Preparing Job Trainers to Describe Knowledge, Skills, and Abilities Measured in an Academic Assessment[1]

Commentary on Computer-Aided Bookmarking (CAB)

Luz Bay

Measured Progress[2]

Contents

Figures

Tables

Introduction

In 2010, Measured Progress was awarded a contract to set standards for academic preparedness as part of an initiative of the National Assessment Governing Board (NAGB) to explore the utility of the National Assessment of Educational Progress (NAEP) as a tool to predict students' academic preparedness for entry into post-secondary education or job training programs. The statement of work for the project provided a Design Document to set standards specifying the Bookmarking method to be used in the judgmental standard setting (JSS) process (National Assessment Governing Board, 2010). Further, there was a request to computerize a specific part of the Bookmarking implementation -- the annotation of knowledge, skills, and abilities (KSAs) required to answer each item correctly -- an integral part of the method. In response, Measured Progress proposed an implementation that not only computerizes the KSA annotations, but implements a process that is fully computer-based. This marked the birth of Computer-Aided Bookmarking (CAB).

The implementation of CAB is consistent with the NAEP tradition of setting achievement levels (ACT, Inc., 2010, 2007). A special feature of the JSS for academic preparedness is that panelists started their training on NAEP, the appropriate NAEP framework, and participated in the drafting of the borderline performance descriptor (BPD) by attending a two-hour webinar and answering survey questions. The first draft was presented to them by the content facilitator who prepared it. The borderline performance descriptor (BPD) was edited and finalized during the on-site meeting with the content facilitator's help.

On-site standard setting activities started with orientation and rigorous training that prepared the panelists for rounds of bookmarking. A set of feedback was presented to the panelists after each round, and where they place their bookmarks on subsequent rounds is based on their understanding of the BPD, their understanding of the items, and information presented as well as discussions between rounds. There were five evaluation questionnaires, each one administered after each major part of the process, as well as a consequences data questionnaire that gave panelists a final chance to consider their recommended cut scores.

The computerization of the Bookmarking method did not only improve the efficiency of the standard-setting process, it also brought to light issues that undoubtedly were always in existence but have not gotten attention. For example, when panelists were asked to annotate the knowledge, skills, and abilities (KSAs) necessary to answer each item correctly, were the panelists' annotations actually related to the KSAs and were they able to properly identify KSAs? The ability to collect this type of data efficiently, which was afforded by CAB, can lead to an implementation that is much more effective.

This paper will provide some details of CAB as background information for this particular implementation of the Bookmarking method. Findings regarding the type of information that JSS panelists regarded as KSA annotations will be presented. Process improvements used in the JSS process based on the findings will be discussed. Lastly, recommendations for further improvements will be presented.

Background: Computer-Aided Bookmarking for NAEP Judgmental Standard Setting for Academic Preparedness

In 2004, the Governing Board began to explore the utility of the NAEP as a tool to predict students' academic preparedness for entry into post-secondary education or job training programs, forming a Technical Panel on 12th Grade Preparedness Research that was tasked with assisting the Governing Board in planning relevant research and validity studies (National Assessment Governing Board, 2009). The Technical Panel recommended a multi-method approach to exploring the feasibility of reporting post-secondary preparedness on the 2009 Grade 12 NAEP scale for mathematics and reading. One of the four methodologies proposed included a series of criterion-based JSS studies to identify reference points on the NAEP scale that indicate academic preparedness for placement in credit-bearing, entry-level courses of the sort that fulfill general education requirements or eligibility for entry to job training programs in specified occupations. Details of the studies are in the *JSS Process and Technical Reports* (WestEd & Measured Progress, 2011; Measured Progress & WestEd, 2012).

Judgmental Standard Setting for Academic Preparedness

The primary objective of the JSS studies was to obtain cut scores on the NAEP scale that represent academic preparedness for entry into credit-bearing college courses or job training programs selected by the Governing Board. The Governing Board selected for inclusion in this project the following six post-secondary activities: 1) college, 2) automotive master technician, 3) licensed practical nurse, 4) pharmacy technician, 5) computer support specialist, and 6) heating, ventilation, and air conditioning technician. In order to maximize standardization of the JSS process across the post-secondary activities, the Governing Board developed a Design Document (National Assessment Governing Board, 2010) to guide all aspects of the project's implementation. The Design Document stipulated the use of a modified Bookmark methodology and a replicate panel design to assist in evaluating reliability of results.

A pilot study to evaluate the methodology, materials, and logistics was also mandated. A total of four sessions were held in 2011. The first session was the pilot study that included the college and automotive master technician panels. The other three were operational sessions and the pairings of post-secondary activities are indicated in Table 1. Also included in Table 1 is the assignment of content and process facilitators as well as the number of panelists who participated. The panelists were teachers in the training program for each post-secondary activity. Particulars of their recruitment are detailed in the JSS Process Report (WestEd and Measured Progress, 2011). The process facilitation structure for each session of the replicate panel design is presented in Figure 1.

