HumRRO

Human Resources Research Organization

Kentucky's Alternate Assessment Attainment Tasks: Findings and Recommendations from Administration Site Visits

Arthur A. Thacker
Emily R. Dickinson
Andrea L. Sinclair
Leslie R. Taylor

Human Resources Research Organization (HumRRO)
10503 Timberwood Circle, Suite 101
Louisville, KY 40223
Phone (502) 339-9331
FAX (502) 339-9432

Draft until released by KDE.

Prepared for: Kentucky Department of Education
                       Capital Plaza Tower, 18th Floor
                       500 Mero Street
                       Frankfort, KY 40501

Abstract

Kentucky's alternate assessment system was designed to test students who could not be assessed, even with accommodations, on the regular Kentucky Core Content Tests (KCCT). The students taking the alternate assessments tend to have significant cognitive and/or physical disabilities. The alternate assessments consist of three types of assessments, Alternate Portfolios (used for reading, mathematics, and science), Transition Attainment Records (a checklist assessment designed to be predictive of success beyond high school), and Attainment Tasks (used for arts and humanities, practical living/vocational studies, social studies, and writing). This study investigated the administration and scoring of the Attainment Tasks.

Each Attainment Task assessment consists of several multiple-choice questions with symbol choices for students to select. The task may also include items designed to mimic the open-response format of items on the regular KCCT. Those items require that the student create a response using several symbols in a systematic and meaningful way. Most students are given three options for the multiple-choice items. Those with more significant communication issues are only given two choices for multiple-choice items and the numbers of symbols available for the open-response-type items are also reduced.

Observations of students being administered the Attainment Tasks and interviews with their teachers and principals resulted in several recommendations for improving the assessments. However, those recommendations fall short of addressing a more significant issue that was revealed during the data collection process. The Attainment Tasks are based on a subset of the regular Kentucky Core Content for Assessment, the guiding document for creating assessments for the KCCT. The content tested by the Attainment Tasks was not sufficiently flexible to be accessible to many of the students being tested. The format of the assessment was also not sufficiently flexible to allow access for many students. In their current form, the Attainment Tasks can only provide believable scores with classroom utility for the highest functioning students assessed. For the remainder of the assessed students, scores are random, misleading, or absent. The strongest recommendation generated by this study is to review the standards and testing practices for the purpose of introducing sufficient flexibility to allow a much greater portion of the students taking the alternate assessment to generate meaningful scores. The current Attainment Tasks fall short of meeting this most basic validity requirement.

KENTUCKY'S ALTERNATE ASSESSMENT ATTAINMENT TASKS: FINDINGS AND RECOMMENDATIONS FROM ADMINISTRATION SITE VISITS

Table of Contents

KENTUCKY'S ALTERNATE ASSESSMENT ATTAINMENT TASKS: FINDINGS AND RECOMMENDATIONS FROM ADMINISTRATION SITE VISITS

Background and Introduction

This study was designed to investigate the validity of the Attainment Task component of Kentucky's Alternate Assessment. The Alternate Assessment is part of the larger Commonwealth Accountability Testing System (CATS). It is specifically tailored for students with disabilities that prevent participation in the regular assessment, even with accommodations. The provision for testing these students with the Alternate Portfolio was established by the Kentucky Education Reform Act of 1990. The KY Attainment Tasks were added in spring 2007.

Attainment Tasks are used to test Alternate Assessment students in Arts and Humanities, Practical Living/Vocational Studies, Social Studies, and Writing. Students are administered Attainment Tasks, in each listed subject, once in elementary school (at Grades 4 and 5), once in middle school (at Grades 7 and 8), and once in high school (at Grades 10, 11 and 12). Regular assessment students take tests in the same subjects in corresponding grades as part of the Kentucky Core Content Tests (KCCT). Attainment Tasks are designed to be "on grade level" and are written to assess a subset of the standards on which the regular assessment is based.

The Attainment Tasks were designed, in part, to comply with the requirements of the Individuals with Disabilities in Education Act (IDEA) of 2004, which requires that disabled students participate in all aspects of the state's accountability system. Because Kentucky assesses regular education students in these subjects, Alternate Assessment students must be assessed in as parallel a manner as possible. Attainment Tasks apply only to the state accountability system, as do regular assessments in these same subjects. NCLB requirements involve only math, reading, and science, which are assessed using alternate portfolios.

The Attainment Tasks include several multiple-choice items and can also include one or two items designed to mimic the open-response items found on the regular KCCT. Open-response items require that students prepare extended answers to complex questions. Students are read the prompts and any accompanying materials (passages, scripts, reference materials). For the multiple-choice items, all potential choices are placed on cutout cards that depict answer choices as both words and symbols. Students select the answer they think is correct by indicating a card. Dimension A students (symbolic communicators) select from 3 choices, while Dimension B students (those pre-symbolic or "starting-with-symbols" communicators) select from 2 choices. The "open-response" items use many of the same symbols as the multiple-choice items. Students select from several symbols to construct a response to an open-ended prompt. If their selections follow a discernable pattern that accurately answers the question, the administrator rules that the item is answered correctly.

This project was designed to examine the administration of the Attainment Tasks for potential threats to the validity of the assessment. The research tasks designed for this investigation included field observations of the administrations of the tasks and interviews with administrators (primarily special education teachers). The goals of the project included providing

an accurate depiction of the test administration processes for several students and describing validity concerns stemming from those observations, as well as gathering test administrator feedback regarding validity issues associated with the tasks themselves and their administration.

This project is directly linked to research questions presented in Kentucky's draft validity argument[1]. Research Question No. 8, "To what extent are students responding to the assessment items as intended?" is addressed. This study specifically investigated the claim that "Students respond to the tasks/prompts as intended (Claim A)" from the Theory of Action included in Kentucky's validity argument. There are several underlying assumptions that are implied by the Claim A listed above. We expect students to respond to the tasks/items in such a way that administrators can judge which response they indicate. We expect student response selections to be deliberate and related to the student's attempt to correctly respond to the tasks/items. We expect students to attend to the materials necessary to respond to the items (e.g. to look at maps when required to answer questions). These basic requirements must be met in order to judge student data to have the potential to represent academic achievement (see Figure 1).

Figure 1. **Assumptions Underlying Claim A**
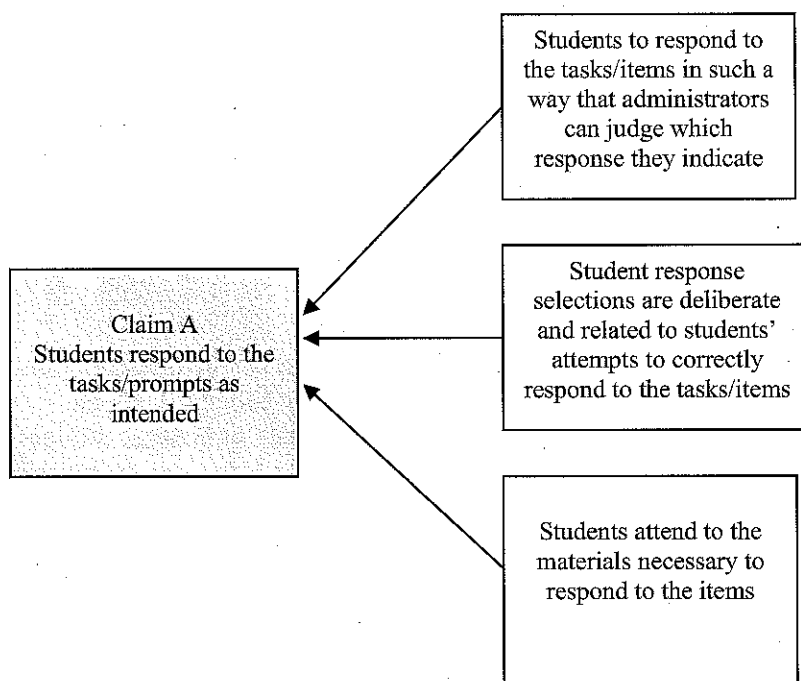
The second claim (Claim B, also from Kentucky's Theory of Action) investigated by this study was that "Construct irrelevant variance is minimized." Claim B is directly linked to Research Question No. 9 from the validity argument, "To what extent are scores accurate and

