Qualitative Analysis of Kansas Alternate Assessment Data Folio Comments

Prepared for the Kansas State Department of Education

December 20, 2011

Diane M. Coffman

Benjamin T. Rutt

Julia Shaftel

Center for Educational Testing and Evaluation

University of Kansas

1122 West Campus Road

Lawrence, KS 66045

Abstract

This qualitative study analyzed reviewers' comments for approximately 800 alternate assessment data folios sampled in 2009 and 2010. Data folios contain the evidence submitted for proficiency determination for students participating in the Kansas Alternate Assessment. Folios in a content area consist of evidence of the student's performance on five content-area indicators. Data collected for an indicator consists of three separate activities with at least five trials for each activity. Reviewers examined the alignment and appropriateness of the selected activities for each indicator, scoring of submitted evidence, and completeness of the folios. Reviewers completed an evaluation rubric for each folio and added comments as they deemed necessary. This study analyzed only the comments to identify themes that may be useful for future professional development.

The Individuals with Disabilities Education Act of 1997 (IDEA), followed by the Individuals with Disabilities Education Improvement Act of 2004, required students with disabilities to be included in all district- and state-wide testing programs. Prior to July 2000, students with Individualized Education Programs (IEPs) were often excluded from state academic testing, particularly students with the most significant disabilities. According to Hanzlicek (2006), up to half of the special education students in Kansas were exempted from state testing by their IEP teams in the mid-1990's. The No Child Left Behind Act of 2001 (NCLB) mandated that all students be included in state assessments aligned with student performance standards for determination of Adequate Yearly Progress (AYP). Consistent with IDEA, NCLB required all students with disabilities, including students with the most significant disabilities, to participate in state assessments.

According to NCLB, up to 1% of all students, specifically students with the most significant cognitive disabilities, are permitted to use an alternate assessment based on alternate achievement standards (AA-AAS) to demonstrate proficiency in their knowledge and skills with respect to state educational standards. The AA-AAS is known in Kansas as the Kansas Alternate Assessment (KAA), consisting of a data folio addressing the same subjects on which other students are tested. According to Hanzlicek (2006), "the standards for these students will differ only in the degree of content and skills rather than in the expected results" (p. 3). By 2009, a data folio or body of evidence was the most common method used in alternate assessments nationwide (Roach, Elliot, & Webb, 2005).

AA-AAS methods such as the KAA pose technical difficulties. Data folios can be difficult to assemble in a consistent and standardized manner across teachers. The requirements that each student's data folio be individualized for that student and that teachers must determine how to assess each selected skill make data collection a challenge. Inconsistency in teacher preparation for the alternate assessment process and data folio compilation underscores the need for continued teacher training in data collection (Hanzlicek, 2006). The scoring of data folios brings new challenges in terms of interpretation of data, scorer training, and inter-scorer reliability (Browder, Karvonen, Davis, Fallin, & Courtade-Little, 2005; Browder et al., 2003; Hanzlicek, 2006; Rabinowitz, Sato, Case, Benitez, & Jordan, 2008).

Furthermore, teachers may question the utility of alternate assessments in general and data folios in particular. In a qualitative study of 10 Kansas special education teachers, Hanzlicek (2006) found that many considered data folio collection a "waste of time" (p. 88), while others reported they never saw the results of their students' data folio assessment so were unable to use them in a timely manner. Only a few teachers used the results for instructional decision making in this study.

Teachers may doubt if their students with various disabilities can produce usable information to include in portfolios. However, Browder and her associates in North Carolina (2005) discovered that teacher training in instructional interventions prior to portfolio collection overcame student disability characteristics, such that "84% of teachers who completed the survey (at project's completion) reported that their students made more progress as a result of being in the project, and none reported that there was less progress" (Browder et al., 2005, p. 278).

The KAA measures standards extended downward from the general curricular standards. The wording of the extended standards and benchmarks is identical to the general standards and their corresponding benchmarks. At the finest grain, the indicator level, performance expectations differ. Indicators within the extended standards are conceptualized and written to be accessible to students with significant cognitive disabilities. Less than 1% of students are eligible for the KAA each year, as determined by their IEP teams.

Teachers evaluate students on five indicators in each subject selected for assessment (math, reading, science, history-government, and writing). IEP teams choose indicators and assign tasks to measure those indicators for each student individually in order to show the performance of the student in each assessed subject for that year. Evidence for data folios is collected during the testing window in the spring semester. After collection, three local scorers evaluate each piece of evidence to determine the extent to which each student demonstrates competence on the chosen indicators. Those scores are compiled to determine the student's level of proficiency for AYP.

Purpose of the Study

The Kansas State Department of Education (KSDE) undertakes an annual review or audit of a selection of data folios in order to monitor implementation, inform professional development, and improve the state's alternate assessment process. The KSDE audit occurs after the assessment window has closed each spring with a sample of approximately 5% of KAA data folios in all subjects. Approximately 24 reviewers meet in Topeka during the summer to inspect these data folios for evidence that the folio was properly compiled and scored. Reviewers use an evaluation rubric that is designed for this task and revised annually. This rubric includes questions such as, "Does skill targeted align with the selected extended indicator?", "Are the evidence labels complete?", and "How would you rate overall compliance with KAA requirements, as evidenced by the presentation and quality of information in this learner's folio?" In addition to rating aspects of each data folio, reviewers provide comments to explain anomalies they find in the data folio or assessment process and to provide suggestions for teachers. Information from and about the reviewed data folios is not returned directly to the teachers who submitted the folios. Instead, this information is used by KSDE to improve the KAA and guide professional development. The initial purpose of this study was to evaluate reviewers' comments on the data folios collected in 2009 and 2010.

