REPORT

Alignment Analysis of Science Standards to ACT College Readiness Standards-Science and ACT-Science Assessments

Kentucky

High School

Norman L. Webb

June 2, 2007

REPORT

Alignment Analysis of Science Standards to ACT College Readiness Standards-Science and ACT-Science Assessments

Kentucky
High School

Norman L. Webb

June 2, 2007

Acknowledgements

John Putnam	Group Leader	WV
Doug Johnson		WI
Jim Leidel		WI
Angela Murdock		KY
Diane Johnson		KY
Eric Thomas		KY
Brian Vesperman	Technical Consultant	

The Kentucky Department of Education funded this analysis. Pamela Rogers, Associate Commissioner, Office of Assessment and Accountability, was the main contact person for the Kentucky Department of Education and oversaw the coordination of the study

Table of Contents

Executive Summary iii

Introduction 1

Alignment Criteria Used for This Analysis 3
Categorical Concurrence 3
Depth-of-Knowledge Consistency 3
Range-of-Knowledge Correspondence 6
Balance of Representation 6
Source of Challenge 7

Findings 8
Standards 8
Alignment of Kentucky CCA 4.1 Standards with the ACT College Readiness Standards-Science 15
Alignment of Curriculum Standards and Assessments 17
Summary of Items Coded to Standards Across Forms 22
Source-of-Challenge Issues 26
Reviewers' Comments 26
Reliability Among Reviewers 26

Summary 27

References 28

Appendix A
Kentucky Science High School Standards and Group Consensus DOK Values

Appendix B
Data Analysis Tables Kentucky High School Science Standards to ACT College Readiness Standards-Science

Appendix C
Data Analysis Tables Kentucky High Science and Mathematics Standards to ACT-Science Assessments

Appendix D
Reviewers' Notes and Source-of-Challenge Comments Kentucky High School Science and Mathematics Standards to ACT Science Assessments

Appendix E
Debriefing Summary Notes Kentucky High School Science and Mathematics Standards to ACT Science Assessments

Executive Summary

A three day alignment institute was held in Frankfort, Kentucky, on May 18-20, 2007, to analyze the alignment between the Kentucky Core Content for Assessment (CCA) 4.1 high school science standards and four forms of the ACT-science assessment. The Kentucky science standards were also compared to the ACT College Readiness Standards-Science (ACT CRS-S). The six reviewers included science content experts and science teachers. Three of the reviewers were from Kentucky and three were from other states. They analyzed the agreement between the Kentucky standards and four forms (62F, 63D, 63E, and 64C) of the ACT-science assessment.

The results indicated that there was poor alignment between the Kentucky CCA 4.1 high school science standards with the ACT CRS-S and with the ACT-science assessments. Reviewers complemented the Commonwealth of Kentucky on very comprehensive and demanding standards. They also indicated that the ACT CRS-S stated and the assessments measured important scientific knowledge. However, the ACT documents attended primarily to science process skills—interpretation of data, scientific investigations, and evaluation of models—rather than content knowledge of different science areas expected in the Kentucky standards (physical science, earth and space science, and biological science).

Reviewers only found 10 of 47 ACT CRS-S standards (21%) that matched any of the Kentucky science standards. These ACT standards corresponded to about 30% of the Kentucky science standards. The ACT CRS-S also had lower depth-of-knowledge levels than expected by the Kentucky science standards. Considering the content in both the ACT CRS-S and the Kentucky CCA 4.1 high school science standards, only about 13% of the content was in common to both documents, 40% exclusive to the Kentucky science standards, and 47% exclusive to the ACT CRS-S.

The alignment between the Kentucky CCA 4.1 high school science standards and the ACT-science assessments was not much better than for the ACT CRS-S. Each of the four assessment forms did not include enough items for one or two of the four Kentucky subdomains. The subdomain without at least six corresponding items varied by assessment form among Subdomain 2 (Earth and Space Science), Subdomain 3 (Biological Science), and Subdomain 4 (Unifying Concepts).

The items that did correspond to the subdomains were generally lower in depth-of-knowledge level than the ceiling level expected by the Kentucky standards. Most of the ACT-science assessment items had a DOK level 2 while over 60% of the Kentucky standards were assigned a DOK level 3. Only one assessment form, 64C, met the Depth-of-Knowledge Consistency criterion for three of the four Kentucky subdomains. Form 62F met the criterion for only one subdomain while the other two forms did not satisfy the DOK criterion for any of the subdomains. None of the four assessment forms had sufficient coverage of the standards under the subdomains to satisfy the acceptable level for the Range-of-Knowledge Correspondence criterion for any of the four subdomains.

Overall, from 14 to 20 items on the 40 item assessment forms would need to be replaced to attain full alignment between the ACT-science assessment and the Kentucky high school science standards. The ACT documents, designed as a prediction tool for success in college, have a different and very limited purpose than to assess or communicate all that students should know about science upon leaving high school as included in the Kentucky high school science standards.

Alignment Analysis of Science Standards to ACT College Readiness Standards-Science and ACT-Science Assessments

Kentucky
High School

Norman L. Webb

Introduction

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Alignment describes the match between expectations and an assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997).

A three-day Alignment Analysis Institute was conducted April 18-20, 2007, in Frankfort, Kentucky. Six reviewers—including science content experts and science teachers—analyzed the agreement between the Core Content for Assessment (CCA) 4.1 in science and four forms (62F, 63D, 63E, and 64C) of the ACT-science test. Three of the reviewers were from Kentucky, one was from West Virginia, and two were from Wisconsin.

The Core Content for Assessment 4.1 (CCA 4.1) used in this analysis is a subset of the content standards in *Kentucky's Program of Studies for Grades Primary – 12*. The Commonwealth of Kentucky assessed for the first time in spring 2007 the content specified in the CCA 4.1. The Commonwealth of Kentucky uses the terminology of *domain*, *subdomains*, and *standards*. The domain of science is subdivided into four subdomains. These subdomains are specified in greater detail by standards. Only standards designated for assessment were included in this analysis. Data for this analysis were entered at the standard level and reported at the subdomain level.

As part of the alignment institute, reviewers were trained to identify the depth-of-knowledge of the standards and assessment items. This training included reviewing the definitions of the four depth-of-knowledge (DOK) levels and reviewing examples of each. Then the reviewers generally participated in 1) a consensus process to determine the depth-of-knowledge levels of the standards and 2) individual analyses of the assessment items. However, the Commonwealth of Kentucky had committees of teachers assign DOK levels to the standards, representing the ceiling level or highest level of complexity for the standard. In this alignment analysis reviewers did review the assigned DOK levels and only found two or three DOKs that were different from the assigned values. The Kentucky DOK values were used in this analysis. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they assessed the degree to which they had coded particular items or types of content to the standards.

To derive the results from the analysis, the reviewers' responses are averaged. Any variance among reviewers is considered legitimate, with the true depth-of-knowledge level for the item falling somewhere between the two or more assigned values. Such variation could signify a lack of clarity in how the standards were written, the robustness of an item that can legitimately correspond to more than one standard, and/or a depth of knowledge that falls in between two of the four defined levels. Reviewers were allowed to identify one assessment item as corresponding to up to three standards—one primary hit (standard) and up to two secondary hits. However, reviewers only could code one depth-of-knowledge level to each assessment item, even if the item corresponded to more than one standard.