Figure 1: Facilitation Structure
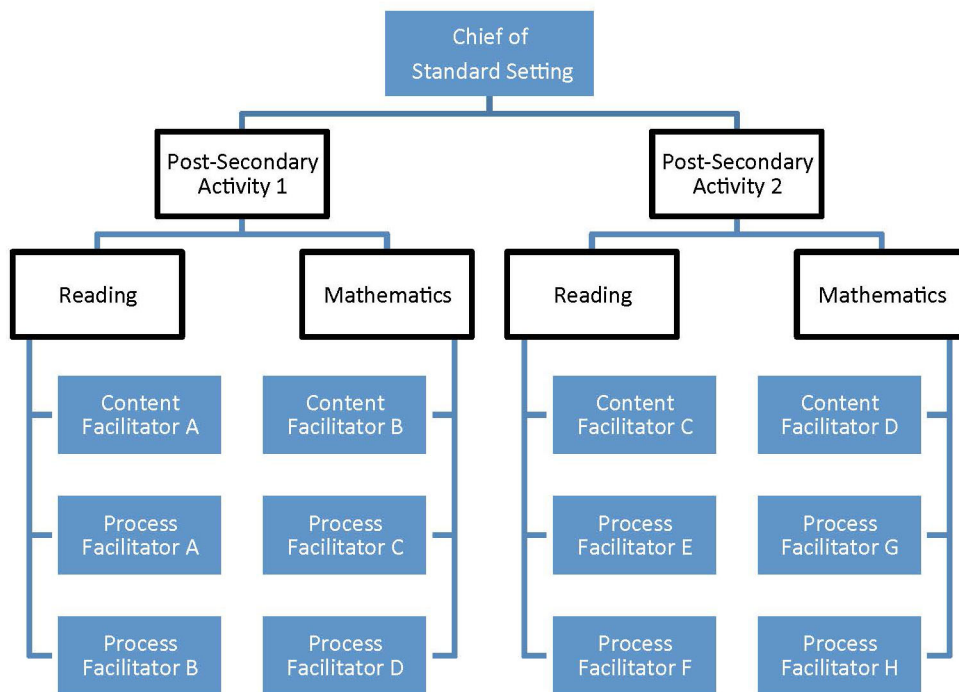
Table 1. Operational Workshop Design

Session	Workshop	Content Area	Content Facilitator (CF)	Replicate Panel	Process Facilitator (PF)	Number of Panelists
Pilot Study	College-Preparedness	Mathematics	CF1	A	PF1	6
				B	PF2	6
		Reading	CF2	A	PF3	6
				B	PF4	6
	Automotive Master Technician	Mathematics	CF3	A	PF5	5
				B	PF6	5
		Reading	CF4	A	PF7	4
				B	PF8	5
Operational 1	College-Preparedness	Mathematics	CF1	A	PF1	10
				B	PF2	10
		Reading	CF2	A	PF3	10
				B	PF4	9
	Automotive Master Technician	Mathematics	CF3	A	PF5	7
				B	PF6	8
		Reading	CF4	A	PF7	5
				B	PF8	6
Operational 2	Licensed Practical Nurse (LPN)	Mathematics	CF1	A	PF1	10
				B	PF2	10
		Reading	CF2	A	PF3	10
				B	PF4	10
	Pharmacy Technician	Mathematics	CF3	A	PF5	10
				B	PF6	8
		Reading	CF4	A	PF7	10
				B	PF8	10
Operational 3	Computer Support Specialist	Mathematics	CF1	A	PF1	10
				B	PF2	10
		Reading	CF2	A	PF3	10
				B	PF4	10
	Heating, Ventilation, and Air Conditioning Technician (HVAC) Technician	Mathematics	CF3	A	PF5	10
				B	PF6	9
		Reading	CF4	A	PF7	10
				B	PF8	9

The Bookmark Standard-Setting Methodology

As prescribed by the Governing Board per the Design Document (National Assessment Governing Board, 2010, p. 1), a modified Bookmark standard-setting method was used for the pilot study and the three operational JSS sessions. The specific variation of the process used was developed for NAEP achievement level setting by ACT, Inc. (ACT, Inc., 2007; ACT, Inc., 2010). Within this process, panelists reviewed assessment items that were ordered by difficulty based on item mapping using a response probability (RP) criterion of 0.67, starting with the easiest item and progressing to the most difficult. Each multiple-choice (MC) item was mapped to a point on the scale where an examinee will have a 0.67 probability of responding correctly. Each constructed response (CR) item is mapped to the scale as many times as there are partial- or full-credit scores. Each mapping represent a probability of 0.67 of obtaining that particular score or higher. The items were physically compiled in a book referred to as an Ordered Item Book (OIB). They evaluated each item against a description of borderline performance until they came to an item they judged to be too difficult for students at the borderline of the performance level. A bookmark was placed immediately preceding that item to locate the cut score. The cut score set by a panelist was the scaled score of the item that immediately preceded the bookmark. Within each group, or panel, individual panelist's cut scores were used to form a group median, which served as the group's cut score. The implementation of the method was modified in the following ways to be consistent with Bookmark-method implementation in previous NAEP achievement level-setting studies:

- Panelists were provided with actual test booklets to show examples of student performance on the assessment at the cut score and at the middle of each achievement level (i.e., Whole Booklet Feedback); and
- A spatially-representative display of items on a student achievement scale was given to panelists to accompany the ordered item books used to place the cut scores (i.e., Item Maps).

The Bookmark process was implemented using two independent replicate panels for each post-secondary area within each content area in an attempt to estimate the reliability of the replicate panels' cut scores. The item pools were divided into two comparable parts, with some overlap between the two parts, which were assigned to replicate panels. Panelists were selected and assigned in order to create comparable replicate panels, and, to the extent possible, panelists were deemed equivalent for each pair of replicate panels (Measured Progress and WestEd, 2012).

An Iterative Process

A JSS process consistent with the intended purpose of the Program for Preparedness Research and as prescribed in the Design Document was implemented in the four separate standard-setting sessions. The basic structure of the sessions consisted of the following:

- Training, wherein panelists were provided with information and instructions to enable them to provide informed judgments;
- Data collection, whereby panelists provided their judgments in three rating rounds; and
- Feedback, whereby panelists were provided with additional information based on the judgments they provided in the prior standard-setting rounds.

Other important elements of the process were the editing and reviewing the BPDs as well a process evaluation questionnaire after each major stage of the process. After the final cut scores were set, panelists filled out a questionnaire about their final recommendations and selected items illustrative of performance of students considered academically prepared. More detailed stages of the process in Figure 2 present the iterative nature of the process.