An additional objective for this project was conceived while conducting qualitative analyses of reviewers' comments from the data collected during the audit. That secondary purpose is to improve and standardize annual data folio examination procedures in order to increase the generalizability of reviewers' conclusions. Teachers and reviewers can be conceptualized as two members of a feedback loop, with the improved capability of each enhancing the proficiency of the other. With greater competence, teachers will do a better job of adhering to the KAA process and producing valid outcomes by more accurately assigning students to proficiency categories. If reviewers follow a standardized procedure in their appraisal of the sample data folios, their results and conclusions will be generalizable to KAA-participating teachers and students across the state. This in turn will enhance and improve professional development for teachers, which brings the loop to a close.

Method

Data Collection

Data collection involved two years (2009, 2010) of reviewers' comments on data folios. Subjects included reading, writing, math, science and history/government. Comments from both 2009 and 2010 were used in this data analysis.

In 2009, reviewers evaluated 506 data folios. The 2009 data folios were organized by subject area: reading (136), math (139), writing (158), science (60), and history/government for high school (13). In 2010 there were a total of 304 data folios, but the subject of each folio was not indicated. In some cases the subject was evident from reviewer comments or the inclusion of a particular indicator number, but in general the researchers could not separate comments by subject for 2010.

Data Analysis

First, two qualitative evaluation team members coded all the 2009 comments independently. All three authors then compared and evaluated the coding systems. As a result, the team established a revised coding system for the 2010 comments. Two researchers coded the data folio comments of the first 50 folios of 2010 using the revised coding system and compared their coded comments to establish inter-rater consistency. After reconciling differences in coding, two researchers each coded approximately half of the remaining 2010 data and recoded the 2009 data as needed. Upon completion of coding, the two researchers selected illustrative comments for each theme from the data and compared their comments to confirm consistency. In general, the team was consistent in coding specific comments. The team did have some inconsistencies in the activity and evidence label themes. An example of an inconsistency was whether to code the instructions for the activity, which are described on the evidence label, in the activity theme or in the evidence label theme. The team resolved this discrepancy and other minor discrepancies by describing their perspectives and finding common ground.

Results

Themes identified from reviewers' comments fell into three major areas: those relating to the assessment process, those relating to the actual data folios received by the reviewers, and other issues. Comments addressing the assessment process involved the selection of indicators, alignment of activities to indicators, appropriateness of activities chosen for data folio inclusion, and accommodations provided (or not provided) to the student. Comments concerning the data folios themselves included information on the evidence labels, organization of the data folios, and scoring by the original three scorers. Other issues comprised non-compliance with KAA process guidelines; general praise for the teachers, activities, or the KAA process; and idiosyncratic comments that didn't fit one of the other categories. Comments from 2009 and 2010 are intermixed in the following sections unless a comment pertained to something unique about one of the years, then the year is noted. If a specific indicator was referenced in a comment, the indicator number is included.

Indicators

Comments regarding indicators occurred with low frequency. These comments were grouped into their own theme because they were distinct from the other reviewer comments and addressed an important aspect of the testing process. This theme proved to be coherent because relatively few reviewers commented on indicators and their comments were highly specific. Comments within this theme included concerns about whether the indicators were too easy and whether or not the indicators were appropriate for the students. Reviewers made virtually no comments questioning whether an indicator was too difficult. Examples of reviewer comments included:

> *Some of the indicators should challenge the student. Most students do not get ALL data collection samples correct.*

> *The intention of the assessment is to push the student to do new things. It shouldn't be an observation of known behaviors.*

> *ES 1.1.1: Student had no way to fail the task; could have asked to identify rough or smooth-not just feel the rock.*

Also included in this code are comments on missing indicators:

One indicator was missing.

Missing indicator sheet listing indicators chosen at the beginning of the data folio.

While most of the reviewer comments were negative, some were positive:

Good use of this indicator for this skill.

Alignment

Alignment was a common theme because it is up to teachers to choose individual activities for their students. The selection of unique activities by individual teachers introduces a lot of variability into the KAA. The activities teachers select do not always align with the indicators they are supposed to assess. For example, while there were no reviewer comments on this particular issue, the researchers noticed that some of the activities corresponding to indicators in the reading standard for literature used expository text materials.

Comments related to alignment were relatively consistent between 2009 and 2010. Comments that fell under the alignment theme were often not specific. Many reviewers simply stated that an activity did not align with the indicator, but they did not give further details about how it did not align:

Activity did not align with indicator.

Evidence did not align correctly.

Other reviewers mentioned what they thought was wrong with the activity. Reviewers wrote that teachers performed the wrong task with their students:

The indicator is collecting data, not interviewing.

The evidence is supposed to be about identifying characters.