---

[1] This document, as well as HumRRO's validity argument specifically associated with this study, have been submitted to KDE but are still under review.

reliable?" Accuracy and reliability can be reduced by construct irrelevant variance. Claim B also implies several inherent assumptions. Teachers (test administrators) must have the knowledge and skills (via administration training or otherwise) necessary to administer the assessments appropriately and consistently. We expect teachers to be able to correctly classify students according to their communication dimension in order to select the version of the test to be administered. We expect teachers to administer the test in a way that allows students to respond and demonstrate whether they have met their academic expectations. Finally, we expect accommodations to be used appropriately and not to introduce construct irrelevant easiness or difficulty, either by their inclusion or omission (see Figure 2).

Figure 2. **Assumptions Underlying Claim B**
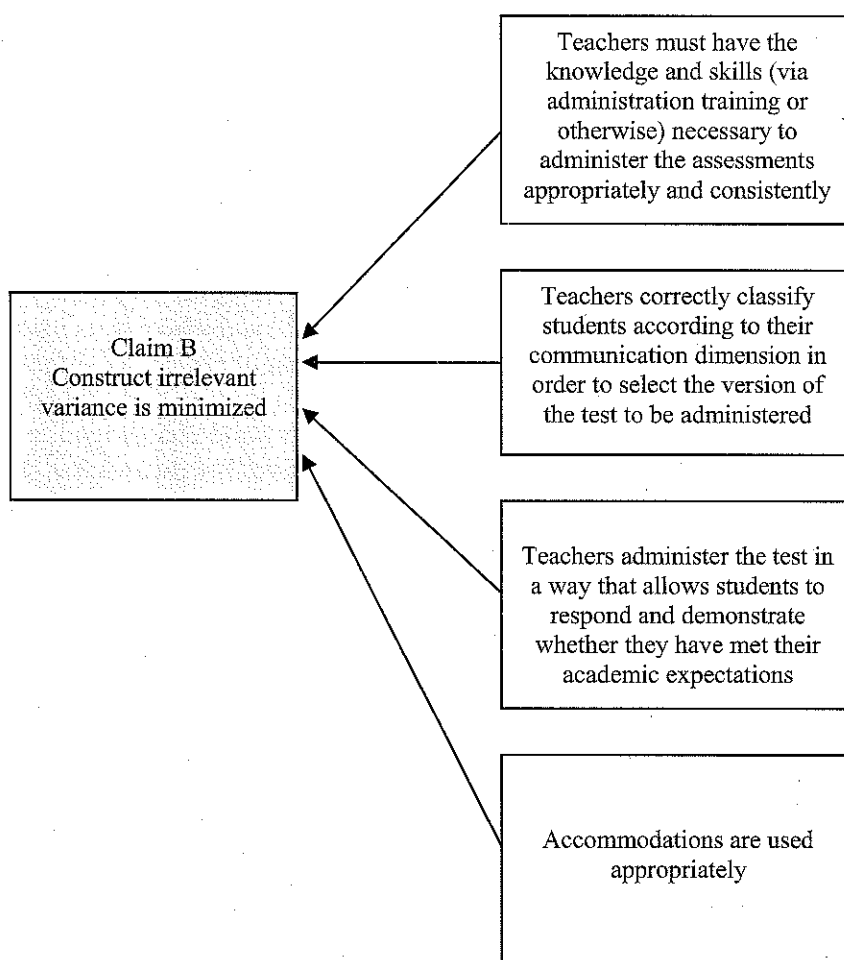
This study was designed to test the assumptions included in the boxes to the right in Figures 1 and 2. By doing so, we also test the veracity of the claims that make up the Theory of Action in Kentucky's validity argument. If these assumptions do not hold, it calls into question the claims, and therefore the validity of the Attainment Tasks, a component of the larger alternate assessment system.

Methods

To develop the sample of schools, HumRRO first selected 9 districts from around the state, using a purposeful sampling technique designed to gather perspectives from geographically dispersed areas. Within those selected districts, HumRRO used data from the previous year's Alternate Assessment administration to identify schools and select a group that included both small and large special education populations, and to ensure that we were able to observe students being tested on all of the grade/subject combinations. HumRRO contacted the selected districts and schools with an introductory letter, and then followed up via phone and email to schedule the visits. When scheduling visits, we indicated that we were interested in seeing as many administrations as possible, while also assuring teachers that we did not want to disrupt their planned schedules.

Ten schools were selected from the 9 districts to participate in the study. They included 2 elementary schools, 2 high schools, 4 middle schools, and 2 schools that had students at all tested grade levels. A total of 21 teachers were interviewed from these schools. The number of teachers interviewed at each school ranged from 1 to 5. A total of 36 students were observed, although some students were observed taking more than one of the Attainment Task assessments. The number of students observed at each school ranged from 1 to 7.

Because of the small numbers of schools, teachers, and students represented in this study, no percentages related to individual questions or topics are presented. This decision was made to guard against improper inferences being made based on percentages that might not accurately represent the larger populations. Percentages were also avoided because the questions asked of teachers were not always uniform. A protocol was used, but researchers were encouraged to pursue topics brought up by teachers that seemed pertinent or informative. For this reason, there may be instances when a particular teacher made a comment that others might or might not have made because of differing interview circumstances. Assigning percentages to these responses would be misleading.

HumRRO developed observation and interview protocols to guide the visits (Appendix A). The observation protocol consisted of guiding questions to help focus the visit, but relied heavily on unstructured, yet detailed note-taking to allow for the emergence of unanticipated themes. The interview protocol drew largely from the training materials provided to the Attainment Task administrators, allowing us to gauge the extent to which they were able to administer the tasks as intended, as well as elicit feedback regarding the appropriateness and feasibility of the administration guidelines. Interviewees were also asked about the appropriateness of the test and expected performance in light of their particular students' abilities. Finally, an open-ended question allowed the interviewees to provide additional information not previously addressed.

Site visit teams consisted of two staff members. When necessary, the team split up within a school to observe concurrent administrations or to conduct interviews with multiple teachers. In most cases, both team members observed the same administrations and conducted joint interviews and then compared notes after the visit. Each team member also compiled a list of

overarching themes from each visit. Following the completion of all visits, project staff met to discuss these themes and develop a cohesive reporting outline.

Limitations

This study used a small non-random sample of schools. As such, it would not be appropriate to assume that findings presented in this report necessarily represent all Kentucky schools. Similarly, while we interviewed all the teachers at each participant school who administered an Attainment Task, it would not be appropriate to assume that responses to interview questions from those teachers necessarily represent all teachers or all special education teachers in Kentucky.

