



# Observation of the Spring 2017 Kentucky Academic Standards (KAS) Science Field Test Administration

**Prepared  
for:** Kentucky Department of Education  
300 Sower Blvd.  
Frankfort, KY 40601

**Authors:** Arthur A. Thacker  
Brittany F. Crawford  
Andrea L. Sinclair

**Date:** June 9 , 2017

# Observation of the Spring 2017 Kentucky Academic Standards (KAS) Science Field Test Administration

## Table of Contents

Introduction .....	2
Background Information .....	2
Method .....	3
Results .....	4
Conclusions .....	8

## List of Tables

Table 1. Test Completion Times for the Kentucky Academic Standards (KAS) Science Field Test.....	6
--	---

# Observation of the Spring 2017 Kentucky Academic Standards (KAS) Science Field Test Administration

## Introduction

The purpose of this study was to (a) observe test administration of a small sample of the Kentucky Academic Standards (KAS) Science Field Test, (b) conduct focus groups with student test-takers, and (c) interview test administrators to gain insight into possible areas to improve upon the field test.

## Background Information

Kentucky teachers, with the support of Kentucky Department of Education (KDE) science content experts, developed the extended response and multiple choice items for the KAS Science Field Test. The items center around complex, cross-discipline science phenomena. One phenomenon may have several associated items. The items were often formatted differently from those that appeared on previous state assessments. Many multiple-choice items had more than the typical four answer choices and, in some cases, there was more than one correct response. Extended response items were presented as fill-in-the-blank, short response, multiple parts, prodded, or diagrams. Extended response and multiple choice items were divided into clusters based on a phenomenon.

The clusters represent real-world phenomenon. Phenomenon were chosen to purposefully cross content areas. This allowed Kentucky's item writers to address cross-cutting concepts directly on the assessment. Each test form included two phenomena, each with several test items. Twelve test forms were field tested to generate a large number of potential operational phenomena and items.

The KAS Science Field Test was administered to all Grade 4, 7, and 11 students, except those taking an Alternate Assessment. The test was administered between March 13<sup>th</sup> - 17<sup>th</sup>, 2017 for students in Grades 4 and 7 and between March 6<sup>th</sup> - 27<sup>th</sup>, 2017 for students in Grade 11.

The KAS Science Field Test was presented to students in a pre-printed, combined booklet (students respond to items directly in test book). Test administrators provided students with necessary testing materials (pencils, scratch paper, erasers, and calculators) and then read directions aloud from the KAS Science Field Test manual. Verbal prompts that encouraged students to ask for clarification were given after each section of the directions. A student honor code and area for required signature were printed on the back page of the student combined booklet. While the administrators read aloud the honor code, students were asked to sign the honor code before testing could commence.

Test administrators directed students to the Extended-Response Questions General Scoring Guide for review. Next, directions were given on how to answer the extended response items that had multiple parts (e.g. Parts A and B). After explaining the extended response items, teachers reiterated correct and incorrect methods for marking answer choices, two of the most important guidelines being—completely fill in the circle corresponding to the correct answer choice for multiple-choice questions and do not make any stray marks on any page of the student combined booklet.

Students were informed that they had 70 minutes to complete the questions from both clusters, and they were not required to stop between clusters. They were also informed that some of the multiple-choice items would require them to choose more than one answer. A picture of a stop sign at the end of the booklet served as a reminder for students that testing was complete at the end of the second set of questions.

Students completed a Student Survey as a class after the last student finished testing. They were informed that survey responses were not a factor in test scores nor would they be identified by individual student. Afterward, all test materials, including the combined test booklets and scratch paper, were collected by the test administrator.

## Methodology

### Procedure

This study was designed as an observational investigation of the administration of the KAS Science Field Test in Grades 4, 7 and 11, followed by an interview with test administrators and focus groups with student test-takers. Two HumRRO researchers participated in each visit to each selected school. For the interviews and focus groups, one researcher asked questions, while the second researcher served as the primary note taker.<sup>1</sup> Upon completing each site visit, the researchers compared notes to ensure that the key themes were captured from the interview with the test administrator and the focus group with the students. Findings were content analyzed and themes identified.