EM 3.1.2: Putting into groups is okay, but it should be by color, shapes, material (wooden/metal) not groups of 2's, 3's, etc.; How do groups of words go with sorting math manipulatives?

EM 3.2.1: The student is not using measuring tools-they are estimating.

ES 1.1.1: Evidence must show exploration of environment, not just following a routine.

ES. 2.1.5: Describes objects by multiple properties; trials were about process not indicator.

ES 3.1.3: The task is about body systems. Description does not match the indicator; talks about weather rather than body systems.

ESS 1.3.1: All samples are question/answer situations that do not relate to basic rights of Constitution (i.e. free speech). Responding to questions does not relate to indicator of basic rights in Constitution.

ESS 3.3.1: Activities of fire drill and lights off for movie do not match 'earth's physical systems'

Comments falling under the alignment theme were also related to how well the teachers described the alignment on the evidence label. The evidence label instructs teachers to explain how the activity aligns to the indicator. However, reviewers often commented that teachers did not provide enough information in their explanations:

Yes the activity was in the clarifying examples, but how for this student does it serve as evidence of understanding the indicator? Be more specific...that is why explanation is necessary in box A to clearly show how the activity relates to the indicator.

Description was more of a student goal, not how the task was aligned to the indicator.

Finally, a small number of reviewers commented when they thought that the activities aligned well with the indicator:

Activities align well with indicators.

Activities

The activity theme was highly diverse and more heterogeneous than most of the other themes. This is largely a reflection of the data and the KAA

process. The kinds of activities teachers select for the KAA are quite varied; therefore, the potential issues related to those activities are numerous. Comments in this theme included concerns about difficulty, appropriateness, administration, and variability. Negative reviewer comments far outnumbered positive reviewer comments, but both are included in this theme since both addressed characteristics of the activities teachers selected. The comments in the activity theme did not significantly differ between 2009 and 2010. Reviewers raised many of the same issues in both years.

Reviewers often expressed their opinion that certain activities were too easy or not appropriately challenging for students, or that the elements of the activity were not age-appropriate for the student. Lack of difficulty was a very common issue in the activity theme:

> *I wonder if activities might be a little too easy if he earns all 5s.*

> *The activities selected seem to show such a wide range of activities from very simple to more complicated. It makes us wonder if some of the activities are too low.*

> *Consider being more respectful (of your student's age) in the character choices [in reading materials and worksheets].*

> *We find it interesting that every student you have received all 5s in all areas, especially considering how low some of your students appear to be.*

> *When students receive all 5s, it makes us wonder if the activities selected were truly at the student's instructional level. Were the activities low-balled?*

Reviewers also indicated when they thought an activity, or the way in which it was done, was too difficult, but such comments were not common:

> *Activities seem rather high for such a low indicator.*

> *Why would you continue to have this student try to print when there are motor issues-have you considered a name stamp and other assistive tech to allow the student to advance?*

EM 3.1.4: When a student fails all trials, the question becomes why? Does the student not understand 'square or round' or the context in which the question is asked?

Another issue reviewers raised was lack of activity variability, either within tasks, between two different indicators, or even between students. Many reviewers also commented that reducing activity variability made the assessments too easy for the students:

Why would you have the student complete the tasks on 3 different dates when she demonstrated mastery on the first time? You need to change the task in some way...always doing the same thing when mastery has been demonstrated does not reveal all that you have taught the student.

Consider varying the task...it would be stronger evidence that the student understands the geometry of this indicator.

Some of the indicators were extremely basic even for a severe student; 15 times of throwing away towels or turning on lights for example.

The weather check sheet was overused and not sufficient data to show student had mastered each indicator. Use different samples for different indicators.

EM 2.1.2: Indicator states variety of formats but all were worksheets with pictorial with shapes.

EM 3.1.2: Assessment task is identical to task from 2.3.2 Need details to indicate how sorting task differed from identifying task in 2.3.2.

EM 4.1.3: Need variation of consequences or events.

Sometimes it is possible to have identical activities because several students are at the same level, but please remember that it should reflect the individual student as much as possible.

Comments regarding activity administration were also included in the activity theme. Reviewers sometimes thought that teachers either administered activities improperly or inconsistently with KAA procedures.

Many of these comments reflect a practice by some teachers to conduct most or all trials of an activity on the same day, which many reviewers did not consider good practice:

> *Working on one trial together should not be part of the assessment.*

> *On data collection, rather than asking the same question five times on the same day with the same question, change at least one of the factors to make the test more valid.*

> *It was undeterminable how the five trials were unique because the description of the trial and the student response make it appear that each trial was identical and all done on the same day.*

Reviewers also made comments regarding functional indicators. Functional indicators, or indicators that measure non-academic behaviors that students perform in their everyday lives, were present in 2009 but were not allowed for the 2010 assessment, and opinion- or preference-based activities were not allowed in either year. Despite this, reviewers were concerned that some selected activities were functional or opinion-based in nature:

> *Pushing a button to get a drink of water is a functional activity.*

> *The first activity is close to being functional and opinion based.*

> *ES 5.1.1: Toileting for cause-effect is not appropriate for assessment.*

> *ER 2.4.2: Behavior during the media presentation might be more appropriate. More attending to media rather than expressing preference.*

Clarifying examples of activities aligned with the extended standards have been available to teachers for many years. However, teachers may have misinterpreted the purpose of an example:

> *The activity is in the clarifying examples, but the twist you put on it adds a dimension of asking for help being measured as opposed to understanding that it is impossible to open the door; that is what should have been measured.*

Reviewers noted when teachers did not appear to understand either the content or the purpose of an indicator and therefore selected an inappropriate activity:

EM 2.1.3: I don't believe you understand the mathematical concept of pattern. Completing adding only complicates it. A math pattern is a relationship between them. Coloring a pattern doesn't meet that.