Figure 2: Basic Structure of the Judgmental Standard-Setting Process
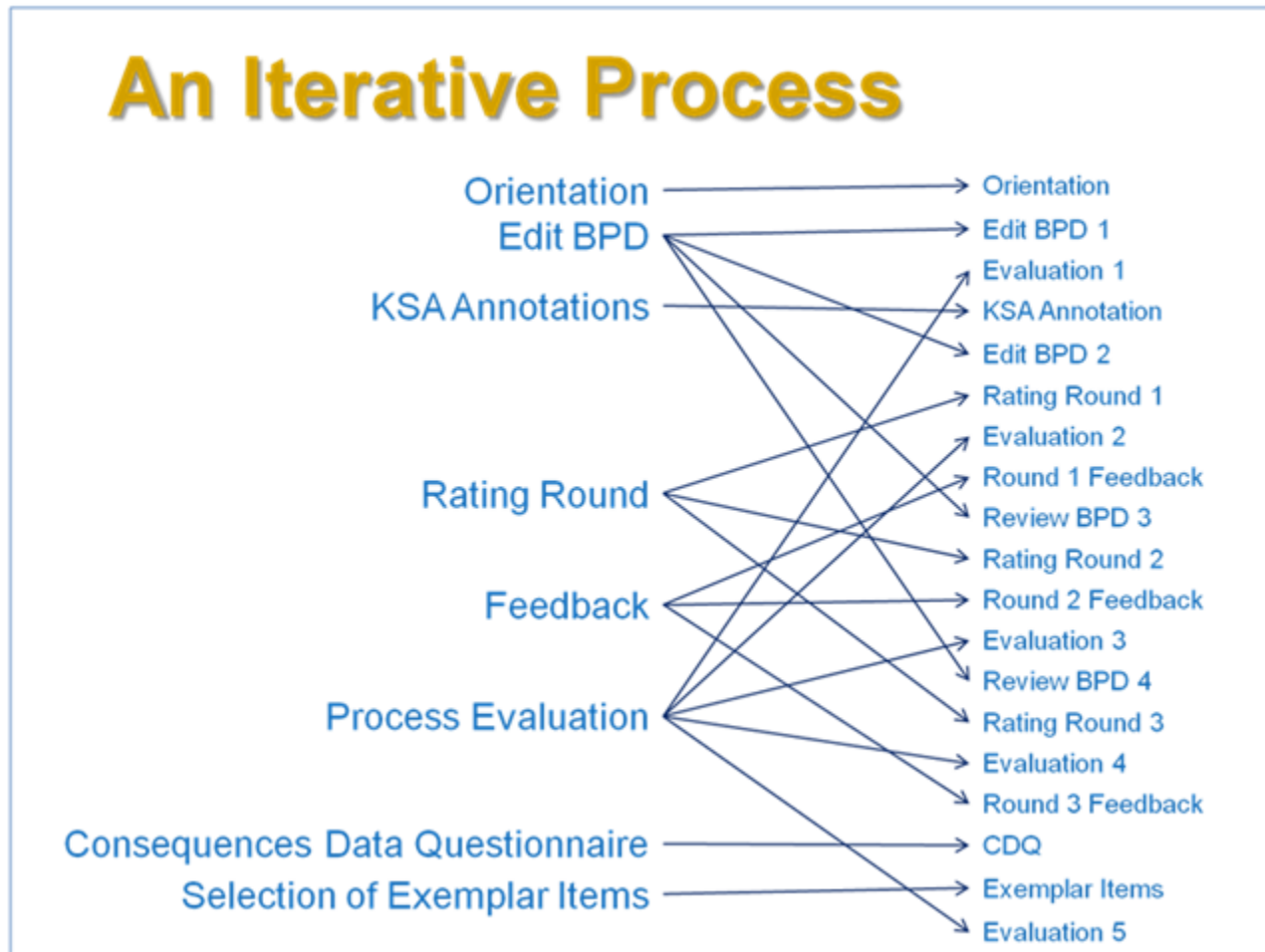
Instructions were provided to the panelists at three different levels of grouping:

- During general sessions;
- Within subject by post-secondary area sessions; and
- Within each replicate panel.

These levels are represented in Figure 3.

Figure 3: Sessions

In general sessions, instructions were given to all panelists. The purpose of general sessions was to provide the same information and instructions to panelists across all panels. The first general session occurred at the beginning of the first day. Other general sessions were held throughout the four-day sessions in order to introduce major parts of the process. All instructions in the general sessions were provided by the Chief of Standard Setting.

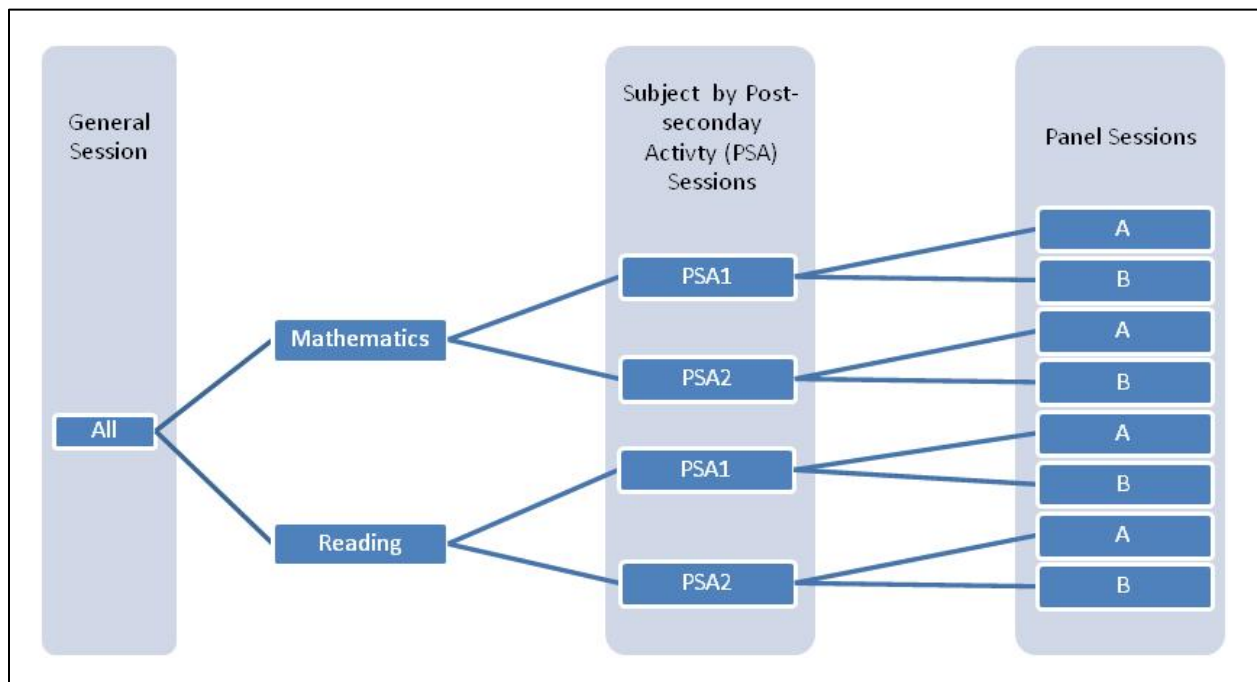The replicate panels were combined when the activity involved editing or gaining a common understanding of the BPDs, as stipulated in the study design. The content facilitators for the academic subject of that post-secondary area facilitated these sessions, and process facilitators noted the edits to the BPDs that were recommended as a result of the panelists' review and discussion.