Reviewers were instructed to focus primarily on the alignment between the commonwealth standards and assessments. However, reviewers were encouraged to offer their opinions on the quality of the standards, or of the assessment activities/items, by writing a note about the item. Reviewers could also indicate whether there was a source-of-challenge issue with the item—i.e., a problem with the item that might cause the student who knows the material to give a wrong answer, or enable someone who does not have the knowledge being tested to answer the item correctly.

The results produced from the institute pertain only to the issue of alignment between the Kentucky standards and the ACT assessment documents and instruments. Note that this alignment analysis of this nature does not serve as external verification of the general quality of the commonwealth's standards or assessments. Rather, only the degree of alignment is discussed in the results. For these results, the means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the means lessened the error that might result from any one reviewer's finding. Standard deviations are reported in the tables provided in the Appendix B, which give one indication of the variance among reviewers.

The present report describes the results of an alignment study of the Kentucky Core Content for Assessment Version 4.1 for high school released in August 2006 and four forms of the ACT-science assessment. The study addressed specific criteria related to the content agreement between the commonwealth standards and grade-level assessments. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the standards and the assessments on the basis of four criteria. Information is also reported on the quality of items by identifying items with sources-of-challenge and other issues. For each alignment criterion, an acceptable level was defined by what would be required to assure that a student had met the Standards.

Categorical Concurrence

An important aspect of alignment between standards and assessments is whether both address the same content categories. The categorical-concurrence criterion provides a very general indication of alignment if both documents incorporate the same content. *The criterion of categorical concurrence between subdomains and assessments is met if the same or consistent categories of content appear in both documents.* This criterion was judged by determining whether the assessment included items measuring content from each subdomain. The analysis assumed that the assessment had to have at least six items for measuring content from a subdomain in order for an acceptable level of categorical concurrence to exist between the subdomain and the assessment. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable subscale for estimating students' mastery of content on that subscale. Of course, many factors have to be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming that the cutoff score is the mean and that the reliability of one item is .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or nonmasters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. Usually states do not report student results by subdomain or require students to achieve a specified cutoff score on subscales related to a subdomain. If a state did do this, then the state would seek a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a subdomain, and as a basis for making some decisions about students' knowledge of that subdomain. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement on the subscale.

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each. *Depth-of-knowledge consistency between subdomains and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively* *as what students are expected to know and do as stated in the subdomains.* For consistency to exist between the assessment and the subdomain, as judged in this analysis, at least 50% of the items corresponding to a subdomain had to be at or above the level of knowledge of the subdomain: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one subdomain of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding subdomain. For example, assume an assessment included six items related to one subdomain and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding standards, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one standard. Some leeway was used in this analysis on this criterion. If a subdomain had between 40% and 50% of items at or above the depth-of-knowledge levels of the standards, then it was reported that the criterion was "weakly" met.

Interpreting and assigning depth-of-knowledge levels to both standards within subdomains and assessment items is an essential requirement of alignment analysis. These descriptions help to clarify what the different levels represent in science:

Level 1 (Recall and Reproduction) is the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple science process or procedure. Level 1 only requires students to demonstrate a rote response, use a well-known formula, follow a set procedure (like a recipe), or perform a clearly defined series of steps. A "simple" procedure is well defined and typically involves only one step. Verbs such as "identify," "recall," "recognize," "use," "calculate," and "measure" generally represent cognitive work at the recall and reproduction level. Simple word problems that can be directly translated into and solved by a formula are considered Level 1. Verbs such as "describe" and "explain" could be classified at different DOK levels, depending on the complexity of what is to be described and explained.

A student answering a Level 1 item either knows the answer or does not: that is, the answer does not need to be "figured out" or "solved." In other words, if the knowledge necessary to answer an item automatically provides the answer to the item, then the item is at Level 1. If the knowledge necessary to answer the item does not automatically provide the answer, the item is at least at Level 2. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Recall or recognize a fact, term, or property.
- Represent in words or diagrams a scientific concept or relationship.
- Provide or recognize a standard scientific representation for a simple phenomenon.
- Perform a routine procedure such as measuring length.

Level 2 (Skills and Concepts) includes the engagement of some mental processing beyond recalling or reproducing a response. The content knowledge or process involved is *more complex* than in Level 1. Items require students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Level 2 item include “classify,” “organize,” ”estimate,” “make observations,” “collect and display data,” and “compare data.” These actions imply *more than one step*. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts. Some action verbs, such as “explain,” “describe,” or “interpret,” could be classified at different DOK levels, depending on the complexity of the action. For example, interpreting information from a simple graph, requiring reading information from the graph, is a Level 2. An item that requires interpretation from a complex graph, such as making decisions regarding features of the graph that need to be considered and how information from the graph can be aggregated, is at Level 3. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Specify and explain the relationship between facts, terms, properties, or variables.
- Describe and explain examples and non-examples of science concepts.
- Select a procedure according to specified criteria and perform it.
- Formulate a routine problem, given data and conditions.
- Organize, represent, and interpret data.

Level 3 (Strategic Thinking) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. The cognitive demands at Level 3 are complex and abstract. The complexity does not result only from the fact that there could be multiple answers, a possibility for both Levels 1 and 2, but because the multi-step task requires more demanding reasoning. In most instances, requiring students to explain their thinking is at Level 3; requiring a very simple explanation or a word or two should be at Level 2. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Experimental designs in Level 3 typically involve more than one dependent variable. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Identify research questions and design investigations for a scientific problem.
- Solve non-routine problems.
- Develop a scientific model for a complex situation.
- Form conclusions from experimental data.

Level 4 (Extended Thinking) involves high cognitive demands and complexity. Students are required to make several connections—relate ideas within the content area or among content areas—and have to select or devise one approach among many alternatives on how the situation can be solved. Many on-demand assessment instruments will not include any assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated in such a way as to expect students to perform extended thinking. “Develop generalizations of the results obtained and the strategies used and apply them to new problem situations,” is an example of a Grade 8 objective that is at Level 4. Many, but not all, performance assessments and open-ended assessment activities requiring significant thought will be Level 4.

Level 4 requires complex reasoning, experimental design and planning, and probably will require an extended period of time either for the science investigation required by an objective, or for carrying out the multiple steps of an assessment item. However, the extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2 activity. However, if the student conducts a river study that requires taking into consideration a number of variables, this would be a Level 4. Some examples that represent, but do not constitute all of, a Level 4 performance are:

- Based on data provided from a complex experiment that is novel to the student, deduct the fundamental relationship between several controlled variables.
- Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions.