EM 2.4.2: A Venn diagram is used to show same and different attributes. The intersecting circle should show the same attributes that a bank and school share. Boy/girl mom/dad may not be best objects to show separate attributes because of family configurations, nontraditional roles or cultural background.

EM 1.1.4: Rote counting implies orally counting. You might state 'count 1-30.'

EM 3.2.2: A visual schedule is not a calendar.

EM 4.1.1: When you ask the student to push play or stop when finished they are following your directions not seeing cause and effect. The same would go with turning lights off or on.

EM 4.1.1: "Push cart" is not a cause and effect task.

ES 4.1.1: Some of the things you selected are not earth materials. Starfish, seed, grasshopper are living things-earth materials are rocks, dirt, water.

ES 4.2.1: You counted looking up at classroom lights as a '+' response. The classroom lights are not objects in the sky.

EW 1.4.1: One-on-one is not sequencing.

ESS 3.1.4: Asking 'Where is the blue water?" "Where is the US?" might be more appropriate to show learning than spinning the globe.

ESS 4.1.2: Not all items are historical figures (needs to be related to person)

EM 2.1.3: Examples 2 & 3 are not really patterns-more like matching.

Finally, some reviewers praised various aspects of the activities including the worksheets, examples, and directions:

> *Worksheets are well constructed and age-respectful.*

> *We liked how you included pictures of activity. It helped us understand better.*

> *Good direction given to student. Very clear.*

Evidence Labels

The evidence label theme included the completeness and quality of the information entered by the teacher. In 2009, many comments dealt with evidence label problems. Revisions to the evidence label were instituted in 2010 and resulted in fewer problems noted by reviewers. Some comments related to the evidence labels were better placed in another theme as the comment dealt with that theme more specifically. Overall, the evidence label theme was more diverse in terms of the variety of comments.

The evidence label accompanies and explains each piece of evidence (student-completed worksheet, performance activity, drawing, photograph, etc.) submitted as proof of the student's competence in completing the indicated skill. Hence, the label requires much necessary information: student's name and grade level, indicator, activity description, instructions given to the student, time/date/location of this skill assessment, accommodations used, and explanation of the link between activity and indicator. Consequently, many comments from the reviewers dealt with missing or incorrect information, such as:

> *Evidence is not consistently marked with an evidence number, date, and indicator number.*

> *It is unclear how many trials there were for the activity.*

> *Instructions are always given to student, even if it is not verbal; please list what was said/shown to student.*

> *EW 1.1.4: If student needs to ask a question for assessment, have instructions say 'student will ask nurse for meds' or 'student will say yes when the nurse asked if they are here for meds.'*

EW 3.1.2: Have your instructions reflect the assessment (i.e. "say hello to the teachers in your room" if your assessment is to say hello to an adult).

EM 3.1.1: Assume you used 3-D shapes (sphere, cube, pyramid) rather than the 2-D shapes listed.

Other comments indicated the label was completed but with minimal information, making evaluation of the data folio difficult for the reviewers. The following comment summarized this issue well:

Evidence tended to follow the letter of the law with KAA regulations, but it is very difficult to see what was done with the student on several indicators because everything was brief and little explanation was made.

Reviewers also commented occasionally on the completeness of the information on the evidence labels. For instance, two reviewers noted:

Thank you for the accurate description and detail on the data documentation part of the label. It made it easy to see what you did with the student

Descriptions are very complete; we easily could understand the nature of the task.

Accommodations

The accommodation theme referred to the type of accommodation, the number of accommodations used, or if an accommodation had been used at all. Any accommodations used on the KAA must be noted on the evidence label. Probably because of revisions to the evidence label, this theme had fewer comments in 2010 than in 2009, and it was a fairly coherent theme. Some comments referenced the type of accommodation that may have been used on the activity but that was not specified on the evidence label:

Please be sure to indicate if a scribe was used (it obviously was).

Please list calculator as an accommodation.