Most of the panelists' tasks were performed at the replicate-panel level. The process facilitator assigned to each replicate panel provided instructions (e.g., for activities such as the rounds of ratings), provided feedback information between rounds, and led discussions about different feedback information.

Item Review of Knowledge, Skills, an   Abilities (KSAs)

Integral to the implementation of the Bookmarking method is the review of each item where the panelists identify the knowledge, skills, and abilities (KSAs) required to respond to each MC question correctly or to score a specific level on a CR item. The KSA review of items is an important step in the process. Understanding what is required to respond correctly to an item is necessary for a panelist to make a judgment on whether a student possessing characteristics described in the BPDs will be able to respond to an item correctly.

For the JSS implementation, the KSA review was done in two stages. First, panelists reviewed each CR item in its entirety. The rational for this was to provide panelists the opportunity to interact with each CR item as a whole and not with one score level at a time. The richness of a CR item might not be evident to panelists if they have to review each item separately for each partial- or full-credit response. For this activity, the Constructed-Response Ordered Item Book (CROIB) was constructed. Unlike the OIB, all the information about a constructed-response item was contained together on consecutive pages within the CROIB. The second stage was for each panelist to review each MC item in the OIB. When a panelist encountered a CR item during the review, his/her comment from the CROIB had to be transferred to the OIB. The description on how this was to be done according to the Design Document is included below:

> Because the panelists will need to record the knowledge and skills identified by going through the CROIB and then adding their notes from the CROIB to the OIB, panelists will find it helpful to use large yellow Post-it Notes to record their notes on each CR item. The notes are for the individual panelist and need only to be informative to the panelist. A separate description is needed for each score point for each CR item. When panelists are finished with an item, they place their notes in the Notes template. The Notes template is a stapled set of 11x17 pages with outlines for accommodating ten Post-its per page. Within each outline is an item handle and OIB page number identifying the Post-it that is to be placed there. (p.22)

Figure 4 below is how a template for a mathematics JSS panel would have looked if we were using paper CROIB and OIB. This sample would have been the first page of approximately six pages. The way that this would have been used is after a panelist annotates the KSAs for each CROIB item on a Post-it, he/she would place that Post-it on the appropriate spot on the template. During the OIB item review, the panelist would transfer the Post-it to the OIB when he/she comes across a CR item.

The Design Document further stated the following:

> The contractor may propose a computer-based method for recording the item descriptions and associating those with the items in the OIB. This note-taking step is labor and time intensive, and a computer-based method might speed the process of both taking notes and associating the notes with the items in the OIB. (p. 22)

Figure 4: Sample Template to Transfer CR KSAs from CROIB to OIB

Page 4 C24_1	Page 9 C1
Page 11 C3_1	Page 13 C2
Page 16 C23_1	Page 20 C21_1
Page 21 C3_2	Page 27 C20_1
Page 28 C4_1	Page 31 C26_1

Computer-Aided Bookmarking (CAB).

Measured Progress developed the Computer-Aided Bookmarking (CAB) software in response to the request by the Governing Board and consistent with the recommendation in the Design Document. In addition to KSA annotations, five other major elements of the JSS process were computerized. The following key activities were computerized with the development of CAB:

- KSA annotations
- Presentation of the Ordered Item Books
- Bookmark placements
- Provision of feedback
- Process evaluation responses
- Selection of exemplar items

CAB is referred to throughout this report, and its documentation can be found in Appendix A of the JSS Technical Report (Measured Progress & WestEd, 2012).

The JSS implementation of this project's Bookmark standard-setting methodology was wholly computerized using Computer-Aided Bookmarking software (CAB) to improve the efficiency of the process and to enhance the quality of the panelists' experience. Each panelist used a netbook computer to perform his or her tasks. All panelists entered their ratings and responses to evaluation questionnaires into the CAB. Panelists also used netbook computers to record the KSAs identified during the item review. The CAB was also used to present consequences data feedback, or impact data.

Although the OIB and CROIB are available on CAB, they were also made available on hardcopies. Panelists were given the option to view the items either on the computer screen or on paper. A handful of panelists initially expressed difficulty in viewing the items on a 10-inch screen. However, at each operational session, the paper copies were soon abandoned in favor of the computer.

Notable features of the CAB are described in the following paragraphs.

KSA Annotations

The CROIB and OIB are presented as item lists on the CAB as shown in Figure 5. Clicking on an item opens a window revealing the item. On the item window, clicking the "Your comment" toggle button opens and closes the window where the panelist can input the KSAs for that item as in Figure 6. Saving the comment and going back to the item list reveals a check mark beside the item that was just commented on.

For the OIB KSA review, a similar item list of all items – the CAB OIB – contains the list of all items. At the initial opening of the CAB OIB, there are already check marks beside the CR items that were previously commented on and those comments are accessible for reading or revising.