Range-of-Knowledge Correspondence

For standards and assessments to be aligned, the breadth of knowledge required on both should be comparable. *The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a subdomain is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.* The criterion for correspondence between span of knowledge for a subdomain and an assessment considers the number of standards within the subdomain with one related assessment item/activity. Fifty percent of the standards for a subdomain had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students’ knowledge should be tested on content from over half of the domain of knowledge for a subdomain. This assumes that each standard for a subdomain should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one standard, the requirement that assessment items need to be related to more than 50% of the standards for an subdomain increases the likelihood that students will have to demonstrate knowledge on more than one standard per subdomain to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the standards. However, any restriction on the number of items included on the test will place an upper limit on the number of standards that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of standards and a large number of subdomains. If 50% or more of the standards for a subdomain had a corresponding assessment item, then the range-of-knowledge correspondence criterion was met. If between 40% and 50% of the standards for a subdomain had a corresponding assessment item, the criterion was “weakly” met.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned standards and assessments require that knowledge be distributed equally in both. The range-of-knowledge criterion only considers the number of standards within a subdomain hit (a standard with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these standards. *The balance-of-representation criterion is used to indicate the degree to which one standard is given more emphasis on the assessment than another.* An index is used to judge the distribution of assessment items. This index only considers the standards for a subdomain that have at least one hit—i.e., one related assessment item per standard. The index is computed by considering the difference in the proportion of standards and the proportion of hits assigned to the subdomain. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a subdomain are equally distributed among the standards for the given subdomain. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the standards hit. Depending on the number of standards and the number of hits, a unimodal distribution (most items related to one standard and only one item related to each of the remaining standards) has an index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate that items/activities are distributed among all of the standards at least to some degree (e.g., every standard has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been “weakly” met.

Source-of-Challenge Criterion

The Source-of-Challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted science standard, concept, or application. Cultural bias or specialized knowledge could be reasons for an item to have a source-of-challenge problem. Such item characteristics may result in some students not answering an assessment item, or answering an assessment item incorrectly, or at a lower level, even though they possess the understanding and skills being assessed.

Findings

Standards

The assigned DOK levels to the standards by Kentucky teachers were used in this analysis. The DOK values were considered as a ceiling level or the highest level that student content knowledge should be assessed. The DOK value for each science standard can be found in Appendix A. Table 1 shows the percentage of standards at each DOK level. The first row shows that the ceiling level for three quarters of the CCA 4.1 science standards was assigned a DOK level 3. The remaining nine science standards had a DOK level 2. These expectations for students are high with the ceiling level for most standards having students engage in strategic thinking and analysis.

A number of the high school science standards had two parts designated by bullets. Reviewers found that many of the ACT College Readiness Standards and the ACT assessment items targeted only one of the bullets, but not both. There was a tendency for one bullet to relate to knowledge of science content while the other bullet related more to a process skill. For example, under the Subdomain 1 (Physical Science) and the organizer "Structure and Transformation of Matter," Standard 1.16 had two parts and given a DOK 3:

1.1.6 Students will:

- identify variables that affect reaction rates;
- predict effects of changes in variables (concentration, temperature, properties of reactants, surface area and catalysts) based on evidence/data from chemical reactions. Rates of chemical reactions vary. Reaction rates depend on concentration, temperature and properties of reactants. Catalysts speed up chemical reactions.

The science reviewers felt the alignment analysis would be more meaningful and precise if items could be coded to either the first bullet (science content knowledge) or to the second bullet (science process knowledge) rather than to just 1.1.6 with a DOK level 3. Reviewers then divided the high school science standards into two parts when one bullet addressed content and the other addressed a process skill. The content knowledge was designated by an "a" after the standard number (1.1.6a) and the process skills were designated by a "b". Some high school standards already had a single focus by being either content knowledge or process skills. In this few cases, reviewers assigned the standard either an "a" or a "b". For example, standard 1.1.5 was designated as 1.1.5a because students were expected to" explain the role of intermolecular or intramolecular interactions on the physical properties." Standard 1.1.1 was designated as 1.1.1b because students were expected to observe and make generalizations from data, a process skill. How the standards were labeled as "a" or "b" can be found in Appendix A. When the high school standards were divided into content and process standards, reviewers assigned two content standards (an "a" standard) with a DOK level 1 and five other content standards with a DOK level 2. The distribution of the DOK levels for the high school science standards when separated into content standards and process standards is shown in the second row of Table 1.

Table 1
Percent of Standards by Depth-of-Knowledge (DOK) Levels for Kentucky Alignment Analysis for High School Science

Assessment/Standard	Total Number of Standards	DOK Level	Number of Standards by Level	Percent within Subdomain by Level
Kentucky High School CCA 4.1 Science Standards	32	2 3	9 23	28 72
Kentucky High School Science Standards Separated into Process and Content Standards	39	1 2 3	2 14 23	5 35 58

If no particular standard is targeted by a given assessment item, reviewers are instructed to code the item at the level of a subdomain. This coding to a "generic standard" sometimes indicates that the item is inappropriate for the grade level. However, if the item is grade-appropriate, then this situation may instead indicate that there is a part of the content not expressly or precisely described in the standards or the content is expected by standards in other grades. These items may highlight areas in the standards that should be changed, or made more precise. If reviewers did not think that there was a match between an assessment item and any standard or subdomain, then they were instructed to code the assessment item as uncodeable.

Table 2 displays the ACT CRS-S standards coded to generic standards or uncodeable by more than one reviewer. The majority of reviewers coded 38 of the 47 ACT CRS-S standards (81%) as uncodeable. Reviewers indicated that the ACT CRS-S primarily were statements about the scientific process under three categories:

I. Interpretation of Data (IOD)
II. Scientific Investigations (SIN)
III. Evaluation of Models, Inferences, and Experimental Results (EMI)

ACT acknowledges that the ACT CRS-S are measured in the "context of science topics students encounter in science courses." However, the 47 statements of expectations in the ACT CRS-S stated activities that students should engage in across the different science topics. For example,

I. 502 (#11 in this analysis) Compare or combine data from a complex data presentation.
II. 502 (#27) Understand a complex experimental design.
III. 502 (#38) Determine whether given information supports or contradicts a simple hypothesis or conclusion, and why.

Reviewers did find a match between the Kentucky science standards and 10 of the ACT CRS-S including, as examples:

I. 304 (#6) Determine how the value of one variable changes as the value of another variable changes in a simple data presentation. (Kentucky Standard 1.1.6b)
II. 504 (#29) Determine the experimental conditions that would produce specified results. (Kentucky Standard 1.1.8b)
III. 501 (#37) Select a simple hypothesis, prediction, or conclusion that is supported by a data presentation or a model. (Kentucky Standard 4.7.2b)

Reviewers' comments indicated that the ACT CRS-S statements were very broad and addressed many embedded or implied expectations found in the Kentucky CCA 4.1 science standards. However, they found little or no alignment between the ACT CRS-S and the Kentucky standards. An explanation given by one reviewer was very typical for all six reviewers when asked if the ACT CRS-S covered the most important topics in the Kentucky standards, "No, the items in their current form are very general in nature and because of that, they do not align very well to the state standards which in contrast are very specific." (Appendix E). Another reviewer commented, "The ACT (standards) were good (standards), but not a good match for the KY standards because the (Kentucky) standards included both process and content items. The ACT focuses on process."