> *Accommodations should be more specific. Who is doing the reading and writing?*

Accommodations could also refer to the quantity of accommodations used by the student. This theme could also show a lack of accommodations used by the student as indicated on the evidence label:

> *List only accommodation used, not all possible.*

> *Does the student really need no accommodations????*

Of a more serious nature were comments that indicated teachers may not have understood the purpose of accommodations or the methods and materials that may be used for accommodations:

> *Accommodations need to be task driven, not the same on all evidence labels.*

> *Specify what you are prompting; if you have to prompt the response isn't valid.*

> *Redirect, restate, staff writing answers, answer orally are accommodations NOT 'hands on activity'.*

> *'Minimal assistance' is not one of the accepted accommodations- describe what is meant.*

> *'Staff will assist' is not one of the accepted accommodations.*

Specific comments about amount of prompting occurred more than a dozen times in the 2010 data:

> *Unlimited prompts are not allowed.*

Occasionally a comment showed that the reviewers noticed the effort to provide accommodations for the student on the part of the teacher:

> *Great explanation of accommodations.*

Organization

This theme included a variety of comments related to organizational issues. There were more comments in 2010 about organization than there were in

2009. Organization comments related to the overall organization of the items in the data folio, ease of understanding the materials, and the inclusion of all needed items. Therefore, comments often defined areas teachers needed to improve upon:

> *Many typos-made it difficult to determine what they were talking about.*

> *Fill in each box, no ditto marks.*

> *Please do not copy on both sides of the paper; front to back, more difficult to evaluate and score.*

Other reviewers asked teachers to include missing information:

> *No administrative signature page.*

> *Use specific forms from the state website.*

Other comments indicated the reviewers noticed when the data folio was well assembled:

> *Very thorough data folio with nice examples and clear directions.*

> *Very easy to read.*

Teachers also occasionally forgot to include student work (name of book student read, picture student drew) that would clarify the activity attempted:

> *There is not a worksheet to show evidence of performance of the skill.*

Reviewers obviously appreciated the inclusion of student samples:

> *It was nice that samples were included.*

> *We enjoyed seeing the student completed worksheets!*

Scoring

The scoring theme included comments on both the scoring done by the teacher of the activities and the original scoring done by the three local scorers. Scoring received a proportionally larger number of comments in

2010 than in 2009. This was a fairly coherent category both years. Often the comments indicated concern that student activities were not scored or were scored incorrectly:

> *Worksheets have incorrect scores marked (No scores should be on the worksheet) and answers are not marked correct/incorrect.*

> *EM 4.1.2: Observation is not an appropriate way to do this (recognition of possible/impossible).*

Other teachers used unique scoring codes without explaining them. From the comments, it appeared that reviewers preferred teachers to use five problems or attempts by the student for each piece of evidence so that the score could be +5/5 or some variation depending upon the student's expertise in the skill:

> *Teacher needs to use utilize the data sheets provided to help clarify the scores.*

> *After looking at the score sheets it looked like the scorers had difficulties figuring percentages where there were more than 5 data points; we suggest that you have a number of problems that were divisible by an easier number.*

Other comments indicated more serious problems which need to be addressed in the future:

> *Test unscorable! No effort by teacher. No data! Only data collected by music therapy.*

> *The scores for this assessment are totally elevated, but this reflects only that the student completed the tasks requested.*

Some comments seemed to show that teachers lack understanding of academic skill indicators versus behavior. The KAA is focused upon meeting academic skill indicators rather than behavioral objectives:

> *The incorrect response indicates the measurement of a behavior. The student must respond correctly or incorrectly. A response of 'no attempt' indicates the measurement of a behavior not the measurement of a skill.*

The reviewers also noted scoring methods that were well understood and completed correctly by teachers:

> *We really appreciate how you scored the worksheets and entered the data in the evidence labels.*

Additional Indicator-Specific Comments from 2009

In the 2009 data folios, many reviewers made comments specific to a particular indicator. Many of these comments could be helpful to other teachers in understanding how to teach and assess these indicators. The following comments are divided by subject and noted by indicator when it was provided by the reviewers:

Reading

ER 1.3.4: If you want each sequence step to count as a separate trial, you need to work it as 'each step' or 'what comes / happens next?' otherwise it is seen as one trial, not 5.

ER 1.4.1: To determine if student recognizes a word from memory, show him 3-4 cards from which to choose the correct word.

ER 1.5.7: Have the student put pictures of expressions with 'raining cats and dogs' with the phrase 'downpour of rain.'

ER 1.5.9: Understanding humor might be better expressed by showing pictures (clown/man in suit)-which is funny? Not everyone things the same jokes are funny. Or use stories or pictures of events that are funny for student to identify, then include data chart of responses. Is the student laughing at the joke or your voice or facial expression?

ER 2.1.1: The evidence is supposed to be about identifying characters. You could have used the same activity but the questions should have been about the characters. For example, instead of 'What grew by Ditto's door?' ask 'Who had flowers by their door?'

ER 2.1.4: Could the students use a picture to align with activities like 'What have we done this morning?'

ER 2.4.1: You could show the student 5 books and say 'pick the one about dogs.'

ER 2.4.2: Use three different books or media to give most exposure possible when appropriate.

Math

EM 1.2.1: Activities did not match numerals. For example, put a numeral on a card and have student match to room number or locker number; perhaps match to phone number; perhaps match 2 cards with the same numeral on it.

EM 2.1.3: We need more detail about how Jackie Robinson, Dog Show, and Jonas Brothers are symbol patterns.

EM 2.1.4: Great change up using music to generate a pattern.

EM 2.3.2: Have the cue match the indicator. If you say 'show me one that is the same' and you hold up the football that would be measuring the indicator.

EM 3.1.4: What are the individual shapes for each week? Suggest stick to one shape, or do 5 individual shapes.

EM 4.2.5: Battleship activity was a clever way to measure the skill.