Figure 5: CAB CROIB

Figure 6: Item Comment Window for KSA Annotations

Bookmarking

For Bookmarking, the same item list for annotation appears, but this time there is a green bookmark that can be placed right before the item that the panelist deems too difficult for a student at the borderline to answer correctly (Figure 7). Going back to the CAB dashboard where panelists can access the functionalities for the different stages of the process, the cut score the she just set is indicated for that round (Figure 8).

Figure 7: Bookmarking

Figure 8: Dashboard

Rater Location Feedback

The Rater Location Feedback is a way of providing panelists interrater consistency feedback (Figure 9). On the screen are the cut scores set by the panelists, thus showing the distribution of the cut scores. The grey bar is the median cut score. Each panelist is able to tell which one is her personal cut score through a secret code that is only known to her, revealed by clicking her name on the upper right hand corner of the screen. The panelist is able to tell where her cut score is relative to the median as well as relative to the cut scores set by the other panelists.

Figure 9: Rater Location Feedback
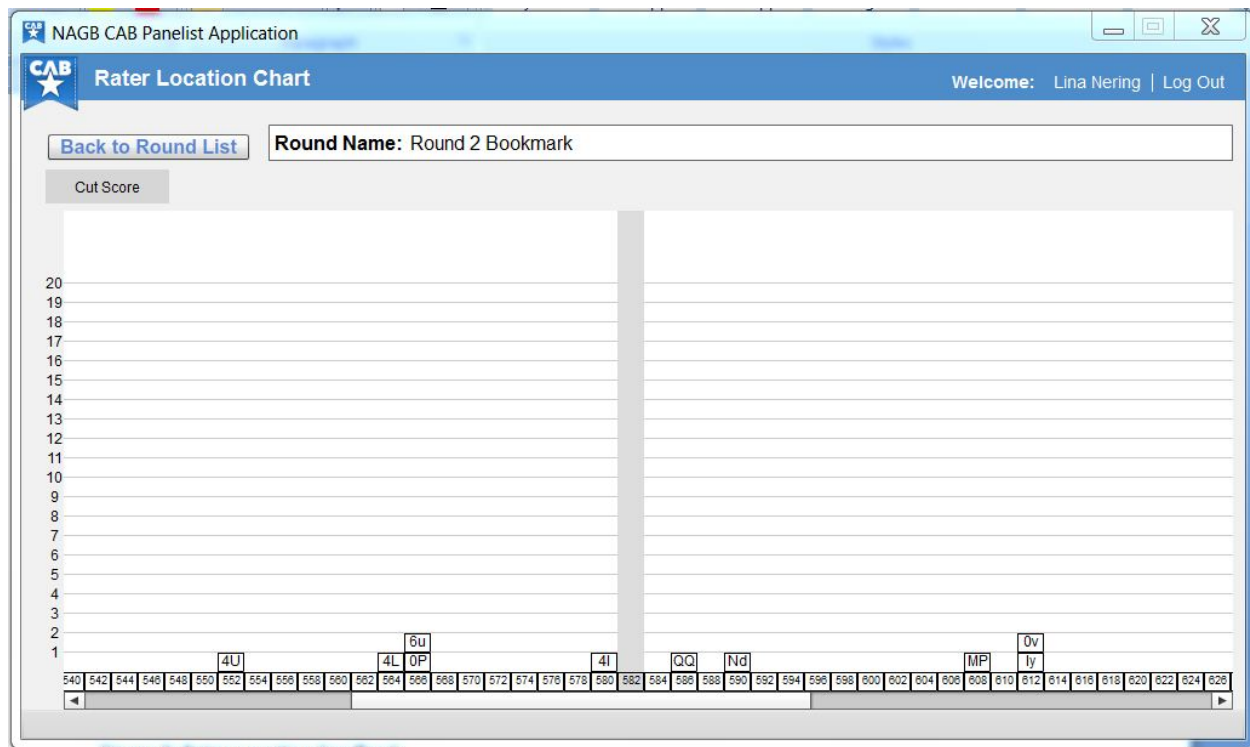
Consequences Data Feedback

As a special feature of the CAB software, the consequences data feedback was also provided using an interactive tool that displays the consequences data resulting from moving cut scores up or down the NAEP scale. The tool also calculates and presents the percentage of students at or above each scale score, as well as indicating on which side of the cut score each item falls as a panelist moves the placement of the cut score. A screenshot of the Consequences Data Feedback tool is in Figure 10. On this figure, a square was put around the slider which is the movable part, and in each circle are pieces of information that changes when the slider is placed in another point on the scale. Moving or sliding the cut score will cause the following four real-time changes on the screen:

1. The academic preparedness cut score.

2. The percentage of high school seniors in 2009 who would be considered academically prepared.

3. The sections of the pie chart that present the percentages of 2009 high school students who would be considered academically prepared (3a) and those who would not be considered academically prepared (3b) if the cut score is set where the slider is currently located.

4. The highlighted items on the vertical chart on the left hand side of the screen which are associated with academic preparedness.

The order of the items in this slider is the same as the order in the OIB. Each item handle is a clickable link to an image of the item as it appears in the OIB. Easy access to the items at or around the cut scores will help the panelists maintain the necessary connection between the cut scores and the BPDs.