Each site visit was conducted during a single day at a school. We have no way of knowing if that day was typical. We cannot know if the students might have performed differently had the assessments been administered on a different day, but we do know that teachers recorded the results of the assessments we observed. We cannot know if the students we observed were representative of the entire school we were visiting, except in the few cases where we were able to observe all of the students taking an Attainment Task for this administration. If we returned to the schools and observed a different set of students taking the tasks, we have no way of gauging how our observations might have been different.

We do not know if having researchers present during the administration impacted student or teacher behaviors during the administration in any way. We asked teachers to monitor student behavior and to ask us to leave if students exhibited unusual behavior or seemed stressed due to our presence. Similarly, we do not know what impact our site visit had on teacher preparation for the Attainment Task administration. Schools knew of our visit well in advance and it is conceivable that some teachers did more preparation for administering the Attainment Tasks than they might have if they were not being observed.

Findings with Recommendations

Four major themes emerged during the data analysis phase. One overarching issue, which will be discussed first, is the appropriateness of the standards and Attainment Tasks for students with the most severe cognitive disabilities. Other issues were related to pre-assessment activities such as training and materials preparation, the administration of the tasks, and the quality of the tasks themselves. The following includes a summary of our findings, along with recommendations for future Attainment Task administrations.

Validity of Score Interpretations

One of the most basic validity questions for any assessment is, "Do the scores students receive represent their achievement within the tested content?" The Attainment Tasks are scored on a raw score metric (students receive a number correct) that is tied to the performance categories Novice, Apprentice, Proficient, and Distinguished (NAPD). The categorization of students is based on a standards setting procedure conducted first for the 2007 administration, then revised each year based on expert judgment. The performance levels and the scores are

designed to "provide information that is useful for teachers in building and maintaining curriculum and instruction aligned with academic expectations (KDE, 2008)." In order to accomplish this goal, the Attainment Tasks must necessarily "yield scores that reflect students' knowledge and skills in relation to academic expectations (KDE, 2008)." Based on the observations and interviews conducted for this study, the Attainment Task scores do not seem to meet these basic assumptions sufficiently for the majority of students taking the assessments.

The problem may stem from the high degree of standardization of the testing format and the lack of flexibility for students of varying cognitive abilities. Kentucky currently uses a subset of the regular assessment standards for the Attainment Task's content definitions. These standards, as written, appeared to be inappropriate for this student population. The only commonly used accommodation for the students we observed was based on their communication. The accommodation, built into the administration guidelines, was to simply remove one of the three response options for the multiple-choice items and reduce the number of symbols for the items designed to mimic the open-response items on the regular assessment. The result was a test that was only accessible for the highest functioning students. For the remainder of students, the scores were random, misleading, or absent. We make this assertion based on the large numbers of students observed during this study who did not answer any items, who responded systematically by always selecting the response on the right or the left, who selected responses prior to the questions being read, who did not ever examine the maps or other materials necessary to respond to the items, who did not seem to know they were being tested and simply returned the symbol cards to their teachers with no indication of any selection of a correct answer, and otherwise failed to demonstrate that they were purposefully selecting an answer for the questions included in the Attainment Tasks. It was rare to observe a student who was obviously trying to perform well on the assessment, and teachers uniformly told us that those were the higher performing students. Consequently, the results of this study suggest that in their current format, the attainment tasks stratify students based on the severity of their cognitive disability. This severely limits the kinds of inferences that might be drawn by teachers and schools based on student scores. Since school accountability indexes are based on aggregations of student performance categories, it also creates a fairness concern for schools with disproportionate numbers of more severely disabled students. Unless the Attainment Tasks are substantially redesigned, the results from this study provide little evidence that teachers and schools would benefit from receiving student score reports. In fact, the evidence collected in this study indicates that providing student score reports could actually be a disservice to teachers and schools, as the scores were often random, misleading, or absent. The following sections contain evidence for these and other validity concerns related to the Attainment Tasks.

Appropriateness of the Attainment Tasks

Life Skills versus Academic Content

One of the main concerns of teachers regarding the Attainment Tasks was the content being assessed. The Attainment Tasks were designed to assess content as described by the Kentucky Core Content for Assessment. The Core Content was written primarily for the regular, non-disabled population of students. While the standards selected from the Core Content for the Alternate Assessment do not include all of the academic content that might be tested on the

regular assessments, the selected standards do represent "academic content." The social studies alternate standards include economics, government, and map skills. The arts and humanities standards include music terminology and dramatic elements. The practical living/vocational studies standards somewhat blur the line between academic content and life skills; they include content related to health and exercise.

Most teachers told us that the content on which the Attainment Tasks were designed held little value for their students. They described students for whom simply conducting the most basic life functions (getting dressed, maintaining hygiene, preparing and eating food, etc.) presented a substantial challenge. In some ways, the value of the content inherent in the standards to later life might be tenuous for regular education students. For example, one might argue that a very small proportion of students (regular or Alternate Assessment students) will ever use the Pythagorean Theorem once they leave high school. However, most regular education students have mastered basic life skills prior to learning about Pythagoras. Alternate Assessment teachers are forced to conduct "academic triage" whereby they must ascertain the most critical needs of their students and attend to them in turn. Regular education teachers perform the same function for their students, but deal with less range in student ability and breadth of potential content topics. For example, most teachers require that students have a basic knowledge of addition and subtraction prior to learning about variables in equations. Many special education teachers described a mismatch between their students' most pressing needs and the requirements of the academic content standards.

Several teachers explained that they struggled with teaching the academic content and that adding it to their curriculum took time away from teaching what they viewed as critical life skills. Others told us that they simply did not devote much effort toward teaching the academic content. They viewed attempts to teach toward the content standards to be unproductive (i.e., students did not benefit from their efforts). Gains in student performance patterns may be limited by the extent to which teachers choose to implement academic curriculum.

Relevant Quotes and Examples

- A teacher stated that a number of her students face serious medical problems, and that to "know about Europe" is not important to them or their families.
- A teacher noted that one particular student had made "remarkable gains," but that this would not be reflected in the student's test score because the "items are not appropriate for our students."
- A teacher stated that the previous years' items were "more appropriate" because they seemed to cover "more life skills."
- A teacher said "this alternate assessment does not truly assess these kids. These kids are in FMD for a reason—because they don't have the same cognitive function as kids in their grade. They need to be working on life skills. I have a kid who just learned how to feed himself. That's major. Finding Rome on a map is not relevant for him." This teacher went on to state that "in the old days, you assessed kids on reading and so forth, but you also assessed life skills."
- A teacher stated that she thinks the items are appropriate for her students and that all her students "will be fine."

Recommendations

Based on the results of this study, the Attainment Tasks in their current form are largely divorced from what teachers view as relevant for their students. If the standards were revised to contain academic content without purposefully excluding life skills and if the Attainment Tasks reflected what teachers viewed as relevant for their students, teachers might be more likely to "buy in" to the standards. For example, items relating to the use of maps could be revised to incorporate relevant life skills (e.g., using a map to navigate a store or to plan a bus trip). Revision of the standards to accommodate the more significant cognitive disabilities would be a strong step in this direction. A more flexible set of standards might lead to more relevant Attainment Tasks and a stronger link with curriculum. For example, Kentucky's alternate assessment standards for reading, math and science contain the selected standard as written in the Kentucky Core Content, a "Critical Function" statement that narrows the content statement to its most important theme, a "Measureable/Observable Skill" statement that provides an example of what the student might actually do to demonstrate achievement, and an "Intent" statement that further refines what is actually expected for the alternate assessment population. The intent statement often reduces the scope of the content statement (e.g. if the content statement says convert units, the intent statement might limit that content to the monetary system). The content described is certainly still academic, but the additional information may increase its relevance for alternate assessment students. The content statements for the Attainment Task subjects do not include similar information, but might benefit from a similar revision. This effort would be further bolstered by strong examples of the integration of the academic content described by the standards into special education curricula and instruction. Teachers struggled to find appropriate applications of the content that they also viewed as teachable.

