Now?
3. How does your school curriculum address the challenge of balancing content coverage and content mastery?
4. What kinds of resources do you need to ensure that your students learn the new standards?
5. What do school and district administrators consider when evaluating the quality of your teaching?
6. Is there any tension between practices associated with success on student achievement goals and those associated with other curricular goals?
7. What types of data do you find most valuable in understanding whether students have mastered the standards?
8. What types of assessments do you use in your classroom? School?
9. How do data impact the decisions you make in your classroom? School?
10. The TellKY survey asks teachers to report hours spent per week preparing for federal, state, and local assessments. What types of activities were you thinking of when you answered this question?
11. In some years schools are given two different reading and math tests in the same subject. For example the ITBS tests have been given in the same year as the KCCT and the SAT 10 tests have been given in the same year as the K-PREP. Do you think students who do well on the KCCT or the K-PREP are also likely to do well on these other assessments?

Resource and Exceptional Child Teacher Questions

1. What accounts for your school's success in helping students to master reading and math standards?
2. How does your school address needs of struggling students? Accelerated students?
3. How does your school curriculum address the challenge of balancing content coverage and content mastery?
4. What do school and district administrators consider when evaluating the quality of your teaching?

5. Is there any tension between practices associated with success on student achievement goals and those associated with other curricular goals?
6. The TellKY survey asks teachers to report hours spent per week preparing for federal, state, and local assessments. What types of activities were you thinking of when you answered this question?
7. What kinds of data do you find most valuable in understanding whether students have mastered the standards?
8. Would you say that education is a priority in this community?
9. If you asked a student in this school why they needed to work hard in school, what would they say?
10. Does the typical student have opportunities for learning after school or during the summer?

Administrator Questions

1. What accounts for your success in helping students to master reading and math standards?
2. What curriculum documents did you use to plan instruction in 2010? Now?
3. How does your school curriculum address the challenge of balancing content coverage and content mastery?
4. How does your school address needs of struggling students? Accelerated students?
5. What kinds of resources do you need to ensure that your students learn the new standards?
6. What do you consider when evaluating the quality of your teachers?
7. What do you district administrators consider when evaluating the quality of your performance?
8. Is there any tension between practices associated with success on student achievement goals and those associated with other curricular goals?
9. What types of data do you find most valuable in understanding whether students have mastered the standards?
10. What types of assessments do you use in your school?
11. How do data impact the decisions you make in your school?
12. The TellKY survey asks teachers to report hours spent per week preparing for federal, state, and local assessments. In your school, what types of activities fit this description?
13. In some years schools are given two different reading and math tests in the same subject. For example the ITBS tests have been given in the same year as the KCCT and the SAT 10 tests have been given in the same year as the K-PREP. Do you think students who do well on the KCCT or the K-PREP are also likely to do well on these other assessments?

Student Population (if time)

14. Would you say that education is a priority in this community?
15. If you asked a student in this school why they needed to work hard in school, what would they say?
16. Does the typical student have opportunities for learning after school or during the summer?

Teacher Survey

Under each statement, mark the box that best describes your views.

The KCCT was a valuable tool in focusing attention of teachers in this school on student learning needs.			
Strongly Disagree	Disagree	Agree	Strongly Agree
Overall, the pressure to ensure that students performed well on the KCCT had a positive effect on the quality of teaching and learning in this school.			
Strongly Disagree	Disagree	Agree	Strongly Agree
In this district student scores on standardized tests are the indicator most valued by administrators in judging the quality of my teaching.			
Strongly Disagree	Disagree	Agree	Strongly Agree
The pressure to ensure that students performed well on the KCCT led me to make instructional decisions that were against my own standards of best practice.*			
Strongly Disagree	Disagree	Agree	Strongly Agree

*Prior to administering the survey, staff explained that this question was not asking teachers whether they had done anything unethical. It was asking whether they made instructional decisions that they would not have otherwise felt were the best for a particular child or class.

For each test preparation activity, mark the box that best describes the amount of time spent by students in your class during the 2010 and 2011 school years.

	Never	Less than once a month	Once a month or more	Once a week or more	Once a day or more
Answer questions in multiple-choice formats					
Answer questions in open-response formats					
Rewrite open-response questions using scoring rubric					
Learn test-taking strategies					
Take timed KCCT-like tests					
Receive awards or rewards for achieving test targets					
Receive additional instruction on KCCT-tested skills not yet mastered					

Please use the back of this paper to add any additional comments you have relevant to the questions asked during the interview.