Figure 10: Consequences Data Feedback
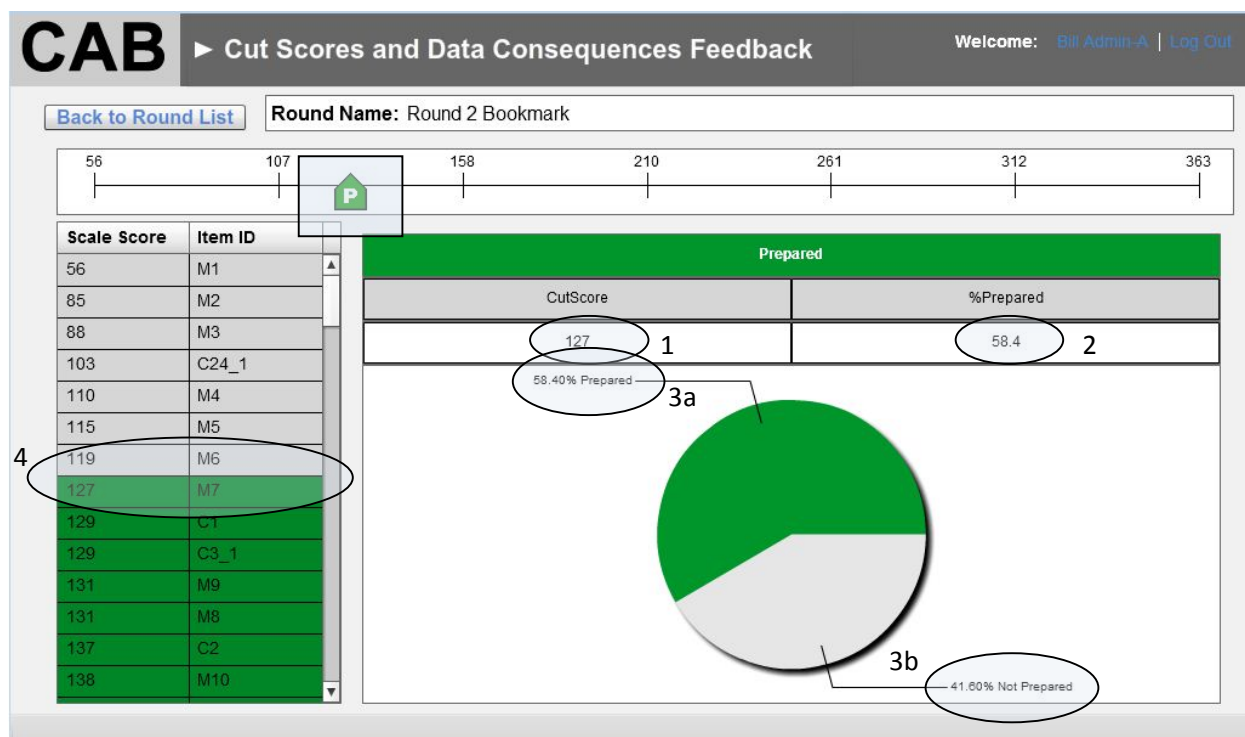
KSA Review Findings

The Governing Board requires for all of its standard setting to have a pilot study prior to operational meetings, and this was not different for the JSS for academic preparedness. A pilot study of the process was implemented using the procedures planned for the operational standard-setting sessions. The purpose of the pilot study was to determine whether modifications for training, instructions, materials, timing, or logistics were needed, as well as to provide an opportunity for the facilitators to practice the process before moving to the operational setting. Several aspects of the process were changed for the operational implementations based on feedback from panelists, facilitators, and observers.

One aspect of implementation that was modified over operational sessions was the amount of training, support, and information provided to the panelists for the KSA review. The ability of the CAB to collect panelists' KSAs provided rich information that led to improvements in the instructions provided. It also provided information on how well panelists were attuned to the language of the framework and the assessment. Information from the CAB regarding KSAs corroborates observations regarding some facilitators' understanding of the panelists' tasks and panelists' understanding of content. Collectively, this information prompted changes in the process of developing KSAs that culminated in a decision to provide item descriptions to

panelists in the final operational session for use in writing KSAs. Table 2 presents the different improvements on the process regarding the KSA review of the items. Details of the KSA review are described below for each standard-setting meeting. The process for the review as intended is described under the pilot study section, and details of the modification are in each subsection for each operational meeting. There were other process improvements as a result of the pilot study that were not directly related to the KSA review. They are not described in this paper, but interested readers are referred to the JSS Process Report (WestEd & Measured Progress, 2011).

Table 2. Guide to Changes in JSS Process Implementation Pertaining to KSA Review

Changes	Operational Session		
	1	2	3
Standardized Instructions			
Addition of a general session before every milestone in the process	✓	✓	✓
Provision of PowerPoint slides to process facilitators	✓	✓	✓
KSA Review Instructions			
Content facilitator models identification or development of KSAs using CR items only	✓	✓	✓
Content facilitator models identification or development of KSAs using MC and CR items		✓	✓
Anchor descriptions provided to facilitators to the lead the panel in a discussion to identify KSAs for CR items	✓	✓	✓
Anchor descriptions provided to panelists for KSA development and review			✓
Bookmarking Instructions Relative to Items Deemed Irrelevant	✓	✓	✓

Pilot Study KSA Review Process

Following the orientation training and discussions surrounding NAEP frameworks and BPDs, panelists began to familiarize themselves with the items in their item pools. Panelists were introduced to the CAB software. For each item, they were then instructed to think about the KSAs needed to correctly answer each item or (for CR items) to score at a specific score point on the rubric. The KSA review process involved multiple stages, including review of CR items alone and with MC items, review with the whole group, review within the table group, and independent review. Each stage was designed to help panelists gain a clearer understanding of what the assessment was measuring and the performance required of students.

Panelists reviewed both a CROIB, which contained only CR items, and an OIB, which contained all the items in the panelists' item rating pools. Both electronic and paper versions of these materials were made available. Facilitators reviewed a few items with the panelists so that panel groups could discuss the KSAs before continuing the KSA review task independently.

Ordered Item Book (OIB). Panelists were provided OIBs (both within CAB and in paper form) containing all of the items in the item pool to which their panel would be exposed during rating rounds. The items were ordered by difficulty from the easiest to the hardest. Item difficulty was based on grade 12 student performance on the 2009 NAEP. Scoring rubrics for CR items were included. CR items appeared multiple times, once for each credited score level. Each OIB included one page for each MC item and at least two pages for each score point of a CR item (i.e., one page for the item and one page for the scoring rubric).