The reviewers found the DOK levels in both sets of standards to be more comparable. Both sets of standards had a range in DOK levels with the Kentucky standards having a little higher proportion of standards with a DOK level 3. Reviewers acknowledged that the purposes for the two sets of standards were not the same. The ACT CRS-S were directed toward college preparation whereas the Kentucky standards were designed to guide the assessment of students' science understanding over the high school curriculum. Overall, reviewers found there to be low alignment between the ACT CRS-S and the Kentucky CCA 4.1 high school science standards. As one reviewer summarized:

> We could not code 37 of the 47 ACT College Readiness Standards to KY's CCA. Although we felt that the KY standards implied that students would have to do some of the process skills delineated in the ACT items, we could not show a direct or strong match for many of the ACT items. Although CCA has some process embedded, there were still more standards that were more content focused. (Appendix E)

Table 2
Items Coded to Generic Standards by More Than One Reviewer, Kentucky Alignment Analysis for High School Science, ACT College Readiness Standards-Science

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT CRS-S		
ACT College Readiness	2	Uncodeable (4)
ACT College Readiness	3	Uncodeable (6)
ACT College Readiness	4	1.1 (6), 2.3 (6), 3.4 (6)
ACT College Readiness	5	Uncodeable (6)

Table 2 (continued)

Items Coded to Generic Standards by More Than One Reviewer, Kentucky Alignment Analysis for High School Science, ACT College Readiness Standards-Science

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT CRS-S		
ACT College Readiness	9	Uncodeable (6)
ACT College Readiness	10	Uncodeable (6)
ACT College Readiness	11	Uncodeable (6)
ACT College Readiness	12	Uncodeable (6)
ACT College Readiness	14	Uncodeable (6)
ACT College Readiness	15	Uncodeable (6)
ACT College Readiness	16	Uncodeable (6)
ACT College Readiness	17	Uncodeable (6)
ACT College Readiness	18	Uncodeable (6)
ACT College Readiness	19	Uncodeable (6)
ACT College Readiness	20	Uncodeable (5)
ACT College Readiness	21	Uncodeable (6)
ACT College Readiness	22	Uncodeable (6)
ACT College Readiness	23	Uncodeable (6)
ACT College Readiness	24	Uncodeable (6)
ACT College Readiness	25	Uncodeable (6)
ACT College Readiness	26	Uncodeable (6)
ACT College Readiness	27	Uncodeable (6)
ACT College Readiness	28	Uncodeable (6)
ACT College Readiness	30	Uncodeable (6)
ACT College Readiness	31	Uncodeable (6)
ACT College Readiness	32	Uncodeable (6)
ACT College Readiness	33	Uncodeable (6)
ACT College Readiness	34	Uncodeable (6)
ACT College Readiness	36	Uncodeable (6)
ACT College Readiness	38	Uncodeable (5)
ACT College Readiness	39	Uncodeable (6)
ACT College Readiness	40	Uncodeable (6)
ACT College Readiness	41	Uncodeable (6)
ACT College Readiness	42	Uncodeable (6)
ACT College Readiness	43	Uncodeable (6)
ACT College Readiness	44	Uncodeable (6)
ACT College Readiness	45	Uncodeable (6)
ACT College Readiness	46	Uncodeable (6)
ACT College Readiness	47	Uncodeable (6)

Reviewers varied by assessment form in the number of 40 items on the ACT-science assessment forms they assigned to generic standards. On Form 62F reviewers assigned each item to a generic standard (Table 3). Whereas on Form 64C reviewers assigned 14 items (35%) to generic standards. On the other two forms, reviewers assigned 28 items (Form 63D) and 23 items (Form 63E) to generic standards. Reviewers' comments indicate that the assessment items focused on processes while many of the Kentucky science standards expected students to explain or describe a concept. For example, a reviewer wrote a comment for item 14 on Form 62F, "Question implies gravitational relationship, but it is a process question. It is not a question that requires students to explain, which is what the descriptors says." The large number of items mapped to generic standards, from 35% to 100%, indicates that the assessment forms only generally corresponded to the content specified in the Kentucky standards. A number of the items on the ACT-science assessment only generally measured content expectations in the Kentucky CCA 4.1 science high school standards without measuring the specific content knowledge expected by the standards.

Table 3

Items Coded to Generic Standards by More Than One Reviewer, *Kentucky Alignment Analysis for High School Science Standards and Four Forms of the ACT-Science Assessment*

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT-Science Assessment		
Science to 62F	1	1.2 (6)
Science to 62F	2	1.2 (6)
Science to 62F	3	1.2 (6)
Science to 62F	4	1.2 (6)
Science to 62F	5	1.2 (5)
Science to 62F	6	1.2 (5)
Science to 62F	7	1.2 (5)
Science to 62F	8	1.2 (5)
Science to 62F	9	1.2 (5)
Science to 62F	10	1.2 (5)
Science to 62F	11	1.2 (5)
Science to 62F	12	1.2 (5)
Science to 62F	13	2.3 (6)
Science to 62F	14	2.3 (6)
Science to 62F	15	2.3 (6)
Science to 62F	16	2.3 (6)
Science to 62F	17	2.3 (6)
Science to 62 F	18	3.5 (5)
Science to 62F	19	3.5 (5)
Science to 62F	20	3.5 (5)
Science to 62F	21	3.5 (5)

Table 3 (continued)
Items Coded to Generic Standards by More Than One Reviewer, Kentucky Alignment Analysis for High School Science Standards and Four Forms of the ACT-Science Assessment

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT-Science Assessment		
Science to 62F	22	3.5 (5)
Science to 62F	23	4.6 (5)
Science to 62F	24	4.6 (5)
Science to 62 F	25	4.6 (5)
Science to 62F	26	4.6 (5)
Science to 62 F	27	4.6 (5)
Science to 62F	28	3.4 (5)
Science to 62F	29	3.4 (5)
Science to 62F	30	3.4 (5)
Science to 62F	31	3.4 (5)
Science to 62F	32	3.4 (4)
Science to 62F	33	3.4 (3)
Science to 62F	34	1.1 (4) 3.4 (5)
Science to 62F	35	1.1 (5) 3.4 (6)
Science to 62F	36	1.1 (5) 3.4 (6)
Science to 62F	37	1.1 (5) 3.4 (6)
Science to 62F	38	1.1 (5) 3.4 (6)
Science to 62F	39	1.1 (5) 3.4 (6)
Science to 62F	40	1.1 (5) 3.4 (6)
Science to 63D	1	3.5 (6)
Science to 63D	2	3.5 (6)
Science to 63D	3	3.5 (6)
Science to 63D	4	3.5 (6)
Science to 63D	5	3.5 (6)
Science to 63D	6	3.5 (4)
Science to 63D	13	1.2 (6)
Science to 63D	14	1.2 (6)
Science to 63D	15	1.2 (6)
Science to 63D	16	1.2 (6)
Science to 63D	17	1.2 (6)
Science to 63D	18	1.2 (6)
Science to 63D	19	1.1 (4)
Science to 63D	20	1.1 (4)
Science to 63D	21	1.1 (4)
Science to 63D	22	1.1 (4)
Science to 63D	23	1.1 (4)

Table 3 (continued)
Items Coded to Generic Standards by More Than One Reviewer, Kentucky Alignment Analysis for High School Science Standards and Four Forms of the ACT-Science Assessment