Writing

Rather than tracing personal info, can the student 'write personal info' (i.e. type, write in boxes, complete missing letters/numbers, put letters in order)?

EW 2.2.4: Written language is about expressive. Tweak activity so student is expressing him/herself to get at the expressive part.

Science

ES 3.2.1: It would have been better to explain what the student actually did to observe the chicks (touch the incubator or turned to look at the box, etc.).

ES 4.1.1: The task was done inside but it should naturally have happened outside. It should have been done on a day when the weather was nice.

ES 5.1.1: To measure cause-effect the directions might be 'when I turn out the lights, what happens?' Answer would be 'the room gets darker.' She could indicate any way she needs.

ES 6.1.3: Do picture cards include snacks that are not healthy so that student is distinguishing between healthy and unhealthy?

History/Government

ESS 4.4.1: Have student identify symbols (i.e. flag, eagle, etc. or how her family celebrates a holiday, put symbol on calendar to mark date, etc.)

Other Reviewer Comments

Some reviewer comments did not fit into the defined themes of the assessment process or the contents of the data folios. Wherever possible, positive reviewer comments were placed within the specific areas they addressed. However, many reviewers gave general praise that was less specific and often targeted towards entire folios. Comments within this subtheme are consistently broad and appeared in similar low frequencies in 2009 and 2010:

Nice data folio – good job

Excellent data folio!!

The teacher did a good job with an obviously low functioning student.

Reviewers sometimes made comments that were very specific, either negative or positive. These comments included statements directed at particular teachers or school districts. Some examples are listed below:

We recommend that everyone involved in completing alternate assessments [in this school district] receive additional training. There are numerous errors that could be eliminated with training from

someone who more fully understands the process. Training should be for administrators as well as for teachers.

It was difficult to determine the skills of the student by looking at this Folio…there were a number of issues: The setting indicates that the student was home-bound. It appears that more training is necessary for the home-bound teacher…We do not feel that there is adequate evidence that the student can or cannot do the items. If a student is not testable, the district might need to consider an exemption for the student.

You might be a good resource to your district for knowing how to put together an exemplary data folio. Passes the stranger test with flying colors…this would be an excellent sample to show other schools/teachers for how to assess a very low functioning, physically involved student.

Finally, reviewers' concerns regarding questionable data folios denote an important assessment issue. Comments in this subtheme are memorable but not very common among the overall responses. Reviewers made comments when they suspected folios may have been assembled improperly and when teachers did not follow established procedures:

It appears that evidence 1 and 2 are the same worksheet with identical markings but on one copy the evidence sample number 3 was crossed out and number 1 written in. Both labels show the date of 1/14. We believe this is the same work used twice.

Data was not collected in the testing window! Data used can only be data collected during the testing window!

Conclusions from Reviewer Comments

After examining the data from 2009 to 2010, there were fewer comments in 2010 about local scoring and organization of the data folios. More of the 2010 comments addressed the teachers' skill and interest in building the data folio. Also, a higher proportion of data folios received comments in 2010.

Reviewers made several comments questioning the age-appropriateness of the materials used with students. Other reviewers noted that the materials used would have fit another indicator better than the one chosen, or that the materials belonged in reading rather than in a math activity. For the purposes of accommodations training, some teachers seemed to believe that an accommodation is the setting, such as a basketball game. Other teachers needed to better clarify how many minutes a student may take in responding to a question or the number of prompts that may be used to encourage a student's response.

Local scorers could also benefit from repeated training in scoring data folios and having planned opportunities to develop inter-rater reliability. Perhaps a checklist of procedures to follow in compiling the data folio would be helpful for the teachers and administrators involved in this effort, as well as encouragement to use the KAA Teacher's Guide from KSDE (2010).

From the data analysis, it was evident that teachers could benefit from professional development and personalized guidance in aligning activities with indicators, use of accommodations, and choosing appropriate indicators for their students. Providing teachers with specific examples in which activities align with indicators as well as examples for which activities clearly do not align with indicators may help to improve the issue of alignment. Communicating with teachers about what accommodations are approved by the state and providing a checklist of accommodations that teachers could easily use in completing the evidence label may encourage teachers to list all accommodations on the evidence labels and thus reduce future issues related to accommodations. Furthermore, teachers could benefit from gaining access to de-identified data folios in which IEP teams chose appropriate and inappropriate indicators. Overall, teachers would benefit by being granted access to both outstanding and mediocre data folios. Otherwise, the issues that reviewers of the KAA brought up are likely to continue.

Recommendations for the Review Process

The process of analyzing and coding reviewers' comments from two years of sample data folios led to an awareness of ambiguities in the review process itself. A tighter link between this process and professional development

would improve communication between reviewers and the teachers and scorers who work directly with data folios.

It would be helpful to articulate the purpose of the folio review. Is it simply to provide feedback to KSDE on what professional development topics need emphasis? If that is the case, and teachers know there are no real consequences for their portfolios, teachers may consider this an unrewarding task. How are the results used and shared with others? What does KSDE do if serious errors and omissions are uncovered in data folios? Is there any follow-up with specific teachers, schools, or district administrators? Has the review process itself been reviewed or evaluated? Reviewers should be accountable for the value and objectivity of their work, just as teachers are held accountable for the quality of their instruction and assessment results.