### Sample

Schools were selected according to the percentage of students scoring Proficient or Distinguished on the 2015-16 Kentucky Performance Rating for Educational Progress (K-PREP) assessment. Specifically, 2015-2016 Mathematics K-PREP data was used for elementary and middle schools and 2015-2016 Writing K-PREP data was used for high schools. Schools were organized by grade level (i.e., elementary, middle, high) and ranked in ascending order according to the percentage of students scoring Proficient or Distinguished at each school. Two schools were chosen from each grade level (for a total of six schools)—one with approximately 25% scoring proficient or higher (indicating a school with room for growth) and one with approximately 75% scoring proficient or higher (indicating a school with good test performance). Schools at the extreme low and high ends of the range were not selected.

Selected schools were relatively large (at least 50 or more students per grade level). Alternative schools and other specialized schools were not considered during selection. Schools were selected from Jefferson County and other surrounding counties, including Bullitt, Oldham, and Shelby Counties.

### Instruments

**Observation Protocol.** Researchers completed an observation protocol while observing the test administration. The observation protocol included sections on (a) basic descriptive information (e.g., school name, date of observation, etc.), (b) clarity of test

---

<sup>1</sup> There was one exception to this; for one of the high schools, only the Project Director (PD) conducted the site visit. This was one of the last schools to be visited; consequently, the PD felt comfortable conducting this visit on his own.

directions, (c) test administration, and (d) general observations. The observation protocol is provided in Appendix A.

***Test Administrator Interview Questions.*** Following test administration, researchers interviewed the test administrators about their test administration experience. Interview questions covered topics such as, effectiveness of the training they received for administering the test, clarity of the Test Administration Manual, and types of questions asked by students. The test administrator interview questions are provided in Appendix B.

***Student Focus Group Questions.*** Following test administration, researchers conducted a brief focus group with the students that just completed the test. The focus group questions asked students to share their overall impression of the test as well as specific questions about the test (e.g., particular questions they found particularly confusing or difficult). The student focus group questions are provided in Appendix C.

## ***Results***

In the results that follow, findings from the observation of test administration, interviews with test administrators, and student focus groups are organized into the following categories:

- Student Questions about the Test
- Test Administrators' Understanding of How to Administer the Test
- Appropriateness of the Time Allotted for Testing
- Student Confusion about Test Content
- Student Confusion about Test Organization/Formatting
- Student Engagement during Testing

### ***Student Questions about the Test***

Throughout the verbal delivery of the KAS Science Field Test directions, students were given the opportunity to ask for further clarification or ask questions that might not have been covered by the script in the test administration manual. Grade 4 students frequently asked, "What science content does this test cover?" Grade 4 students also seemed to have difficulty understanding the difference between the KAS Science Field Test and K-PREP assessments. Although the method for marking correct answers was described during the verbal directions, this still caused confusion with several of the Grade 4 students. Students seemed confused as to why they were unable to make stray marks in the booklet. One additional aspect that puzzled students was whether they needed to "bubble in" the correct answer choice within the booklet or on a separate form, considering that they were instructed not to mark in the booklet. Many questions related to test organization and formatting, such as, "How many questions are there?" "How long will this take?" and "How long is each section?" Several students asked if there was a Part A, B, and C because they did not understand what the word "cluster" meant.

Grade 7 students' initial questions focused on answer selection and test organization and formatting, although there were fewer questions at the middle school level. Students asked if they could write their answers in the booklet (could they circle answer choice A, B, or C?) or

should they skip to another page in the booklet to write their answers. Students also asked whether they should continue to the second cluster upon completion of the first cluster, or if they should stop after completion of the first cluster.

Questions from Grade 11 students also focused on test content and test organization and formatting (e.g., “How many questions will be on the test?”). Also, Grade 11 students were unsure when they were to use their calculators. During the observation of Grade 11 students’ test administration, little to no calculator usage was observed, as students seemed unsure when they were permitted to use their calculators or why they were allowed at all. Many test forms required little or no computation.