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT-Science Assessment		
Science to 63D	24	2.3 (5) 4.6 (5)
Science to 63D	25	2.3 (5) 4.6 (5)
Science to 63D	26	2.3 (5) 4.6 (4)
Science to 63D	27	2.3 (5) 4.6 (4)
Science to 63D	28	2.3 (5) 4.6 (4)
Science to 63D	34	4.6 (2)
Science to 63D	36	4.6 (2)
Science to 63D	37	4.6 (2)
Science to 63D	38	3.5 (3)
Science to 63D	39	4.6 (2)
Science to 63D	40	4.6 (2)
Science to 63E	12	4.7 (6)
Science to 63E	13	4.7 (6)
Science to 63E	14	4.7 (6)
Science to 63E	15	4.7 (6)
Science to 63E	16	4.7 (6)
Science to 63E	17	4.7 (6)
Science to 63E	18	4.7 (6)
Science to 63E	25	1.2 (6)
Science to 63E	26	1.2 (6)
Science to 63E	27	1.2 (6)
Science to 63E	28	1.2 (6)
Science to 63E	29	1.2 (6)
Science to 63E	30	1.2 (6)
Science to 63E	31	1.1 (6)
Science to 63E	32	1.1 (6)
Science to 63E	33	1.1 (6)
Science to 63E	34	1.1 (6)
Science to 63E	35	1.1 (6)
Science to 63E	36	2.3 (4)
Science to 63E	37	2.3 (4)
Science to 63E	38	2.3 (4)
Science to 63E	39	2.3 (4)
Science to 63E	40	2.3 (5)
Science to 64C	12	4.7 (4)
Science to 64C	13	4.7 (4)

Table 3 (continued)
Items Coded to Generic Standards by More Than One Reviewer, Kentucky Alignment Analysis for High School Science Standards and Four Forms of the ACT-Science Assessment

ACT Assessment/Standard	Assessment Item	Generic Standard (Number of Reviewers)
ACT-Science Assessment		
Science to 64C	14	4.7 (4)
Science to 64C	15	4.7 (4)
Science to 64C	16	4.7 (4)
Science to 64C	23	3.4 (5)
Science to 64C	24	3.4 (5)
Science to 64C	25	3.4 (5)
Science to 64C	26	3.4 (5)
Science to 64C	27	3.4 (5)
Science to 64C	36	1.1 (6)
Science to 64C	37	1.1 (6)
Science to 64C	38	1.1 (6)
Science to 64C	39	1.1 (6)

Reviewers' debriefing comments also highlight some issues in matching the assessment items and the standards. These comments can be found in Appendix E. Reviewers' debriefing comments explain in more detail the differences between the Kentucky standards and the ACT science forms.

Alignment of Kentucky CCA 4.1 Standards with the ACT College Readiness Standards-Science

The ACT CRS-S were organized in three categories:

I. Interpretation of Data (IOD)
II. Scientific Investigations (SIN)
III. Evaluation of Models, Inferences, and Experimental Results (EMI)

The ACT CRS-S were statements about the process of standards that were intended to be measured in the content of science topics students are to encounter in science courses.

Each of the three science categories were further divided into the level of standards corresponding to a score range on the ACT set of assessments (EXPLORE, PLAN, and ACT). The standards indicating the level of performance for the lowest score range (13-15) were numbered 201 to 202 and had standards only under Category I. The standards indicating the level of performance for the highest score range (33-36) that apply only to the ACT were numbered with 700's—701 to 703. The four levels in between were labeled in the 300's (16-19 scores), 400's (20-23 scores), 500's (24-27 scores), and 600's (28-32 scores only on the PLAN and ACT). For this analysis the ACT standards were numbered sequentially. What assigned number corresponded to what ACT standard can be found in Table 9 of Appendix B.

The alignment between the Kentucky CCA 4.1 high school science standards and the ACT CRS-S was very general at best and very little if the specific standards were considered. As noted above, reviewers coded most of the ACT CRS-S as uncodeable (81%). On the average, reviewers only coded nine or ten ACT standards as corresponding to Kentucky CCA 4.1 science standards. Reviewers indicated that most of these nine or ten ACT standards could be matched, in part, to two or three Kentucky standards. On the average, reviewers had 18 hits between the Kentucky standards and the ACT CRS-S (Table 4). Reviewers judged that five of the ACT standards mapped to Kentucky standard 1.1.6b (predict effects of changes in variables based on evidence/data from chemical reactions). Two of the ACT standards mapped to Kentucky standard 1.1.8b (justify conclusions using evidence/data from chemical reactions) and one mapped to 1.1.6a (identify variables that affect reaction rates).

Note that many of the Kentucky science standards had a process part and a content part. Reviewers used “a” to designate the content part and “b” to designate the process part of the expectation. Under Subdomain 1 (Physical Science), reviewers found seven ACT standards that mapped to Kentucky standards, one of the ACT standards had two hits. Five of the ACT standards mapped to 1.1.6b, two to 1.1.8b, and one to 1.1.6a. It should be noted that seven of the eight hits for Subdomain 1 was with process standards. For the other three Kentucky science subdomains, reviewers found only two to four corresponding ACT CRS-S standards. Reviewers judged that all of these ACT standards corresponded to the either process parts of the Kentucky standards or to the general statements for a grouping of standards.

The level of complexity of the ACT CRS-S standards that corresponded to the Kentucky science standards was below the level expected by the Kentucky standards. Of the 10 ACT CRS-S standards that corresponded to Kentucky standards, only 28% on the average (or about 3 of the 10) had a DOK level what was the same or higher than the DOK of the corresponding Kentucky standard. In general, the Kentucky standards were at a DOK level 3 (strategic analysis) whereas the ACT standards were rated at a DOK level 1 or 2.

Reviewers judged that the ACT CRS-S standards only corresponded to the about one-fourth of the Kentucky high school science standards—38% of the standards under Subdomain 1, 21% under Subdomain 2, 33% under Subdomain 3 and 14% under Subdomain 4 (Table 4). Considering the expectations stated in both documents, 27% was in common while 47% was addressed in the ACT CRS-S and not in the Kentucky standards and 40% was addressed in the Kentucky standards and not in the ACT standards. In general, as indicated by the Balance of Representation index, the ACT CRS-S standards that did correspond to 11 of the Kentucky standards did not over emphasize any one of the standards. Overall, the analysis indicates there was low alignment between the ACT CRS-S and the Kentucky CCA 4.1 high school science standards.

Table 4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria as Rated by Six Reviewers for the Kentucky CCA 4.1 High School Science Standards to the ACT College Readiness Standards-Science

Science Standards to ACT College Readiness	***Alignment Criteria***			
Subdomains	*Categorical Concurrence* (Avg. Hit)	*Depth-of-Knowledge Consistency* (Percent at and above)	*Range of Knowledge* (Percent of Kentucky Standards with ACT Standards)	*Balance of Representation* (Index Value)
1. - PHYSICAL SCIENCE	8.17	4	38	0.77
2. - EARTH/SPACE SCIENCE	2.33	0	21	0.90
3. - BIOLOGICAL SCIENCE	3.67	0	33	0.91
4. - UNIFYING CONCEPTS	3.83	23	14	0.78
Total	18	5	27	0.84

Alignment of Curriculum Standards and Assessments

Table 5 displays the number of items and points for each assessment form. In the analysis that follows, all items were assigned equal weight and a value of one.