Relevance of Standards

The current Alternate Assessment standards are a subset of the regular Core Content for Assessment standards. Many teachers and principals made the point that these standards may not be appropriate as written for many Alternate Assessment students. They explained that the content, as currently written, is simply beyond the capacity of many students to learn. Teachers expressed frustration regarding how to teach the content to lower performing students. They were unsure how to create curricula and instruction to allow students access to the content. Some explained further, that while content teachers were happy to help, they were often unsure how to adapt their teaching methods to allow Alternate Assessment students access to the content.

Relevant Quotes and Examples

- A teacher stated that she supports "standards-driven education, but not the standards that have been chosen." She described the standards for the Attainment Tasks as "not connected to one another," and stated that "there should be a building on of the standards."
- A teacher stated that the "standards selected for the assessment were not chosen wisely. They are not appropriate for this population."

- A teacher stated that "My As [Dimension A] performed better than my Bs [Dimension B]. The test didn't mean anything for the Bs. The alternate assessments are a waste of time for these kids. She then went on to state, "if someone can show me how to get them to understand this stuff, I'll do it."
- A teacher said, "I liked the test better this year than last year. Some of the questions were more interesting, at least from my perspective. But that doesn't matter to my students. They don't get any of this anyways."
- A teacher stated, "We are working on how to use a napkin. You can't go from that to the Pythagorean theorem."
- A teacher stated "I think my kids did real well on the Attainment Tasks."

Recommendations

It may be helpful to provide strong examples of lessons, accommodated to varying degrees, for each of the selected content standards at each tested grade level. It may also be worthwhile to examine other states' Alternate Assessment standards, as well as Kentucky's alternate assessment standards documents for other subjects, and consider a revision of the current Kentucky documents. Students might benefit if the standards could account for the type or severity of their disability in some way and still allow for the assessment of student achievement. A tiered system, for example, with tiers describing an increasingly rigorous interpretation of the standards, might help provide access to a wider range of students.

Teacher "Buy In"

Many teachers did little preparation for administering the Attainment Tasks. Some only opened the envelope containing the tasks on the day of the assessment, sometimes only a few minutes prior to starting the test. These teachers did not develop any supplemental aids or make any modifications to the Attainment Tasks in order to improve their students' access to the tasks, even though their training encourages such accommodations. Only a few of the observed administrations were modified/accommodated in any way. When asked why they did not make any modifications to the assessment, teachers either explained that the format of the test was acceptable as it was, or that it would not have mattered if they accommodated further or not, as they described their students as not having the cognitive capacity to access the test.

Some teachers also continued administering the assessment despite obvious student discomfort. After the administration, these teachers explained that they wanted to "get it over with" and that they did not believe that giving the students a break would have improved their students' responses. For several students, their responses appeared to be unrelated to their understanding of the item content. Many simply did not answer the items at all. Others always chose the response on the right or left. Many students also seemed to select their responses randomly, or to select the most visually appealing or familiar answer symbol. We should note that while it appeared that these students did not make purposeful attempts to select the correct responses, we cannot know with certainty whether or not they were attending to the items, nor is there a definite way of determining if further accommodations for these students would have made any difference in their response patterns.

The opinion that students cannot perform well on the Attainment Tasks or learn the associated content was pervasive among many teachers. This opinion was used as justification for decisions to spend little time preparing for the administration. It was also used to justify why some teachers spent little time teaching the content of the Alternate Assessment standards.

Relevant Quotes and Examples

- A teacher stated that the instructions for "allowable" modifications are "not appropriate" and "don't understand the full set of limitations of these kids." The teacher went on to say that the system is "forcing (us) to break our ethics to get answers out of the kids."
- A teacher stated, "because of my population of students, even if I had modified the assessment, they still wouldn't have understood. She went on to say that "these kids are FMD; they are very low functioning. They just can't do this content."

Recommendations

Prior recommendations related to improving the relevance of the academic content should improve teacher buy-in as well.

Attainment Tasks Preferred to Portfolios

One of the more curious side-effects of poor teacher buy-in is that most teachers preferred the Attainment Tasks to portfolios. They described several reasons for their preference, none of which included the utility of the scores or the impact of the Attainment Tasks on curriculum or instruction. Their preference was instead related to the relatively little effort required to administer the Attainment Tasks and that the Attainment Tasks (and their associated content) did not impact the curriculum as much as the portfolios. Portfolio preparation takes a great deal of classroom time. Teachers described portfolio preparation as "taking time away from teaching." The Attainment Tasks took less time away from teaching. Teachers were also pleased that the materials for the Attainment Tasks came to them pre-made and that they were not responsible for creating the tasks.

Some teachers also mentioned that they thought the Attainment Tasks were more "standardized" than the portfolios. All students receive essentially the same items. This was considered fairer than the portfolio system, even though the teachers did not believe that scores on the Attainment Tasks represented their students' abilities. Teachers viewed the complexity and supports components of the portfolios as measures of "teacher creativity" or teacher assessment savvy. They do not perceive a similar teacher-component to students' scores on the Attainment Tasks.

Relevant Quotes and Examples

- One teacher said, "I like the AT. The other part [portfolios] I have to spend hours and hours preparing." She went on to say, "I like that we get it and it's done with."
- Another teacher told us, "The AT are just as accurate as the portfolios. Why not just do AT for all (grade/subjects)?" as she expounded on the general inaccuracy of the alternate assessments.
- Finally a teacher said, "I like the AT better than the portfolios. It requires less time and effort. The portfolio pulls me away from teaching more."

Recommendations

Teachers' preferences for the Attainment Tasks over the portfolios were not based on the quality of the Attainment Tasks, but on concerns they had with the portfolio system. Their preference should not be viewed as an endorsement. Teachers' statements reveal an overall lack of confidence in the Alternate Assessment system. Data that indicate the system produces valid and reliable results (for any component, assuming such indications exist) should be summarized and publicized in teacher-friendly language.

Inordinate Stress for Students

While testing can be stressful for any student, some disabilities can exacerbate this stress. Certain manifestations of autism, for instance, may make the administration of the Attainment Tasks problematic. The testing situation takes students out of their regular or routine environment. In several cases, testing was done in an unfamiliar room with a variety of unfamiliar potential distractions. The nature and format of the assessment may also have been unfamiliar to the student.

The tests can be lengthy. The arts and humanities tests have an especially large amount of text for teachers to read aloud. All observed Grade 5 students completed both the arts and humanities and social studies tests in one sitting. In some instances, this required as much as a full hour of testing to complete. While test instructions allow for the two tests to be given at different times, that option was not used among the sampled students. Many students complained that they were tired and/or bored during testing. Many placed their heads down on their desks or acted out before testing was completed. Observers did see teachers give students short breaks between tests or test sections on occasion, but these breaks were very short (less than 5 minutes). The current method of administering the Attainment Tasks seems to create a test of excessive length for some students.