### ***Test Administrators’ Understanding of How to Administer the Test***

For the most part, test administrators described little difficulty in navigating the corresponding manual to deliver verbal directions to students. Most described the Science Field Test administration as comparable to the administration of the K-PREP assessments.

Test administrators suggested that the verbal directions be moved to the front of the manual as it was difficult to locate during the test administration. Test administrators indicated that it was also difficult to determine the location of the student signature for the code of conduct agreement (on the back of the booklet), and it was confusing as to when the students were supposed to complete the Student Survey (individually as they finished testing or as a class when the last student has finished testing?).

Test administrators suggested that it would be helpful to know how many questions were on the test, as well as the structure and content covered by the test. They indicated that this would help them answer several of the students’ initial questions. In particular, for Grade 11, the manual didn’t explain how or when the calculators were to be used and this left test administrators and students unsure of why/if they were needed.

A few of the test administrators did not understand where the students should stop working (between clusters or at the end of the test). This confusion contributed to the students’ confusion regarding breaking between clusters and the stop sign at the end of the test booklet.

Test administrators asked that the terms “combined booklet” and “cluster” be reconsidered for Grade 4 students, as these students spent a large amount of time trying to understand what these words and phrases meant rather than focusing on the subsequent directions. In general, Grade 4 test administrators felt that there were too many directions to deliver prior to testing.

### ***Appropriateness of the Time Allotted for Testing***

To help determine if 70 minutes is an adequate amount of time for Grade 4, 7, and 11 students to complete both clusters of the Science Field Test, the researchers noted the time at which students completed the test. Across the schools visited, Table 1 displays the average number of minutes it took for the first 2-3 students to complete both clusters and the average number of minutes it took for all but 2-3 students to complete both clusters across the grade levels. The amount of time necessary for students to complete the Science Field Test appears to decrease from Grade 4 to Grade 11. It should be noted that for both of the observed elementary schools there were more than three students in Grade 4 who had not completed the test at the end of 70 minutes. Furthermore, during the focus groups some students (across grade levels) reported not having time to complete their tests because they were taking their time and answering each

item carefully. Two students with a reader accommodation were observed. Both students finished in approximately 55 minutes.

**Table 1. Test Completion Times for the Kentucky Academic Standards (KAS) Science Field Test**

<b>Grade Level</b>	<b>Completion</b>	<b>Minutes<sup>2</sup></b>
<b>Grade 4</b>	<b>First 2-3 students completed both clusters</b>	<b>44</b>
	<b>All but 2-3 students completed both clusters</b>	<b>More than 3 students had not completed the test within 70 minutes</b>
<b>Grade 7</b>	<b>First 2-3 students completed both clusters</b>	<b>30</b>
	<b>All but 2-3 students completed both clusters</b>	<b>65</b>
<b>Grade 11</b>	<b>First 2-3 students completed both clusters</b>	<b>22</b>
	<b>All but 2-3 students completed both clusters</b>	<b>55</b>

### **Test Content**

For several of the test items, vocabulary proved to be problematic for Grade 4 students. Many of the test administrators and students felt the test items contained vocabulary unfamiliar to Grade 4 students. For example, “phenomenon” was frequently described by both test administrators and Grade 4 students as problematic for this grade level. During the verbal directions, students were told that items were divided into clusters based on phenomenon; however, many students didn’t know what this word meant and this created a distraction during the remainder of the directions. Other problematic vocabulary for Grade 4 students included: claim (often associated with English content rather than science content), clusters, consistency, influence, and supported<sup>3</sup>. Although students may have understood what the test item was asking them to do, word choice made it difficult for some students to move past an initial reading of the question.

During the student focus group, students were asked if there were specific items with which they struggled. Many of the Grade 4 students described having trouble understanding test items asking them to interpret bar graphs. It also was unclear at times how many questions were associated with each given passage—that is, the distinction between sets of items was sometimes unclear.

Although the phenomena within the clusters were designed to create a storyline for groups of items, Grade 7 students described the “stories” (passages) in the booklet as unclear. It was mentioned that several of the passages grouped together in a cluster were unrelated to one

<sup>2</sup> Caution should be taken in interpreting these averages. They represent only two schools and should be considered as rough estimates only.