Table 5
Number of Items and Point Value by Grade for Kentucky Assessments, Grade 12

ACT Assessment/Standard	Number of Items	Number of Multi-Point Items	Total Point Value
ACT Science62F	40	0	40
ACT Science63D	40	0	40
ACT Science63E	40	0	40
ACT Science64C	40	0	40

The results of the analysis for each of the four alignment criteria are summarized in Tables 6.1-6.4. More detailed data on each of the criteria are given in Appendix C, in the first three tables. With each table and for each grade, a description of the satisfaction of the alignment criteria for the given grade is provided. The reviewers' debriefing comments provide further detail about the individual reviewers' impressions of the alignment.

In Tables 6.1-6.4, "YES" indicates that an acceptable level was attained between the assessment and the subdomain on the criterion. "WEAK" indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. "NO" indicates that the criterion was not met by a noticeable margin—10% over an acceptable level for Depth-of-Knowledge Consistency, 10% over an acceptable level for Range-of-Knowledge Correspondence, and .1 under an index value of .7 for Balance of Representation. If the assessment was judged to have fewer than six items for a subdomain, then “NO” is entered for Categorical Concurrence and the average number of items assigned to the subdomain by the reviewers is given in parentheses. If the assessment had fewer than two items that measured content related to the subdomain, the subdomain was considered as not being assessed. If the subdomain had too few of corresponding items to be considered assessed, then the values for the other three alignment criteria could not be accurately computed and were considered “Not Applicable” (NA).

The analysis indicates there is poor alignment between the Kentucky CCA 4.1 high school science standards and the four forms of the ACT-science assessments. Reviewers made clear that both the Kentucky science standards and the ACT-science assessments were good. The Kentucky standards expected students to do important science. The ACT-science assessments measured important areas in science that students should be able to do. However, the ACT-science assessments were confined primarily to the measuring students’ knowledge of scientific processes which were only a part of the Kentucky standards.

The greatest alignment issues between the Kentucky CCA 4.1 high school science standards and the ACT-science assessment were with the DOK levels of the assessment items and the Range-of-Knowledge Correspondence. The ACT-science assessments only addressed from 5% to 41% of the standards under the four subdomains. Three forms of the assessment were judged not to have a sufficient number of items (six or more) to make a reliable judgment of students’ proficiency of Subdomain 2 (Earth and Space Science). Form 62F was judged to have too few items targeting Subdomain 4 (Unifying Concepts). The alignment between the Kentucky science standards and the ACT-science assessments did vary by form. But overall, the ACT-science assessments targeted too few of the standards under each of the four subdomains and the complexity of too many of the assessment items were lower than the complexity expected by the Kentucky standards. The alignment with each assessment form is discussed in more detail below.

Science Standards to ACT Science 62F

The alignment of the Kentucky CCA 4.1 high school science standards and Form 62F of the ACT-science assessment was low. Even though the assessment had a sufficient number of items that had content related to three of the four subdomains (1, 2, and 3), reviewers only found assessment items that corresponded to one to four of the standards under each domain. An acceptable level for Categorical Concurrence was not reached for Subdomain 4 (Unifying Concepts). Reviewers only found four items that mapped to standards under this domain, below the six items required to have an acceptable level. The Depth-of-Knowledge Consistency criterion acceptable level was only reached by Subdomain 3 (Biological Science). For the other three subdomains, less than 25% of the corresponding items had a DOK level that was comparable to the DOK level of the targeted standard.

An acceptable level for the Range-of-Knowledge Correspondence of 50% of the underlying standards with at least one corresponding item for a subdomain was not attained for any of the four subdomains. Only 5% (Subdomain 4) to 35% (Subdomain 1) of the underlying standards had a corresponding item. The Balance was acceptable for all subdomains. That is, of the standards targeted under a subdomain the items did not overemphasize any one of the standards. A total of 20 items (50% of the items on the assessment) would need to be replaced to attain full alignment—seven items for Subdomain 1, four items for Subdomain 2, two items for Subdomain 3, and seven items for Subdomain 4.

Table 6
Summary of Acceptable Levels on Alignment Criteria for Kentucky CCA 4.1 High School Science Standards and the ACT-Science Assessment

Table 6.1
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria Kentucky CCA 4.1 High School Science Standards to ACT-Science Assessment Form 62F

Science Standards to ACT Science Form 62F	***Alignment Criteria***			
Subdomains	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
1. - PHYSICAL SCIENCE	YES	NO	NO	YES
2. - EARTH/SPACE SCIENCE	YES	NO	NO	YES
3. - BIOLOGICAL SCIENCE	YES	YES	NO	YES
4. – UNIFYING CONCEPTS	NO (4.17)*	NO	NO	YES

*NO (4.17) Insufficient number of items to meet the Categorical Concurrence acceptable level. The number of parentheses is average number of items.

Science Standards to ACT Science 63D

The alignment between the Kentucky CCA 4.1 high school science standards and Form 63D of the ACT-science assessment also was found to be low. Form 63D did have a sufficient number of items for Subdomains 1, 3, and 4, but was judged to only have 4.67 items on the average as corresponding to Subdomain 2 (Earth and Space Science). The assessment and none of the four subdomains adequately attained an acceptable level for the Depth-of-Knowledge Consistency criterion, except the criterion was nearly met for Subdomains 1 and 4. The assessment only had items that targeted one or two of the standards (12% to 27%) under each subdomain, too few to have an acceptable level for Range. All of the items that did correspond to Kentucky standards were adequately distributed among the standards that did have a corresponding item. A total of 19 items would need to be replaced to attain full alignment—three items for Subdomain 1, three items for Subdomain 2, seven items for Subdomain 3, and six items for Subdomain 4.

Table 6.2
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria Kentucky CCA 4.1 High School Science Standards to ACT-Science Assessment Form 63D

Science Standards to ACT Science Form 63D	***Alignment Criteria***			
Subdomains	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1. - PHYSICAL SCIENCE	YES	WEAK	NO	YES
2. - EARTH/SPACE SCIENCE	NO	NO	NO	YES
3. - BIOLOGICAL SCIENCE	YES	NO	NO	YES
4. - UNIFYING CONCEPTS	YES	WEAK	NO	YES

Science Standards to ACT Science 63E

The alignment between the Kentucky science standards and Form 63E of the ACT-science assessment varied some from the prior to forms, but was still considered low. Reviewers found a sufficient number of items only for Subdomains 1 and 4. The assessment only had four items that corresponded to Subdomain 2 (Earth and Space Science) and five items that corresponded to Subdomain 3 (Biological Science). The Depth-of-Knowledge Consistency criterion was not met by any of the four subdomains. The percentage of items corresponding to each subdomain that had a DOK level that had a DOK level that was the same or higher than the corresponding standard ranged from 3% (Subdomain 3) to 32% (Subdomain 1). The Range criterion was not met for any of the four subdomains with from 9% (Subdomain 4) to 33% (Subdomain 1) of the underlying standards targeted by assessment items. As for the other forms, the assessment items did have good Balance and did not over emphasize any of the standards. A total of 17 items would need to be replaced to attain full alignment—four items for Subdomain 1, three items for Subdomain 2, three items for Subdomain 3, and seven items for Subdomain 4.