For a small subset of students, the administration of the Attainment Tasks seemed to cause a great deal of stress. One student complained throughout the test and pulled her hair out in sizable chunks during the administration. When the teacher administering the test was asked later about the behavior, the researchers were told that the student had never exhibited that particular behavior prior to test administration. Another student screamed and slid out of her chair. She then tried to escape from the exam room. The teacher administering the exam was required to

pick the student up from the floor on several occasions to complete the test. Still another student shook her hands throughout the exam and vocalized a high-pitched tone during the test. One young man was disturbed and scared by one of the passages, as evidenced by a phrase he repeated regularly after the passage had been read to him. A high school student protested throughout the entire assessment, yelling that he hated the test and getting up from his seat to leave the room on several occasions. He could only be convinced to stay for the duration of the test by the promise of a food reward upon completion.

Relevant Quotes and Examples

- One teacher was observed allowing several short breaks throughout the administration of the Attainment Task. Each time the student completed a section, he was allowed to play with toy for approximately a minute. This appeared to keep the student calm and allow for the completion of the test in a single session.
- One student left his seat after nearly every item on the assessment and announced his intention to leave the room. It did not appear that he ever attended to the items on the test, but rather stayed and participated as a means to receive a snack.

Recommendations

It may be helpful to provide more or clearer guidance for when and how to break up the administration of the tests. Provide more or clearer guidance on when it is inappropriate to continue testing and when it is not. Reduce the length of the tests themselves, especially the long reading passages. Improve bias and sensitivity review[2] for the passages accompanying the test items and for the items themselves.

Pre-Administration Issues

Training of Test Administrators

Teachers were required to complete an online training module and take a qualifying quiz prior to administering the Attainment Task. The purpose of the training module and quiz was "to ensure appropriate administration and scoring of the Attainment Tasks" (http://www.ihdi.uky.edu/KAP/index.asp). All teachers we observed and interviewed indicated that they had successfully completed both the online training module and the qualifying quiz. During interviews, we asked each teacher to indicate whether the online module provided adequate training. Most teachers indicated that the training was adequate. Only a couple of the teachers reported that the online training was inadequate. One teacher indicated that she would have liked to have received additional "hands-on" training with someone "in person." Another teacher indicated that she had received sufficient training only because the co-op in her district had provided supplemental training

---

[2] Bias and sensitivity review did occur for the Attainment Tasks, but it was conducted internally by KDE personnel in a very short time. Passages may not have been reviewed thoroughly. (Personal communication with Rhonda Sims and Phyllis Shuttleworth from KDE after the completion of data collection)

above and beyond what was provided by the on-line training module. She stated that without the additional training provided by the co-op, the on-line training would have been insufficient.

Despite the fact that nearly all of the teachers indicated that the online training was sufficient, our observation of the Attainment Task administration provided evidence that was more congruent with the sentiment of the two teachers who indicated that the training was insufficient. We observed multiple instances in which teachers did not follow the directions in the training module and/or appeared unaware of what was allowable. Many inconsistencies described elsewhere in this report provide a strong indication that training did not result in a standardized administration of the assessment. It was common to observe tests administered where the first response read was always correct. A teacher was observed administering both Dimension A and B versions of the assessment to a student. Teachers commonly read the "do not read" portions of their scripts to students. Teachers provided students with incorrect information during the assessment (e.g. I'm not allowed to tell you if you got the answer right.). The most salient evidence of ineffective training was the lack of modifications to the Attainment Task materials. Most teachers administered the Attainment Task materials without making any modifications, even though the training included content on materials modification. The most elaborate modification we observed was a teacher who recorded the response options into a voice-activated "switch" upon which the teacher placed the symbols; however, this modification in no way improved the student's access to the assessment. The student was completely non-responsive throughout the duration of the assessment. A few teachers attached the symbols to a poster board that they held up in front of the student. Lastly, one teacher reported that she planned to puncture holes in the symbols for a deaf and blind student. No other modifications were observed or reported by the teachers we interviewed. Many teachers indicated that they were uncertain as to what types of modifications were permitted; consequently, for fear of "breaking the rules" they opted not to make any modifications. This suggests that the training module would be improved by more and clearer information regarding when and how to make appropriate modifications, perhaps with more specific examples.

Several other issues also called into question the adequacy of the training: 1) We observed several instances in which teachers read aloud the italicized, "not to be read aloud" portion of the script. 2) We observed a teacher who administered both the Dimension A and Dimension B version of the Attainment Tasks to the same student. 3) We observed multiple teachers who did not vary the placement of the answer choices. As such, the correct response was always located in the same position. 4) We did not observe any instances of teachers splitting up the tasks into smaller time chunks, even though students were clearly tired (e.g., One student said she was "worn out." Another student put her head down on the table and started yawning. Other students became increasingly fidgety/distracted through the duration of the assessment). 5) A couple of students asked if they got the item correct, but the teachers told the students that they were "not allowed" to reveal the correct answers, even though the training module indicates otherwise. 6) None of the teachers reduced the number of picture symbols for the writing prompts even though Step 7d in the training module indicates that it is acceptable to reduce the number of symbols as long as the specified proportion of correct and incorrect symbols are presented. Not reducing the number of symbols resulted in an unwieldy amount of symbols to manage; we observed several instances in which there was insufficient space to display all the symbols in front of the student in a way that permitted the student to view them

without shuffling through stacks of symbols. 7) Lastly, we observed considerable confusion and inconsistency among teachers on the scoring of the writing prompt items. For example, one student responded to a writing prompt with a personal story, but without referring to the symbols provided. His teacher asked him to use the symbols, but he indicated that the symbols weren't relevant to his story. The teacher ultimately scored his response as incorrect, though he did provide a reasonable response to the prompt. Administration inconsistencies were the rule rather than the exception among the participant teachers.

Relevant Quotes and Examples

- A teacher asked the observer how to administer and score the open-ended question.
- A teacher described the training as "adequate, but not helpful."
- A teacher stated that she liked the online training because it was "not really long" and because it was "good to not have to travel."
- When asked what modifications she made to the attainment task one teacher said, "You can't modify the Attainment Task too much. You have to read what's in the script."
- Another teacher said, "I cut out the symbols. That's all I was allowed to do."

Recommendations

To help alleviate these sources of construct irrelevant variance, the following may be helpful: 1) provide additional training, perhaps with a "live" trainer, so that teachers better understand/know the instructions for administering the Attainment Task, 2) provide more in-depth examples of the types of modifications that are acceptable so that teachers will feel more comfortable making modifications to the assessment materials, and 3) provide more specific instructions on how to score the writing prompt items so as to reduce/eliminate the confusion and inconsistency surrounding the scoring of these items.

Delivery of Assessment Materials

Another source of construct irrelevant variance may result from Attainment Task materials not being delivered in time to allow teachers sufficient time to prepare materials for administration. Additionally, the delivery of incomplete assessment materials is problematic. Fortunately, all interviewed teachers reported that they received their Attainment Task materials on time and that the Attainment Task materials were complete. We mention this because several teachers noted that this was a marked improvement over last year. A few teachers noted that the number of Alternate Assessment students a teacher is responsible for teaching should be taken into consideration when determining what constitutes an appropriate amount of prep time. The more students a teacher has, the more time needed for preparation/modification of materials. In addition, some modifications to the current organization of the assessment packets might reduce the required preparation time, as well as some of the error possibly introduced into the administration. Currently, answer symbols are provided on several sheets of paper, each with a numeric code. Each test item lists the codes for its corresponding answer symbols, and several

answer symbols are used across multiple items. The majority of teachers' preparation time was spent cutting out the answer symbols and organizing them along with the appropriate items.