<sup>3</sup> Specific science content vocabulary is not included here, as it represents an intentional part of the measured construct and including those terms might expose test item content.

another, although it should be noted that each form included two unrelated phenomena, so the students may have mistakenly assumed that the phenomena should both have come from the same content domain.

Grade 11 students felt there was not enough information to correctly answer many of the items. A couple of the extended response questions were described as vague. This may be an indication that the students are accustomed to science assessments where they only interpret text and data, rather than recall content learned in class. Discussion with students did not clarify this issue.

Some students noted that on a few multiple choice items, answer choices were very similar (only one word difference among response options). This made selecting a correct answer tricky for several of the students, as selecting a response became more of a reading/attention issue. Students reported trying to tease out the differences in response options rather than thinking about the best answer choice. This was also particularly difficult for students with read aloud accommodations because all answer choices sounded the same and test administrators were unable to do voice modulation.

Another issue that seemed to prevent students at each grade level from (quickly) providing a correct answer was asking students to draw a picture. A couple of items asked students to illustrate or draw their answers. Students who felt they didn't possess the ability to draw these objects worried they didn't answer the question effectively or accurately.

In the Student Survey, there was one item that asked about the student's understanding of the test items. This was presented as a "yes or no" question, but some students stated that they wanted to indicate that they understood some of the items but not others.

### ***Test Organization/Formatting***

Regardless of grade level, a concern frequently mentioned was confusion about where to answer Parts A and B of the extended response items. One Grade 4 student suggested a single block of space should be provided for students with a clear indication of where to begin answering Part B. One Grade 11 student suggested providing more space for each part initially, rather than having the student go back to write in the space for Part A. Other Grade 11 students mentioned that the large amount of space provided for the extended response items caused anxiety because they felt they needed to fill up the entire section. One test administrator commented that asterisks should not be used in the extended response items for Grade 4 because Grade 4 students do not understand what the symbols are supposed to indicate.

Across grade levels, students reported being confused about being able to select more than one correct answer. Not all students were clear on when they were supposed to "bubble in" more than one answer choice. Test administrators had to make announcements to remind students that they needed to bubble in correct answers only, as several students were crossing out incorrect answers, too. Crossing out incorrect response options is a common test taking strategy for the K-PREP assessments, where students' choices are recorded on an answer sheet and the test booklet is not scanned at all. Students consistently reported that the most difficult questions were the multiple-choice items where students had to select three answers. One test administrator suggested that it might be good to add a practice problem to demonstrate this item format.

There were a couple of test items that prompted students to compare graphs; however, students said the figures weren't labeled, titled, or numbered so it was difficult to know exactly which figures were being compared.

It was not uncommon for students at each grade level to stop after completing the first cluster because they were unclear about where the stop sign was located or if they were required to break before moving on to the next section of the test. Grade 7 students also mentioned rushing through a portion of the test because, with the extra pages in the back of the test booklet, they assumed there were more items to complete than there were.

A Grade 7 test administrator expressed concern that students were unable to take notes and leave other marks in their test booklets. Not being able to leave comments in the margins, underline key words and phrases, or cross out incorrect answers changes the test-taking strategies that students have been taught to use since they started state testing.

For many students, having to flip back and forth in the test booklet was a source of frustration. Several students described flipping back and forth between the scenario and items as confusing and time consuming.

Finally, it was also suggested that the Student Survey be provided on a separate page, so that booklets could be collected and students would be able to read after finishing their tests.

### ***Student Engagement During Testing***

Overall, test administrators and students provided positive feedback for the first administration of the KAS Science Field Test. Comments from test administrators included, "I like the way the science test is going, it creates deep thinking," "Presenting content as a story is an enormous improvement. It feels less like a test to the kids. Without stories and analogies, kids don't learn," and "This is a good starting point. I'm impressed." The variety in the test items and tasks the students were asked to complete was also described as an improvement in the KAS Science Field Test, as this variety peaks the interests of many different students and improves students' overall level of engagement. Item scenarios and situations were also described as "relatable." Students reported having to use their critical thinking skills to answer the questions on the test.