Table 6.3
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria Kentucky CCA 4.1 High School Science Standards to ACT-Science Assessment Form 63E

Science Standards to ACT Science Form 63E	***Alignment Criteria***			
Subdomains	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
1. - PHYSICAL SCIENCE	YES	NO	NO	YES
2. - EARTH/SPACE SCIENCE	NO	NO	NO	YES
3. - BIOLOGICAL SCIENCE	NO	NO	NO	YES
4. - UNIFYING CONCEPTS	YES	NO	NO	YES

Science Standards to ACT Science 64C

The alignment between the Kentucky science standards and Form 64C of the ACT-science assessment was a little better than for the other three forms, but was still considered as low. Reviewers only coded one item to a standard under Subdomain 2 (Earth and Space Science). This is too few of items to consider that the assessment form measured any content knowledge related to Subdomain 2. The other criteria were not computed for Suddomain 2. The assessment did have an acceptable level on the Depth-of-Knowledge Consistency criterion for the other three subdomains. Reviewers judged that some of the assessment items had a DOK level 3 and the assessment targeted more standards with a DOK level 2. Form 64C, as for the other forms, was still too narrow in content when compared to the Kentucky standards. The percentage of standards under each subdomain targeted by the assessment varied from 13% (Subdomain 4) to 41% (Subdomain 1). The Balance was good. A total of 14 items would need to be replaced to attain full alignment—only one item for Subdomain 1, five items for Subdomain 2, two items for Subdomain 3, and six items for Subdomain 4.

Table 6.4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria Kentucky CCA 4.1 High School Science Standards to ACT-Science Assessment Form 64C

Science Standards to ACT Science Form 64C	***Alignment Criteria***			
Subdomains	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
1. - PHYSICAL SCIENCE	YES	YES	WEAK	YES
2. - EARTH/SPACE SCIENCE	NT (1.0)	NA	NA	NA
3. - BIOLOGICAL SCIENCE	YES	YES	NO	YES
4. - UNIFYING CONCEPTS	YES	YES	NO	YES

Summary of Items Coded to Standards Across Forms

Table 7 summarizes the items coded by at least two reviewers to a standard for each of the four science forms.

Table 7
Summary of Assessment Items Mapped to Kentucky Core Content for Assessment 4.1 Standards for High School Science by Two or More Reviewers for Four Forms of the ACT-Science Assessment

Standard	Grp DOK	ACT Science Test 62F					ACT Science Test 63D					ACT Science Test 63E					ACT Science Test 64C				
1	3																				
		34-(4)	35-(5)	36-(5)	37-(5)	38-(5)	19-(4)	20-(4)	21-(4)	22-(4)	23-(4)	31-(6)	32-(6)	33-(6)	34-(6)	35-(6)	36-(6)	37-(6)	38-(6)	39-(6)	
1.1.	2	39-(5)	40-(5)																		
1.1.1b	2																18-(4)	19-(6)	20-(6)	21-(5)	22-(6)
1.1.5a	2																				
1.1.6	3																				
1.1.6a	1																				
		28-(5)	29-(5)	30-(5)	31-(5)	32-(5)						6-(5)	7-(5)	8-(5)	9-(5)	10-(5)					
1.1.6b	3	33-(5)	34-(2)									11-(4)									
1.1.7a	2																				
1.1.8	3																				
1.1.8a	2																40-(6)				
1.1.8b	3	33-(6)																			

Table 7 (continued)

Summary of Assessment Items Mapped to Kentucky Core Content for Assessment 4.1 Standards for High School Science by Two or More Reviewers for Four Forms of the ACT-Science Assessment

Standard	Gp DOK	ACT Science Test 62F					ACT Science Test 63D					ACT Science Test 63E					ACT Science Test 64C				
		1-(6)	2-(6)	3-(6)	4-(6)	5-(5)	13-(6)	14-(6)	15-(6)	16-(6)	17-(6)	25-(6)	26-(6)	27-(6)	28-(6)	29-(6)					
		6-(5)	7-(5)	8-(5)	9-(5)	10-(5)	18-(6)	19-(2)	20-(2)	21-(2)	22-(2)	30-(6)									
1.2	3	11-(5)	12-(5)				23-(2)														
1.2.1	3																				
1.2.1a	2																				
												36-(6)	37-(6)	38-(5)	39-(5)	40-(4)	6-(6)	7-(6)	8-(6)	9-(6)	10-(6)
1.2.1b	3																11-(5)				
1.2.2a	3																				
2	3																				
2.3	3	13-(6)	14-(6)	15-(6)	16-(6)	17-(6)	24-(5)	25-(5)	26-(5)	27-(5)	28-(5)	36-(4)	37-(4)	38-(4)	39-(4)	40-(5)					
2.3.1a	3																				
2.3.2a	2																				
2.3.3a	2																				
2.3.6	3																				
2.3.6a	2																				
2.3.6b	3	23-(6)	24-(6)	25-(5)	26-(5)	27-(6)	6-(3)														
2.3.7a	3																				
2.3.8b	3																				

Table 7 (continued)

Summary of Assessment Items Mapped to Kentucky Core Content for Assessment 4.1 Standards for High School Science by Two or More Reviewers for Four Forms of the ACT-Science Assessment

Standard	Gp DOK	ACT Science Test 62F					ACT Science Test 63D					ACT Science Test 63E					ACT Science Test 64C				
3	3																				
		28-(5)	29-(5)	30-(5)	31-(5)	32-(4)											23-(5)	24-(5)	25-(5)	26-(5)	27-(5)
		33-(3)	34-(5)	35-(6)	36-(6)	37-(6)															
3.4	2	38-(6)	39-(6)	40-(6)																	
3.4.1a	3																				
3.4.3a	2																				
3.4.5	3																				
3.4.5a	2																				
							7-(6)	8-(6)	9-(6)	10-(6)	11-(6)						28-(6)	29-(6)	30-(6)	31-(6)	32-(6)
3.4.5b	3						12-(6)										33-(6)	34-(6)			
3.4.7b	2																				
		18-(5)	19-(5)	20-(5)	21-(5)	22-(5)	1-(6)	2-(6)	3-(6)	4-(6)	5-(6)										
							6-(4)	36-(2)	37-(2)	38-(3)	39-(2)										
3.5	3						40-(2)														
3.5.1a	3																				
3.5.2b	3											1-(6)	2-(6)	3-(6)	4-(6)	5-(6)					
4	3																				

Table 7 (continued)

Summary of Assessment Items Mapped to Kentucky Core Content for Assessment 4.1 Standards for High School Science by Two or More Reviewers for Four Forms of the ACT-Science Assessment