Relevant Quotes and Examples

- A teacher stated that it took two hours to get the Attainment Task materials set up for two students and that it would be "too much" for a teacher with more students.
- A teacher stated that "for an assessment, this required a lot of work for the teacher before you can actually administer it."
- A teacher stated: "why do we have to make materials? Why don't they come pre-prepared?" This teacher also stated that she got all the materials in time and "just cut up the existing materials." She also mentioned increasing the font size and cutting the words away from the symbols for a visually impaired student.
- A teacher said, "I like the Attainment Task because it's a pre-made packet. It's easy."

Recommendations

Each item could be presented along with its answer symbols, and any repeated answer symbols could be presented separately for each item with which they are associated. It also might be helpful to consider printing the answer symbols on heavier stock paper with the answer choices presented in a pre-determined order. It did not appear that any students were particularly helped by the presentation of the answers on separate cards, and the most often seen modification was to affix the symbols to a larger sheet of paper or poster board. Teachers could be trained that the symbols could be cut apart, if needed, though they should be presented to the student in the prescribed order. This would cut down on preparation time for the large numbers of participating teachers who did little to modify the assessment.

Administration Issues

Inconsistencies across Teachers in Attainment Task Administration

During the course of the Attainment Task observations and interviews, we observed many inconsistencies among teachers in their administration of the Attainment Tasks. Administration inconsistencies likely stem, in part, from training issues; however, because these inconsistencies were observed during the actual administration of the assessment they are discussed in this section. Nonetheless, the discussion of these inconsistencies will likely inform issues to be addressed in Attainment Task training as well.

Despite the scripted nature of the Attainment Tasks, we saw many deviations in its administration across teachers. One source of difference stemmed from teacher facility in reading scripts and associated passages. Some teachers were much more adept at narrating or storytelling. These teachers used appropriate voice inflection when reading passages, their pauses came at appropriate times, they used different voice intonations to denote multiple characters, and their overall reading of the script was very fluid. This was in contrast to other teachers who stumbled over the script, spoke in a monotone fashion, mispronounced words, and

had awkward pauses. Because the length of passages tended to be quite long, poor narration was particularly problematic; we, as observers, had difficulty following along with the readings. Students likely had difficulty following as well. This was evidenced by non-verbal cues such as playing with their hair, picking at their fingernails, looking at pictures on the walls, etc., and verbal cues such as sighing, yawning, and commenting that they were "bored." If teachers are doing a poor job of reading the script/passages, then this would likely impact students' understanding of the material, which could impact their performance on the Attainment Task.

Another inconsistency related to teachers' re-reading of key parts of the script. Some teachers were very explicit about re-reading question stems and response options. Others read the question once and then placed the symbols in front of the student without reading aloud the words associated with the symbols. Also, some teachers would re-read passages each time questions related to that passage were asked, whereas others would read a passage only once.

There was also inconsistency across teachers in whether they tried to relate the Attainment Task material to something with which the student was familiar. For example, when reading particular passages some teachers would pause to make comments like, "this is like the time you did _____" or "remember the time we talked about _____?" These insertions served as a way to grab the students' attention and also to make students think about what the teacher was reading and relate it to something with which they were familiar. This appeared to be an effective strategy for those teachers who used it, but again, it was not used consistently across teachers. Some teachers also took greater liberty in orienting students to key pieces of information. For example, one teacher pointed to where the key was located on a map.

There was also variability in whether one or two test administrators were present during the Attainment Task administration. In most cases, an individual teacher administered the assessment. In a few cases, an aide was also present. In these instances, the aide was there to help calm the student and/or to provide back-up if the student became disorderly. The aides did not assist in the actual administration of the Attainment Task. In a third scenario, we observed two special education teachers administer the Attainment Task in tandem. One teacher would read the script and record the student's scores and the second teacher would prepare and arrange the picture symbols. This arrangement resulted in the smoothest administration of the Attainment Tasks that we observed.

Lastly, there was variability across teachers in their placement of correct response options. Some teachers varied the placement of the correct response, whereas other teachers always placed the response options in the same order in which they appeared in the assessment packet. This introduces a threat to the accuracy of the scores. For example, if the leftmost response option is always correct and the student always picks the answer on the left (which we observed multiple times), then that student's score would not be a valid representation of his/her mastery of the content being assessed. As discussed in the above section, this is an issue that may warrant improved training amongst teachers administering the Attainment Task.

Relevant Quotes and Examples

- A teacher administered both the Performance A and Performance B versions of the test to one Dimension A student.
- A teacher mispronounced one of the response options. During the same administration, a student was presented with an incorrect response option (a response option meant for a different item).

Recommendations

In order to improve the construct irrelevant variance introduced by the variability in teachers' adeptness at narrating, one suggestion would be to use voice recordings of professional speakers. Using recordings would standardize the narration of the material across administrations and thereby eliminate the construct irrelevant variance introduced by the variability in teachers' narrating abilities. The use of recorded narration might also generate other construct irrelevant variance (are students attending, are devices playing the narration loudly enough, can students ask for narration parts to be repeated), so this suggestion should be weighed in light of both potential positive and negative aspects.

More guidance should be provided about which parts of the script should be read and re-read, as well as what types of deviations from the script are acceptable. Reductions in length of the teachers' reading materials should also help. Recommending that two teachers administer the Attainment Task in tandem seems worth considering as guidance for the administration as well.

Test Materials Difficult to Manage

Many teachers voiced complaints about difficulty coordinating and arranging all of the "pieces" of assessment material (i.e., the cut-out symbols, the script, the maps, etc). Those complaints bore out in our observations. In nearly every administration, we observed teachers fumbling through symbols trying to locate the appropriate symbols for each question (the exception was the two teachers who administered the Attainment Tasks in tandem). This was problematic because it resulted in a considerable lapse of time between the reading of the question and the placement of the response options. Some teachers re-read the question after assembling the response options, but many did not. Consequently, it is possible, or even likely, that many students had forgotten the question by the time the response options were laid out. The writing prompt items were the most problematic due to the sheer number of symbols associated with those items. Teachers did not appear to be aware of the fact that they could reduce the number of symbols for the writing prompts, as we saw no one implement this option. There were so many symbols associated with the writing prompt items that not all of the symbols fit on the wheelchair trays of those students sitting in wheelchairs. Moreover, several teachers grew visibly frustrated while struggling to find the appropriate symbols for a given question, and some made negative comments during the administration related to the unwieldy nature of the assessment materials. We also observed several cases in which symbols were mistakenly placed with an item. In fact, there were two instances where teachers failed to display the correct symbol as a response option at all.

The disorganization of the materials is due, in part, to the fact that most teachers used the same set of materials for testing each student. Consequently, after each administration, the materials were out of order. Many teachers did not reorganize the materials before administering the assessment to the next student; hence, the reason for the fumbling. However, there was still some fumbling even among the teachers who organized the materials prior to each administration. Because the same symbols are used for multiple questions, it can be difficult to keep track of which symbols go with which items quickly and efficiently.

Relevant Quotes and Examples

- A teacher misplaced a response symbol and took roughly a minute to locate it. The teacher did not repeat the question once the response symbol was located.
- A teacher who had two students testing at the same grade level combined the answer symbols from the two testing packets so that she would not have to shuffle to locate repeated answer symbols.