There was, however, some indication of student frustration during testing. For example, students expressed frustration about having to flip back and forth within the test booklet to answer questions; they reported that it was frustrating and time consuming. Also, Grade 4 test administrators feared that students were using too much "cognitive energy" trying to understand the directions given prior to the test. This is especially problematic for poor readers who were asked to read the directions silently, review the scoring guide, and then come forward with any questions they might have. Finally, one of the most common complaints from test administrators and students was the number of extended response questions a student was expected to complete in one sitting. This was described as a large amount of writing for many students. One test administrator allowed students to take quick "brain breaks" to prevent frustration. Students were often observed stopping to rub their hands apparently due to fatigue from writing.

### **Conclusions**

Overall, the KAS Science Field Test administration appears to have gone smoothly based on the limited number of site visits included in this study. Test administrators generally reported

that the test administrator manual was easy to navigate, and both test administrators and students reported that the test required students to use critical thinking skills.

There were also some findings that may indicate areas for potential improvements. However, because this study was based on a small sample (two elementary schools, two middle schools, and two high schools), the findings and recommendations that follow should be interpreted with this caveat in mind. Additional corroborating evidence should be sought prior to making substantive changes.

### *Summary of Findings and Potential Recommendations for Consideration*

- Test administrators indicated that the test administration manual was, overall, clear and easy to follow. However, moving the script to the front of the manual and clarifying the location for the student signature code of conduct agreement were noted as two possible areas for improvement.
- Students commonly asked test administrators how many questions were on the test and what content is covered by the test. **Recommendation:** It may be helpful to create a “Student FAQs” supplement to the test administration manual. In this instance, for the question, “How many questions are on the test?” the answer might be, “The test is designed to be completed within 70 minutes and includes \_\_\_ multiple-choice items and \_\_\_ extended response items.” An exact number or a range could be specified. For the question, “What does this test cover?” the answer might be, “this test is designed to assess your understanding of the science content on which you’ve received instruction” or similar.
- Across grade levels, there was some confusion and frustration about the formatting/organization of the test. For example, students were unsure of (a) whether they should stop between clusters, (b) whether they could make notes and cross-out incorrect response options in their test booklets, and (c) where to answer Parts A and B of the extended response items. Also, students expressed some confusion about the items with multiple correct answers and confusion about some of the terminology (e.g., cluster” vs. “part,”). They also exhibited some frustration with having to frequently flip back and forth in the test booklet to find the necessary information to answer test questions. Some of this confusion and frustration (which can introduce construct irrelevant variance) likely stems from the students’ unfamiliarity with the test format. **Recommendation:** Consider having students complete a practice test and/or tutorial prior to the actual test administration to familiarize them with the test format/organization before they take the operational test. Such a tutorial/practice test should also define key test terminology for students, such as “cluster,” “part,” “claim,” and “phenomenon.” It may also be helpful for test administrators to review the tutorial/practice as well, to help increase their familiarity with the test prior to administration.
- The reading load may be high for the typical fourth grade student. One test administrator indicated that students were expending too much “cognitive energy” trying to read and understand all the instructions. **Recommendation:** Look for ways to streamline the instructions so that they are more concise. Improving the conciseness of the instructions is likely to be beneficial for all grade levels, but may have the greatest benefit for Grade 4 students.