One of the questions that arose during this qualitative study was whether the comments that were recorded were representative of problems in Kansas data folios in general or whether they were simply the result of reviewers' idiosyncratic perspectives and opinions. In order for the outcomes of the folio audit to be generalizable, or applicable to folios that were not part of the appraised sample, the process must be well defined and the consistency of the reviews and prevalence of the identified folio problems must be established. In other words, if a particular student data folio exemplifies a shortcoming that could be addressed through professional development, any reviewer or team of reviewers should be equally capable of identifying and describing that shortcoming. Without evidence of the skill of the reviewers and the similarity between reviewers, or inter-reviewer reliability and consistency, it is difficult to know whether the problems found in these folio samples reflect individual reviewer knowledge and skill or systemic issues. Because there may be unevenness in terms of which problems were noted and commented upon by individual reviewers, or disagreement between reviewers about what constitutes a problem, one cannot draw conclusions about the pervasiveness of these problems and how much professional development should be devoted to them. Some of the folio issues identified through this review are clearly serious and demand attention, while others may be uncommon or unimportant, or they may represent the tip of the iceberg of deficiencies that have gone largely unnoticed.

For example, not all folios received comments from reviewers, even though some of those folios were rated as having serious shortcomings. Comments about the specifics of those inadequacies may have been useful for professional development. More to the point, constructive comments about how to rectify observed problems would be directly applicable to teachers in the field. What is the intended purpose of the comments? If the review criteria are clearly defined and comprehensive, then folios that meet those criteria may not need additional comments. However, if any criteria aren't met, then comments would be useful to describe deficiencies and recommend solutions.

The procedures for this audit of data folios have not yet been described. The annual review has only recently been implemented, which suggests that now is an excellent time to define the process in order to make stronger inferences from this activity. An incremental process of refinement each year will strengthen the conversation between teachers and reviewers and continually improve the quality of the feedback that reviewers provide.

To begin with, it would be useful for a panel of teachers to evaluate a selection of de-identified data folios and reviewers' evaluations to give advice on the usefulness of the review outcomes and suggestions for how reviewers could provide more guidance to struggling teachers. This would close the feedback loop and could be a precursor to professional development offered to all teachers on how to become a folio reviewer. Another suggestion would be to have teachers who submitted weak data folios defend their actions to the reviewers. This would open a discussion with reviewers about issues such as different interpretations of alignment and selections of tasks.

Next, the process by which reviewers are selected and trained should be described. This not only refers to individual reviewers, who should probably be experienced teachers of this population to begin with, but also the method of organizing reviewers into review teams. Two main methods of managing variation are random assignment and stratified assignment of the variable feature. Random assignment would consist of randomly allocating all reviewers to teams, probably followed by reconstitution of teams from time to time—from once an hour to once a day, for example. The alternative is control of the feature by assigning reviewers to teams on the basis of their

expertise, such as by content area knowledge or age and grade level teaching assignment. A related question concerns the optimal size for teams. There is doubtless a trade-off between the deeper scrutiny larger teams may offer versus the efficiency of smaller teams processing a similar number of folios each hour. Smaller teams would enhance active participation by all members, though the converse is that smaller size may increase differences between teams and hence inconsistency of reviews.

Given that many or most special education teachers for this population probably have primary special education qualifications and secondary qualifications, if any, in tested content areas, perhaps content area knowledge should be distributed across teams rather than concentrated. On the other hand, if the makeup or expectations of data folios differ by grade band, perhaps teams should represent restricted grade bands (elementary, middle school, high school) and then score folios only for those grades. A crucial activity will be to have a set of data folios reviewed by pairs of teams so that inter-team reliability can be established. Each review team would be trained to find the same problems and give the same feedback to teachers. Reliability could be maintained across years by training on anchor folios selected to illustrate key problems and issues.

Another issue is the team recorder. Selection of the recorder appears to be the most salient variable in the whole process, since the recorder controls the amount and quality of the information obtained from the team. Is this a volunteer or is this position assigned? Do recorders receive any additional training? Variability in folio evaluation, particularly the optional comments, may be a function of how the recorder attends to this task. Perhaps the position of recorder should be voluntary since not all reviewers have strong keyboarding skills or are able to listen and take notes simultaneously. Perhaps each team should use two recorders for assessment of reliability, or the recorder role should rotate among team volunteers to avoid fatigue. Recorders could also rotate among teams, particularly if recorders have unique skills or training that improves their performance in that role. Rotation could bring a level of consistency to team output. Specific training for that function could also be considered.

Some of the reviewer and team questions recommended for consideration include:

- How are reviewers evaluated for their expertise to perform the folio appraisal?
- Are there minimum requirements to serve as a reviewer?
- Has systematic reviewer training been prepared?
- How can it be maintained and improved from year to year?
- What is the optimal team size?
- Should all teams consist of the same number of reviewers?
- Should teams be organized by content area expertise or grade level, or should they be randomly assigned?
- Should teams have rotating or static membership?
- Do reviewers from a district serve on the same team or on the same team with a supervisor from that district?
- How is the team recorder chosen?
- Are specific skills or training required of the recorder?
- Should two or more recorders trade off to avoid fatigue?
- Should recorders rotate among teams?