Standard	DOK	ACT Science Test 62F					ACT Science Test 63D					ACT Science Test 63E					ACT Science Test 64C				
		23-(5)	24-(5)	25-(5)	26-(5)	27-(5)	24-(5)	25-(4)	26-(4)	27-(4)	28-(4)										
4.6	3						34-(2)	36-(2)	37-(2)	39-(2)	40-(2)										
4.6.1a	3																				
4.6.2a	3																				
4.6.2b	2																1-(6)	2-(6)	3-(6)	4-(5)	5-(5)
4.6.4a	3																				
4.6.5a	3																				
							29-(4)	30-(4)	31-(3)	32-(4)	33-(4)										
4.6.7b	2						34-(4)	35-(4)													
4.6.8a	3																				
4.6.9a	3																				
4.6.10	3																				
4.6.10a	1																				
4.6.10b	3																				
4.6.11a	2																				
												12-(6)	13-(6)	14-(6)	15-(6)	16-(6)	12-(4)	13-(4)	14-(4)	15-(4)	16-(4)
4.7	3											17-(6)	18-(6)				17-(3)				
4.7.1a	3																				
												19-(2)	20-(2)	21-(2)	22-(2)	23-(2)					
4.7.2b	3											24-(2)									
4.7.3a	3																				
4.7.5a	3																				

Source of Challenge Issue Comments

Reviewers were instructed to comment about any items that contained an inappropriate source of challenge. Their comments can be found in Tables (grade).5 in Appendix D. Only one reviewer indicated any source-of-challenge issues for any item. One reviewer commented on item 23, Form 63E, indicating a potential issue in reading a diagram. On Form 64C one reviewer identified a mismatch between the Kentucky standard and five items (28, 29, 30, 33, and 34). This mismatch was not a source-of-challenge issue with the items, but rather an issue with the relationship of the items to the Kentucky standards. In general, reviewers found the items to be in good form.

Reviewers' Comments

Along with source-of-challenge issue comments, reviewers were asked to provide any other notes they may have. These comments can be found in Tables (grade).7 in Appendix D. Reviewers were required to write a comment for any item they assigned to a generic standard. Since reviewers assigned a relative large number of items to generic standards there are several comments by reviewers. These notes explain in more detail why an item did not match any of the Kentucky science standards.

After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the comments made by the reviewers are given in Appendix E. The debriefing notes are reviewers' summary describing to what degree the assessment targeted the most important content in the standards and was at an appropriate depth-of-knowledge level. The summary debriefing comments describe in some detail the process that the group of science reviewers used to divide the Kentucky standards into process standards and content standards to communicate better the relationship of the ACT-science assessment to the Kentucky CCA 4.1 high school science standards. One reviewer summarized the opinion of the group, "It is the opinion of the Science alignment team that the ACT, although a valid predicting tool for college preparedness, is not an appropriate assessment to measure master of the KY Core Content of Science as expressed in their standards."

Reliability Among Reviewers

The overall intraclass correlation among the six science reviewers' assignment of DOK levels to items was reasonable for six reviewers for ACT-science assessments (Table 8). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. On the science assessment, the reviewers' assignment of a DOK level to each item was so similar that there was very low variance among the reviewers. In these cases (Forms 62F and 63E), the pairwise comparison is more appropriate to use to judge reviewer consistency. Overall, all of the reliability measures of reviewers' assignment of DOK levels to assessment items and the ACT CRS-S standards was .69 or higher. These values are considered reasonable for six reviewers.

A pairwise comparison is used to determine the degree of reliability of reviewers coding at the standard level and at the subdomain level. Both the standard and subdomain pairwise comparison values are reasonable and comparable to those for other alignment studies. The pairwise agreements for assigning items to Kentucky standards are a little lower than for some alignment studies because reviewers assigned items to more than one standard. Reviewers can have exact agreement on the primary standard, but can disagree on the secondary standard.

Table 8
Intraclass and Pairwise Comparisons of the Assigned Level of Complexity, Standard, and Subdomain for the Kentucky Alignment Analysis of the High School Science Standards with the ACT College Readiness Standards-Science and Four Forms of the ACT-Science Assessment

Standard to ACT Assessment	Intraclass Correlation	Pairwise Comparison:	Pairwise: Standard	Pairwise: Subdomain
Science to ACT Science 62F	.56	.69*	.76	.83
Science to ACT Science 63D	.82	.85	.68	.74
Science to ACT Science 63E	.74	.78*	.78	.84
Science to ACT Science 64C	.82	.73	.82	.86
Science to ACT College Readiness	.98	.96	.76	.80

***** The pairwise comparison was used rather than the Intraclass Correlation

Summary

A three day alignment institute was held in Frankfort, Kentucky, on May 18-20, 2007, to analyze the alignment between the Kentucky Core Content for Assessment (CCA) 4.1 high school science standards and four forms of the ACT-science assessment. The Kentucky science standards were also compared to the ACT College Readiness Standards-Science (ACT CRS-S). The six reviewers included science content experts and science teachers. Three of the reviewers were from Kentucky and three were from other states. They analyzed the agreement between the Kentucky standards and four forms (62F, 63D, 63E, and 64C) of the ACT-science assessment.

The results indicated that there was poor alignment between the Kentucky CCA 4.1 high school science standards with the ACT CRS-S and with the ACT-science assessments. Reviewers complemented the Commonwealth of Kentucky on very comprehensive and demanding standards. They also indicated that the ACT CRS-S stated and the assessments measured important scientific knowledge. However, the ACT documents attended primarily to science process skills—interpretation of data, scientific investigations, and evaluation of models—rather than content knowledge of different science areas expected in the Kentucky standards (physical science, earth and space science, and biological science).

Reviewers only found 10 of 47 ACT CRS-S standards (21%) that matched any of the Kentucky science standards. These ACT standards corresponded to about 30% of the Kentucky science standards. The ACT CRS-S also had lower depth-of-knowledge levels than expected by the Kentucky science standards. Considering the content in both the ACT CRS-S and the Kentucky CCA 4.1 high school science standards, only about 13% of the content was in common to both documents, 40% exclusive to the Kentucky science standards, and 47% exclusive to the ACT CRS-S.

The alignment between the Kentucky CCA 4.1 high school science standards and the ACT-science assessments was not much better than for the ACT CRS-S. Each of the four assessment forms did not include enough items for one or two of the four Kentucky subdomains. The subdomain without at least six corresponding items varied by assessment form among Subdomain 2 (Earth and Space Science), Subdomain 3 (Biological Science), and Subdomain 4 (Unifying Concepts).

The items that did correspond to the subdomains were generally lower in depth-of-knowledge level than the ceiling level expected by the Kentucky standards. Most of the ACT-science assessment items had a DOK level 2 while over 60% of the Kentucky standards were assigned a DOK level 3. Only one assessment form, 64C, met the Depth-of-Knowledge Consistency criterion for three of the four Kentucky subdomains. Form 62F met the criterion for only one subdomain while the other two forms did not satisfy the DOK criterion for any of the subdomains. None of the four assessment forms had sufficient coverage of the standards under the subdomains to satisfy the acceptable level for the Range-of-Knowledge Correspondence criterion for any of the four subdomains.

Overall, from 14 to 20 items on the 40 item assessment forms would need to be replaced to attain full alignment between the ACT-science assessment and the Kentucky high school science standards. The ACT documents, designed as a prediction tool for success in college, have a different and very limited purpose than to assess or communicate all that students should know about science upon leaving high school as included in the Kentucky high school science standards.

References

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement, 25*(1), 47-55.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.