Constructed-Response Ordered Item Book (CROIB). Panelists were also provided CROIBs, both within CAB and in a paper book, which contained all CR items and score points. The CR items were ordered by item difficulty of the full-credit response, from easiest to most difficult. In the CROIB, each item was included only once. Having panelists review the KSAs of CR items using the CROIB provided them the opportunity to interact with each CR item as a whole and not with one score level at a time. The following information was included in the CROIB:

- The page with the item, which included an information box with the item ID and the page numbers where the item's highest score point can be found in the OIB.

- The scoring rubric. (Note: KSAs were written for credited responses only.)

- An example of a student response at each score level, including incorrect (noncredited) responses.

Panelists used the CAB software for the first time during each group's review of CR items, during which panelists viewed and noted the KSAs required to receive a specific score point or higher. Exact instructions (with screenshots) given to the panelists are provided in the Facilitator Handbooks in Appendix K of the JSS Process Report (WestEd and Measured Progress, 2011). Panelists were also introduced to item maps for their use during the OIB KSA review.

Because there was not sufficient time for each panelist to review all items individually, panelists were assigned specific items to review. For mathematics, panelists in each panel (i.e., A and B) were assigned between 59 and 61 score points to review. For reading, panelists in each panel were assigned between 65 and 67 score points to review. Each item was included in the item list of at least two panelists in each table group; thus, each panelist had an opportunity to interact with each item during table group discussions.

Stages of KSA Review. The KSA review was done in stages to ease the panelists into the activity. The process facilitators modeled the item review task for a sample of about four items that illustrated the various types of scoring rubrics associated with the constructed response items. They began with an easy item and progressed to more difficult items. For each constructed response item, they asked panelists to identify and make comments on what students need to know and be able to do to get full credit on the item; then they asked panelists to identify and make comments on the knowledge and skills needed to earn successively lower scores on the item. Comments were entered into the CAB software. The panelists received training to use the CAB in a general session immediately before this session. For reviewing the rest of the CR

items, panelists independently reviewed the items identified with a star on the CAB CROIB item list.

For reviewing the OIB at the next stage, panelists used an Item Map which had been printed for them. The Item Map showed all NAEP items in the item pool for the replicate panels ordered from easiest to hardest. On this map, the difficulty of an item was mapped to a scale value. A student scoring at that scale value had a 67% chance of getting that item correct. The item map, therefore, showed "how much" more difficult one item is than another. It is color-coded to identify the domain (e.g. algebra for math or literary for reading). Panelists independently reviewed the knowledge and skills required by multiple-choice items in the item pool by entering comments for each item where they see a red star in CAB.

Finally, the items in the OIB were discussed by panelists in each table group. Again, items are considered sequentially, beginning with the easiest. Panelists shared their ideas about the knowledge and skills needed to get each item correct and edited or added to their comment notes as they saw fit.

KSAs Identified by Panelists. The primary purpose of having panelists input their KSA comments in CAB was to have their CR annotations automatically available to them in the OIB. Having all KSA comments in the database afforded us (standard setters/ process implementers) the ability to examine their comments. Because they were readily available, a quick visual inspection of their KSA comments could be done without much effort. Not so in a paper-based implementation of Bookmarking: panelists handwrite their notes on the physical OIB, and for the chief of standard setting to be able to examine their notes one has to either flip through each panelist's OIB or have the notes entered into a database.

A visual inspection of panelists' KSA annotations revealed three unintended and undesirable behaviors from the panelists.

- Some panelists' comments are mostly regarding whether items are relevant or irrelevant to their training program. Figure 11 is an example of a panelist's KSA annotations.

- Some panelists' comments on the items are not about knowledge, skills, and abilities required to answer the item correctly (if it's MC) or to provide a response that will receive a certain score of higher (if it's CR). Some of the comments are simply on the difficulty of the items (e.g., "this item is difficult"). Figure 12 is an example of a panelist's KSA annotation that is not pertinent to KSAs.

- Some panelists commented on the items relative to the BPDs. Figure 13 is an example of this type of item comment. It was determined that this is not a failure on the part of panelists to understand the task, but a failure on the part of some facilitators to understand the intent of this part of the process; thus, the wrong instruction was given to the panelists.

The above findings led us to implement process improvements for the operational sessions.

Figure 11: Example of a Panelist's Comments on Items that are mostly on Relevance

little relevance to auto

not relevant to auto

some relevance to auto

somewhat related to auto

somewhat related to automotive;O2 readings

requires critical thinking skills

not relevant to auto

somewhat related to auto;medium difficulty

somewhat related to auto;medium difficulty

question framework similar to ASE

not relevant to auto

requires some critical thinking

somewhat related to auto

related to auto

may be used in alignments

relevant question

somewhat related to auto; need to be able to read spreadsheets

not relevant to auto

not relevant to auto

don't seem relevant to auto; reubric seems fair

not relevant to auto

not relevant to auto

not relevant to auto

not relevant for auto

not relevant to auto

somewhat related

not related to auto

don't see relevance for auto; rubric seems fair

not relevant to auto

not related to auto

not needed for auto

not related to auto

relevant question'fair rubric

somewhat related to auto

not relevant to auto

important task to know

not relevant to auto

formula not needed, but interpretation of data important

not related to auto

relevant quesion

relevant to carpentry not auto

not relevant to auto

not relevant to auto

not relevant to auto

relevant to carpentry not auto

not relevant to auto

relevant to carpentry not auto

not really relevant for auto

Figure 12: Example of a Panelist's Comments that are not about Knowledge, Skills, or Abilities Required to Answer an Item Correctly

Most of our students would quickly lose interest in this story as they would not see any relevence for them. Therefore they would not make a good effort at responding to this question.

response above the level that I would expect of our students

I don't feel this is relevent for our students

OK for our students

Appropriate question for entry level

This level is appropriate for technical students at our level.

I don't feel this is relevent to technical students

Not relevent for our students

Students should be at least at this level.

Hopefully students would be at this level.

This would be an appropriate question for our students

Students should be at this minimum level

Good concept for our students to understand

No correct answer because there is no brochure

Ok for our students

Must understand the statement but not explain its relationship

Once again above the level that entry level students have

Not sure how this relates to the technical reading knowledge required of our students

Once agian this is an invalid response given the material presented. If a student responded this way to this material the answered response would be wrong.