Recommendations

One suggestion to improve this disorganization would be to print out a unique set of symbols for each question and to have those symbols printed on the same sheet of paper as the question. With this arrangement, there would be no need to cut out the symbols[3], and if there were no need to cut out the symbols, then this would ensure that the wrong symbols would never be mistakenly placed with an item. Alternatively, if the current system remains in place, then the best option would be to administer the Attainment Tasks in tandem with another teacher as described in the above section. Minimally, the reused symbols could be replicated for each item with which they were associated to eliminate teachers' searching for the correct symbol set for each item.

Scoring Inconsistencies

While touched on in the above sections, the issue of scoring inconsistencies is one that warrants additional elaboration. It is important to point out that all responses to the Attainment Task items are scored dichotomously, either as "1 = correct" or "0 = incorrect." As such, this "right/wrong" dichotomous scoring rubric is a very basic one. Moreover, when asked if they had any problems scoring the items, nearly all teachers reported that they did not. Nonetheless, we observed three points of concern related to scoring. First, there is confusion surrounding the scoring of the writing prompt items. The scoring of the multiple-choice items is very straightforward in the sense that the student either selects or does not select the correct symbol. The writing prompt items, however, are much less straightforward because students' right/wrong responses are determined based on whether they select the appropriate subset of symbols. We observed inconsistencies across teachers in their scoring of the writing prompt items for students who: 1) do not use symbols, 2) use most, but not all of the symbols correctly, 3) select one symbol, or 4) select all of the symbols. There was also varying interpretation across teachers in terms of what constituted the correct subset of symbols. Neither the on-line training module nor the Attainment Task packet provided specifics for how to score these types of responses. A

---

[3] This would also save the teachers from having to use their time to cut out pictures.

second issue that relates to the accuracy of student scores is that some teachers did not vary the placement of the correct response option. This is particularly problematic for students who, for example, always select the leftmost response. If the leftmost response is always the correct response, those students would receive perfect scores on the Attainment Tasks, which, obviously, would not be valid representations of their mastery of the Attainment Task content. Third, there was considerable variability across teachers regarding the extent to which they required some indication that the student was attending to the test. Some teachers would score the first symbol the student selected (or the symbol to which the student's hand got the closest) even if the student was aimlessly pointing. Other teachers would not score the student's first response if they believed the response was not a purposeful one. Some of these teachers would wait a couple moments to see if the student would reselect a symbol and/or they would repeat the question a second time. There is reason to believe that the latter approach may produce more valid scores than the former approach.

Relevant Quotes and Examples

- A teacher scored a student's response to the open-ended item as incorrect, though the student attended to the question and answered appropriately. The student did not, however, make explicit use of the answer symbols provided and was scored 0 for not having done so.
- Teachers did not vary the placement of the answer symbols.

Recommendations

Additional clarification is needed for scoring of writing prompt items, and the appropriateness of the dichotomous scoring rubric may need to be re-evaluated for these items. The order of placement of answer symbols should be predetermined and teachers should be trained to present items in this order. It may be helpful to emphasize in training the importance of ensuring that the recorded response is the intended response. Feedback could be elicited from teachers about how they ensure that the response they record is the intended response, and incorporate these tips into training materials.

Item Quality Issues

Item Bias and Quality Review

Some Attainment Tasks contained content that was inappropriate (e.g., violent imagery), or that was likely unfamiliar to students (e.g. jargon). Other items exhibited inconsistencies between the answer choice symbols and the visual aids or tools that were provided as part of the item, or among the symbols representing the answer choices (e.g., differences in colors, fonts). Attainment Tasks should be held to the same level of bias and quality review as regular assessment items to ensure that item-level characteristics such as wording, formatting and context are not introducing construct-irrelevant difficulty or construct-irrelevant easiness (Lane, 1999).

Relevant Quotes and Examples

- On an item that required students to access a visual aid, the visual aid and the corresponding answer symbol did not match in color, size, or font.
- Some sets of answer symbols included at least one item that stood in sharp contrast to the other answer symbol choices (e.g., one response symbol in bold colors).

Recommendations

For future administrations of the Attainment Tasks, conduct a formal review of potential item biases in content, language, structure and formatting[4].

Item Length and Phrasing

Some items and accompanying reading passages contained "technical" language or jargon, words likely unfamiliar to students in this population. Teachers were often unsure of when simplifying terms was appropriate, and seemed to err on the side of not changing the language of the items. Item scripts were at times awkward and unnecessarily long or complex. By the time teachers read through the item introduction and arrived at the item stem, many students were no longer paying attention. Other items were accompanied by long reading passages, requiring a significant amount of memorization on the part of the student.

Relevant Quotes and Examples

- A teacher placed the passage she was reading so that the student could see the passage and watch as the teacher pointed to what she was reading. This appeared to help the student maintain focus on the passage.
- A student appeared very distracted while the teacher read aloud two passages that precede some items. The student did not appear to attend to the last sections of the passage.
- One item contained very specialized language that would be used by only a small, mainly young, subpopulation of Kentuckians. Teachers were unsure how to read the words appropriately and may not have recognized the context of the item based on the words. They often stumbled over the language and mispronounced the words.

Recommendations

It may be helpful for future Attainment Tasks to contain simplified language and shortened passages where possible, along with increased training for teachers on how to appropriately modify the language and length of items and passages. Another option might be to include symbols within the passage texts and allow students to follow along while the teacher reads aloud.

---

[4] As noted above, the item review for the Attainment Task items was much less rigorous than for the regular KCCT.

Answer Symbols

Though a few teachers indicated that their students were familiar with the symbols used in the Attainment Tasks, it appeared that many of the symbols presented as part of the Attainment Tasks were unfamiliar to students. A large portion of the symbols appeared to come from the same set, though there were many that were clearly pulled from another source. In some tasks, two answer symbols were presented in black and white whereas a third symbol was brightly colored or contained a picture that might have been attractive to the student, but may or may not have been relevant to the construct being measured. Certain abstract concepts were not well represented by the symbols selected. In some cases, students might have benefited from only hearing the answer choices rather than seeing the symbols.

Relevant Quotes and Examples

- A student appeared to respond to the oral presentation of the answer choices, rather than accessing the symbols.
- A teacher stated that the students use symbols in class, but that many of the Attainment Task symbols were "inappropriate."
- One teacher stated that her student, who was able to read, was actually distracted by the symbols. She reported that if she had the opportunity to do it again, she would not have used the symbols.

Recommendations

Future versions of the Attainment Tasks could benefit from a review of the quality and consistency of the symbols used.

Discussion and Conclusion

HumRRO has been charged with the role of "evaluator" of the validity of the Alternate Assessment System in Kentucky. The meaning of evaluation in the context of assessment validation has been defined as "a systematic examination of interpretations and uses occurring in and resulting from an assessment or accountability system" (Ryan, 2002, p. 9). There are various sources of construct irrelevant variance that may threaten the validity of the interpretations and uses of test scores. In our role as validity evaluators, HumRRO systematically examined several sources of construct irrelevant variance that pose a potential threat to the validity of the interpretations and uses of test scores derived from the Attainment Tasks.

As a part of CATS, the Alternate Achievement Standards Assessment, including the Attainment Task component, aims to allow students with the most severe physical and cognitive disabilities to access a test of grade-level content, thus allowing their scores and resulting proficiency level classifications to be included in the calculation of an overall indicator of school-level progress toward proficiency goals. Thus, a student's proficiency rating derived from performance on the Alternate Assessment should be representative of that student's proficiency on grade-level content and, when considered in combination with other students' proficiency ratings from either the Alternate Assessment or the Kentucky Core Content Test (KCCT), should

provide evidence of whether or not a school is adequately meeting its students' instructional needs. Scores on the Alternate Assessment alone (separate from the KCCT) will also be used as evidence of school-level progress in serving the particular needs of this student subpopulation.