- Across both elementary schools, more than three students had not completed the test within the allotted 70 minutes. **Recommendation:** Consider expanding the time allowed for completing the test for Grade 4 and dividing the test into two separate sections.
- Grade 11 students expressed confusion about when they could use their calculators. **Recommendation:** Provide clearer guidance on when calculator use is allowed or structure items so calculators are not needed, except for specific accommodations.
- Both students and test administrators indicated that the number of extended response items that students were expected to respond to in a single setting was “a lot.” **Recommendation:** One test administrator allowed students quick “brain breaks” between the extended response items. This might be a useful strategy to implement for the operational test. We suggested earlier that the Grade 4 test might be split into two sections.
- Students indicated some concerns with the clarity of certain test items. For example, graphs that were to be compared were described as missing key labels/titles/numbering, and response options that differed by only one word were described as overly tricky (i.e., had more to do with ability to read carefully than one’s understanding of the content). Also, some students expressed concern that they would not score highly on the items that required them to draw a picture because of their poor drawing ability. **Recommendation:** Item statistics should be reviewed by group (e.g. graphing items, drawing items, etc.) for poor statistics (e.g., low p-values, negative point-biserials) and any anomalous items reviewed by content experts to determine whether items are functioning as intended. These analyses should include scrutiny of distracter performance as well as classical test statistics.

Finally, several interviewees mentioned areas for improvement regarding the student survey.

**Recommendation:** If the student survey were provided on a separate page, test booklets could be collected as students finished their tests, which would allow students to complete the assessment and then read silently as soon as they were finished. Also, there was one item that asked about the student’s understanding of the test items. This was presented as a “yes or no” question, but some students stated that they wanted to indicate that they understood some of the items but not others. Consider adding a “some” response option.

## Appendix A: Observation Protocol

# 2017 Kentucky Academic Standards Science Field Test Observation Form

<p>Date _____</p> <p>District/School _____</p> <p>Principal _____</p> <p>Type of Testing: Large Group (Arena style) or Classroom</p>	<p>Staff _____</p> <p>Grade _____</p> <p>Teacher _____</p> <p>Number of Students _____</p> <p>Accommodated: Yes or No</p>
<p><b>OBSERVATION OF THE DIRECTIONS</b></p>	<p><b>COMMENTS:</b></p>
<p>After the teacher read the directions, if students asked questions, capture the questions.</p>	
<p>Did the teacher understand the directions?</p>	
<p>Could the teacher answer the student's questions?</p>	
<p><b>OBSERVATION OF THE ADMINISTRATION</b></p>	<p><b>COMMENTS:</b></p>
<p>Record the start time of the test.</p>	
<p>Record the time when the first 2-3 students completed both clusters.</p>	
<p>Calculate the amount of time when the first 2-3 students began to finish.</p>	
<p>Record the time when all but 2-3 students have NOT completed both clusters.</p>	
<p>Calculate the amount of the time when all but 2-3 students had NOT completed both clusters.</p>	
<p>Observe if the students are using the information provided to answer the questions.</p>	

How many students asked the teacher to clarify the test questions?	
Were the students generally engaged?	
Did any of the students seem frustrated? Why?	
<b>GENERAL OBSERVATIONS</b>	<b>COMMENTS:</b>
Were there any comments about the construction of the clusters (readability, image quality, response space, etc.)?	
Capture any general comments made by the students or the teacher.	

## Appendix B: Test Administrator Interview Questions

1. Did you feel prepared to administer this test today? If yes, what contributed to this (e.g. training)? If no, what could have helped you feel more prepared (e.g., improvements to training)?
2. Was the Test Administration Manual clear and easy to follow? Please explain.
3. Did the instructions you read to students cover all the necessary information that students needed to take the test? If no, what was missing?
4. What are the most common questions students have been asking during test administration?
5. Do you feel that students have the right amount of time to finish the test?
6. Do you think you will be able to use the data/results from this assessment to inform your classroom instruction? How?
7. From what you've observed, do you feel the content on the test aligned with the content students were exposed to in their science curriculum this school year?
8. In your opinion, does the investment of time in preparing, administering, and scoring the test pay off for both students and teachers? How?
9. Have you administered the science test with accommodations for any students? If so, was the guidance clear on how to implement the accommodations?
10. Is there anything else you would like to share about the test?

## Appendix C: Student Focus Group Questions

1. Was there anything on the test that you didn't understand or that didn't make sense? If so, what was it?
2. Tell me what you thought about the test overall.
3. Did the test questions make you think deeply/ critically about science?
4. Were there any questions that you really struggled with? What made these questions more difficult than other questions?
5. Do you have any ideas about how this test could be improved?
6. Is there anything else you would like to tell us about this test?