Another set of issues concerns the review process itself. With 300-500 data folios to be reviewed each year, what resources are required? How many people and how many teams are needed to conduct a valid review and produce usable results for KSDE? What are the key issues for evaluation? Each year's outcomes can inform the subsequent audit as part of the feedback loop in which review outcomes are used for continuous program improvement. For example, the initial review in 2009 revealed that teachers sometimes did not collect data until the final days of the testing window. For 2010, enhancements to the KAA process included the requirement that indicators for assessment be selected and approved by an administrator before December 31, 2009. Because of the previous use of functional activities, which were no longer permitted in 2010, administrators also had to sign off on a statement that no functional activities were included. These improvements reduced the number of problems in these areas.

Some questions for reflection include:

- How long did it take to complete the folio audit in 2009 and 2010?
- How many folios must be processed each hour?
- Is there a decline in the quality of reviews conducted later in the day?

- What critical issues were identified in the 2010 audit that will be targeted in 2011?
- When reviewers have questions or disagree, what resources are available to resolve disagreements?
- Have former reviewers been solicited for their feedback on the process?
- Has the evaluation rubric been critiqued by reviewers and by teachers in the field?
- Are there critical items that are missing from the rubric or are there unnecessary questions?
- When should comments be made about a data folio?
- Are there situations in the review process that should require comments?

All of these questions should be considered guides for future planning and not issues that must all be resolved immediately. Each year, small improvements will lead to enhanced professional development for all teachers. Greater transparency of the review process will give teachers in the field better knowledge about what's expected of their folios, while teacher feedback on the review process will in turn improve its utility. Finally, if reviewer outcomes are analyzed in the future, it will be easier to judge whether their conclusions represent a balanced and reliable assessment of the quality of KAA data folios as a whole.

As an example of improvements apparent in this study, the revised evidence label generated fewer problems and comments in 2010 than were made in 2009. Strong conclusions about the prevalence of specific issues throughout the state, such as improper or inappropriate tasks, evidence, and scoring, can be drawn when results of the annual review are obtained from consistent reviewer training and auditing procedures. Effort should be made to obtain more substantive and useful generalizations from the annual review data and to share these conclusions with teachers of KAA students.

Appendix

2010 KAA Qualitative Codes and examples

Comments for the KAA process

1. **Indicators:** Whether the indicators are appropriate for the student (too easy, too hard) and whether the indicator should be in use (e.g. functional indicators).
2. **Activities:** Includes comments on all aspects of activities (whether the activity was administered correctly, whether the activity was too easy or too hard, praise of activities, etc.)
3. **Alignment:** Did the activity align with the indicator?
4. **Accommodations:** Any comments regarding student accommodations.

Comments for the Evidence folios

5. **Evidence labels:** reflects quality/completeness of the evidence labels and whether fields were filled out correctly (unless the comment is more specific and better accounted for by another code such as scoring).
6. **Scoring:** reflects whether raters were able to score the assessments, whether pass/fail criteria was included, and all other comments related to scoring the assessments.
7. **Organization:** Reflects more general organizational issues such as whether all materials were present, in order, were easy to read, etc.

Miscellaneous

8. **General praise**: for example "nice portfolio!" or "good job!"
9. **Idiosyncratic comments:** very specific rater comments that do not fit anywhere else
10. **Questionable data folios:** When raters suspect that teacher is not correctly following the KAA system or questionable assessment practices were followed. While the intent of the teacher may not be apparent, some comments are clear indications of improper activities.

References

Browder, D. M., Karvonen, M., Davis, S., Fallin, K., & Courtade-Little, G. (2005). The impact of teacher training on state alternate assessment scores. *Exceptional Children*, *71*, 267-282.

Browder, D. M., Spooner, F., Algozzine, R., Ahlgrim-Delzell, L., Flowers, C. & Karvonen, M. (2003). What we know and need to know about alternate assessment. *Exceptional Children*, *70*, p. 45-61.

Hanzlicek, V. A. (2006) *The Kansas Alternate Assessment and instructional planning for special education teachers: A case study of implications for students with severe disabilities* (Doctoral dissertation), AAT 3223355, DAI-A 67/06, Dec 2006. Retrieved from ProQuest Dissertations and Theses database.

Kansas State Department of Education, Special Education Services. (2010). *Kansas Alternate Assessment teacher's guide 2010-2011.* Retrieved from http://www.ksde.org/LinkClick.aspx?fileticket=X80HpN0L90Q%3d&tabid=2384&mid=8879

Rabinowitz, S., Sato, E., Case, B. J., Benitez, D., & Jordan K. (2008). *Alternate assessments for special education students in the Southwest Region states* (Issues & Answers Report, REL 2008–No. 044). Washington, DC: U.S. Department of Education, Institute of Education

Sciences, National Center for Education Evaluation and Regional

Assistance, Regional Educational Laboratory Southwest. Retrieved

from http://ies.ed.gov/ncee/

Roach, A. T., Elliott, S. N., & Webb, N. L. (2005). Alignment of an alternate

assessment with state academic standards: Evidence for the content

validity of the Wisconsin Alternate Assessment. *The Journal of Special

Education*, *38*, 218-231. doi: 10.1177/00224669050380040301