Based on the results of this study, for the scores derived from the Attainment Task to be valid indicators of student achievement, several sources of construct irrelevant variance would need to be addressed and reduced significantly. For the students we observed, many of their scores were at best random, and at worst misleading. Except for a very few high performing students, this study provides little evidence that the Attainment Task scores represent student achievement in any real way. The standards measured by the Attainment Tasks are simply a subset of regular content standards—which is inappropriate for this assessment and population according to most of the teachers interviewed. The Attainment Tasks have not been exposed to adequate levels of review and revision (Personal communication with Rhonda Sims and Phyllis Shuttleworth from KDE after the completion of data collection), and contain errors, inconsistencies and, in one case, inappropriate content. While several teachers told us that the amount of work required to prepare the Attainment tasks was far less than to prepare portfolios, the amount of work required by teachers administering the assessment is substantial (assessments are taken one at a time, require that the teacher arrange coverage for classes for each administration, require pre-organizing and cutting out materials, and would be even more work if the teachers designed customized accommodations for each student). The training provided has not been sufficient to produce the balance between flexibility and standardization that the Attainment Tasks had aspired to achieve. The teachers we observed were unable, and in some cases unwilling, to make the necessary modifications that would allow their students to access the test. And in several cases, it would be fair to say that no amount of test-level modification would result in an assessment that offered a meaningful way for these students to demonstrate what they know and can do. Several students simply did not respond at all during test administration.

Teachers themselves indicated a lack of confidence in the utility of scores derived from the Attainment Task. Their reasons related to their beliefs about the level of difficulty of the Attainment Tasks and the appropriateness of the tested content for their students, as well as students' perceived lack of access to the test, which is typically in direct relation to the severity of their physical and/or cognitive disabilities. Teachers spoke of students' scores as "random," "blank," or "luck-driven," rather than as indications of achievement or as useful tools for instructional or curricular planning.

Conversely, teachers seemed very concerned about the impact that these scores could have on the school, and what they considered to be a false impression of lack of progress, as demonstrated by low Attainment Task scores. Teachers liked to point out that their students have made, and continue to make, progress throughout the year, but that these kinds of gains are not measured by the Attainment Tasks. As a result, teachers generally expected their students to score below Proficient on the Attainment Tasks, but were not ready to take this as a meaningful indicator of their students' lack of achievement or of their own inability to meet students' educational needs. Some teachers went so far as to characterize the results of the test as being a painful reminder to students and their parents of the fact that they "don't measure up" when held

to similar standards as general education students. Many teachers spoke of themselves and of their students and their families as feeling "set up for failure."

There were indications that the Alternate Assessment is having some positive consequences for schools serving students with severe disabilities by strengthening the focus on academic content in special education classrooms, though some teachers admitted that they do not teach (or only cover very minimally) content tested by the Attainment Tasks (e.g., music, art, and drama). Whether a focus on academic content is imposed at the expense of life-skills education is still a point of contention among practitioners. Teachers who are skeptical of the appropriateness of academic content for their students might be more amenable to the idea if the standards were revised to include core or "intent" statements similar to the standards as written for subjects assessed using the portfolio (math, reading, and science). They might benefit from a "boiling down" of the standards and delineating a clearer path (similar to a learning progression) on which a significantly disabled student could make gains toward demonstrating proficiency.

The meaning and utility of Attainment Task scores is a concern. It appears that the Attainment Task scores offer little or no useful information about the content knowledge of a large proportion of the students tested, thus rendering any inferences about student achievement or school performance made from these scores to be incorrect, or at least incomplete. This is especially true of the Dimension B students, who have little intentional communication, and for whom the Attainment Task scores seem least valid. Finally, there is potential for a pejorative effect on both teachers and students and their families, if the focus on test scores reduces attention and focus from other areas in which students do make progress.

Kentucky's strict interpretation of federal NCLB and IDEA legislation, though representative of a strong belief in and a commitment to recognizing the potential of every single student, has resulted in a system that is at best insufficiently flexible, and at worst inconsistent with realistic expectations for many students with significant cognitive disabilities. The extent to which Kentucky might change the current system, and remain compliant with regulations while meeting the practical needs of teachers and students, is unknown. However, efforts to increase the flexibility of the assessment and the access of students to the content have the potential to improve the validity and utility of Attainment Task scores. Addressing many of the recommendations presented in this report would move Kentucky closer to having an Alternate Assessment that is, arguably, a more valid measure of student performance on grade-level content. Addressing those recommendations without also significantly revising the standards themselves and/or the general format and flexibility of the Attainment Tasks would still leave a large portion of the tested population with meaningless or misleading scores. Our strongest recommendation is to review the standards and/or testing practices for the purpose of introducing sufficient flexibility to allow a much greater portion of the students taking the alternate assessment to generate meaningful scores. The current Attainment Tasks fall short of meeting this most basic validity requirement.

References

Kentucky Alternate Assessment Program. (2008). *Attainment Task administration training module*. Retrieved from http://www.ihdi.uky.edu/KAP/index.asp.

Kentucky Department of Education. (2006). *Core Content for Assessment. Version 4.1*. Frankfort, KY: Author.

Kentucky Department of Education. (2008). *The KY Alternate Achievement Standards Assessment Validity Argument*. Draft document.

Lane, S. (1999). *Validity Evidence for Assessment*. Reidy Interactive Lecture Series.

Ryan, K. (2002). Assessment Validation in the Context of High-Stakes Assessment. *Educational Measurement: Issues and Practice, 21(1), 7-15*.

Appendix A

Observation Protocol and Teacher Interview Protocol

Observation Protocol
Attainment Task Study

School _____   Teacher_____

Task Administrator_____   Observer_____

Administration number _____

Grade level/Content area(s) observed_____

Summary of Observation:

Guiding questions

1. Describe the testing environment (classroom or other room, noise level, were others present?).
2. Describe how testing materials were presented (laid out within student's reach, computer adapted).
3. How does the teacher interact with the student?
4. How closely does the teacher follow the prescribed instructions?
5. How long did the entire testing time take?
6. Did anything that the teacher did (or did not do) influence the responses of the student?
7. How engaged was the student during the administration?
8. How did the student access and respond to the test items (what means of communication did the student use to respond to the items?)
9. Were the materials prepared by the teacher to assist students with the tasks appropriate?
10. Were there any interruptions to the administration process?
11. Was the task administered by someone other than a teacher (e.g., counselor)?

Teacher Interview Protocol
Attainment Task Study

School _____   Interviewee_____

Interviewer_____

Did you experience any problems receiving the testing materials from the district (were they complete, correct grade/content levels)?

Did you prepare any materials to accompany the Attainment Tasks for any students? Can you give examples? How did you come up with these aids?

Were you given adequate time to prepare the assessment materials? How long did it take?

Were the instructions on how to modify the assessment materials clear and appropriate?

Did you have to make many modifications to the items? What types? Were the answer symbols provided appropriate for your students, or did you need to switch them out often?

Did you have to make many modifications to the script? What types of modifications? Are there ways that the script could be changed that would make it more appropriate for your students?

Did the answer key make sense? Did you have any problems scoring the items?

Did you receive adequate training to administer the assessment? Did you complete the online quiz?

Were the items appropriate for your particular students?

Overall, how have your students performed on the Attainment Tasks?

Do you have any additional comments about the Attainment Tasks?