Endnotes

- ¹ American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association, 1999. P. 9.
- ² Human Resources Research Organization. *Documenting Instruction and School Learning Environments in Kentucky Schools, Year 1*. Alexandria: HumRRO, 2007. Pp. 38-42.
- ³ Human Resources Research Organization. *Exploring Patterns in School Achievement from KCCT to K-PREP: The Role of School-Level Socioeconomic Status*. Alexandria: HumRRO, 2013.
- ⁴ National Governors Association Center for Best Practices, Council of Chief State School Officers. *Common Core Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Web. Nov. 11, 2013. P. 7.
- ⁵ *Ibid*, P. 36.
- ⁶ Kentucky. Dept. of Educ. *Kentucky Performance Rating for Educational Progress 2011-12 Technical Manual*. Web. Nov. 11, 2013. P. 53.
- ⁷ Koretz, Daniel. *Alignment, High Stakes, and the Inflation of Test Scores*. National Center for Research on Evaluation, Standards and Student Testing, Los Angeles: Center for the Study of Evaluation, 2005.
- ⁸ Auty, Bill. "Re: FW: OEA Assessment study – draft." Message to Roger Ervin. Nov. 18, 2013. Email.
- ⁹ Hambleton, Ronald. "re: clarification." Message to Deborah Nelson. Nov. 22, 2013. Email.
- ¹⁰ Polikoff, Morgan, Andrew C. Porter, and John Smithson. "How Well Aligned Are State Assessments of Student Achievement With State Content Standards?" *American Educational Research Journal*, 48(4), 2011. Pp. 965-995.
- ¹¹ Kentucky. Dept. of Educ. *Commonwealth Accountability Testing System 2010-11 Technical Report*. Web. Nov. 11, 2013. P. 51.
- ¹² Baker, Eva, and Robert Linn. "Validity Issues for Accountability Systems." In Susan H. Fuhrman and Richard Elmore, eds. *Redesigning Accountability Systems for Education*. New York: Teachers College Press, 2004. P. 53.
- ¹³ *Ibid*.
- ¹⁴ Koretz, Daniel. *Alignment, High Stakes, and the Inflation of Test Scores*. Los Angeles: National Center for Research on Evaluation, Standards and Student Testing Center for the Study of Evaluation, 2005.
- ¹⁵ Kentucky. Dept. of Educ. *Kentucky Performance Rating for Educational Progress 2011-12 Technical Manual*. Web. Nov. 11, 2013. P. 53.
- ¹⁶ Herman, Joan. "The Effects of Testing on Instruction." In Susan H. Fuhrman and Richard Elmore, eds. *Redesigning Accountability Systems for Education*. New York: Teachers College Press, 2004. Pp. 141-166.
- ¹⁷ Stecher, Brian. "Consequences of Large-scale, High-stakes Testing on School and Classroom Practice." In Laura Hamilton, Brian Stecher, and Stephen Klein, eds. *Making Sense of Test-Based Accountability in Education*. Santa Monica: RAND, 2002.
- ¹⁸ Pedulla, Joseph, Lisa Abrams, George Madaus, Michael Russell, Miguel Ramos, and Jing Miao. *Perceived Effects of State-mandated Testing Programs on Teaching and Learning: Findings From a National Survey of Teachers*. Boston: National Board on Educational Testing and Public Policy, 2003.
- ¹⁹ Hamilton, Laura, Brian Stecher, and Stephen Klein, eds. *Making Sense of Test-Based Accountability in Education*. Santa Monica: RAND, 2002; Koretz, Daniel. *Alignment, High Stakes, and the Inflation of Test Scores*. Los Angeles: National Center for Research on Evaluation, Standards and Student Testing Center for the Study of Evaluation, 2005; Herman, Joan. "The Effects of Testing on Instruction. In Susan H. Fuhrman and Richard Elmore, eds. *Redesigning Accountability Systems for Education*. New York: Teachers College Press, 2004. Pp. 141-166.
- ²⁰ Stecher, Brian. "Consequences of Large-scale, High-stakes Testing on School and Classroom Practice." In Laura Hamilton, Brian Stecher, and Stephen Klein, eds. *Making Sense of Test-Based Accountability in Education*. Santa Monica: RAND, 2002.
- ²¹ *Ibid*, P.18.
- ²² Kidwell, Karen. Personal communication to Deborah Nelson. Oct. 18, 2013.
- ²³ Danielson, Charlotte. *The Framework for Teaching Evaluation Instrument 2013 Edition*. Web. Nov. 11, 2013.
- ²⁴ Klein, Stephen, Laura Hamilton, Daniel McCaffrey, and Brian Stecher. "What Do Test Scores in Texas Tell Us?" IP-202. Santa Monica: RAND, 2000.
- ²⁵ Hambleton, Ronald, RM Jaeger, Daniel Koretz, Robert Linn, J. Millman, and SE Phillips. *Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991-1994*. Frankfurt: Legislative Research Commission, Office of Education Accountability, June 1995.

- ²⁶ Koretz, Daniel, and Sheila Barron. *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica: RAND, 1998.
- ²⁷ Jacob, Brian. *Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments*. NBER Working Paper #12817. Cambridge: National Bureau of Economic Research, 2006.
- ²⁸ Jacob, Brian. "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*. 89(5-6), 2005. Pp. 761-796.
- ²⁹ Santelices, Maria, and Mark Wilson. "Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning." *Harvard Educational Review*, 80(1) (2010): 106-133.
- ³⁰ Kentucky. Office of Education Accountability. *Compendium of State Education Rankings, 2011*. Legislative Research Commission Report. Dec. 2011.
- ³¹ Kentucky. Dept. of Educ. *How to Develop a Standards-Based Unit of Study*. Web. Aug. 11, 2013.
- ³² Kentucky. Dept. of Educ. Commonwealth Accountability Testing System Technical Report 2008-09 Technical Report. Dover: Measured Progress, n.d.