Ok for our students

Ok for our students

Appropriate for our students

The responses for full comprehension contain statements that are not in the job opportunity reading material. There is no reference to completing in black ink.

Hopefully students would be at this level.

I do not feel this is relevent to our students reading ability.

Appropriate for our students

Students should be at this level.

This simply test the students on their knowledge of the definition of the word

It would be nice for the students to be at this level.

Our students should be able to obtain this response from the material

OK for our students

This is a good example for technical students. the article has relevent information and a good response shows the studeht's ability to read, understand, and pull relevent information out to help form a conclusion.

This is still above the level I would expect

would hope that students are at this level.

A good representative reading for technical students. Shows their ability to pull relevent information from the material.

This would be at the level of our students

I do not feel this is a good measure of technical students ability to locate, relate, and retrieve relevent material required in technical areas.

This is above what I would expect of entry level students

Not an appropriate article for our students

This requires a complete and thorough comprehension of the text

Would be nice if students were at this level but most would not be

Not a good reading article to capture our students ability to obtain relevent and useful technical information.

This is a good example of requiring a student to analyze a statement and draw conclusions.

Above the level I would expect for our students

Way above our level.

Way above what I would expect from entry level students

Figure 13: Premature Bookmarking

a minimally prepared student should be able to complete this question for full credit even if unable to do the math they should be able to figure it out and provide correct answer
a minimally prepared student should definitely be able to calculate the perimiter of a square
basic math skills necessary for auto students
basic probability can figure out with basic fractions no understanding of prob needed
demonstrate understanding of the relationship between congruent angles and parallelism, alternate exterior angles should be able to give a partial explanation would not expect full understanding
level of geometry not directly applicable
might not be able to answer requires higher level understanding of geometric relationships
minimally prepare student should answer correctly should be able to calculate percentages perform multi-step calculations correctly order steps
minimally prepared should be able to answer chart may be confusing rather than helpful leading to quick response rather than thoughtful
minimally prepared student should be able to answer
minimally prepared student should be able to answer partially student should be able to sketch a vector would not expect a northerly component
minimally prepared student would not need to do this demonstrate understanding of sin calculate square root
minimally prepared student would not need to know this
minimally prepared will not need to know this basic question but terminology and concepts will be unfamiliar
minimally prepared would not be able to answer this dont know if they need to more difficult in that many layers of info are included lots of data to track and compare
minimally prepared would not be able to answer this no graph increases difficulty removing the ability to estimate answer
must absolutely recognize that information is missing
no clue not applicable
not applicable to automotive more dificult because there is no graphical representation
not directly applicable but should be able to answer

This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

The annotations presented above in Figures 11—13 are just examples of comments from three panelists – each one a proliferation of a specific undesired way that panelists commented on items rather than the actual knowledge, skills and abilities required to respond to OIB items correctly. Observations by designated process observers, in this case two Governing Board staff members – one of which is the Contracting Officer's Representative – as well as information shared during daily facilitator debriefings revealed two main sources of the discrepancies between what panelists were expected comment on and their actual comments, as well as the understanding of some facilitators regarding the task for the panelists at that specific stage of the process. In either case, the issue is on instructions – the instructions received by panelists, as well as the instruction provided by facilitators based on their understanding of the task. Thus, process improvements coming from the pilot study were more on providing instructions and less on process logistics, per se.

Operational Session 1

Beginning in the first operational session, a general session was included prior to every major stage of the process. (In the pilot, the general sessions are the orientation, and whenever there is a new functionality of CAB that needs to be introduced to the panelists.) Holding multiple general sessions in a standard setting meeting is consistent with the NAEP achievement levels setting tradition. However, this was not prescribed in the Design Document. Holding general sessions speaks to the standardization of the process. As it relates to the KSA review, only some facilitators instructed their panelists to gauge whether each item will be responded to correctly by a student performing at the borderline of academic preparedness. Effectively, the facilitator for one group instructed the panelists to decide where they will set their individual cut score for round 1 as they review the items in the OIB. This anomaly was clarified with the facilitators during the debriefing meetings during the pilot study. It was also reinforced during the facilitator meeting the day before the first operational session. To ensure that panelists are less affected by variation of instructions from different facilitators, and also to make them less susceptible to a misunderstanding on the side of the facilitator, all instructions are initially provided during the general session. Thus, panelists hear the same instructions during the general sessions. When they get into their respective rooms, the instructions are repeated by the process facilitators. Additionally, PowerPoint slides were prepared for and provided to each facilitator for them to use in providing instructions to their respective panels. As expected, panelists felt that instructions were repetitive but they understood the necessity. Giving the panelists in the operational sessions a heads-up that they would feel that instructions were repetitive significantly reduced the grumbling.

Another enhancement to the KSA review of items was instructing panelists on what not to do when identifying the KSAs required to respond to an MC item correctly or to receive a certain score level or higher. Instructions to panelists included the following:

- Do NOT write your opinion of the item or rubric.

- Do NOT write about how the KSA relates to the BPD.

- Do NOT rate the items as "hard" or "easy."

- Do NOT rate the items as "needed/relevant" or "not needed/irrelevant" for your particular post-secondary area.

Furthermore, for items that measure KSAs that are not relevant to the requirements of the panelists' programs or courses, panelists were instructed to make note of that item but not to rate all items in terms of relevance.

In addition to more standardization measures, the KSA review of items was enhanced by adding an additional session to the agenda where the content facilitator introduced the panelists to the KSA review activity while in their subject groups. Using items from the released blocks of NAEP items, the content facilitators led the panelists in discussion of what students should know and be able to do in order to receive credit for their responses to CR items, modeling the thought process for making comments regarding the KSAs required for each item. The content facilitators were provided with examples of KSA descriptions, and they encouraged panelists to suggest ways to describe the KSAs needed for each level of response. The anchor descriptions were prepared for NAGB for another project.

Because the modeling of the thought process for KSA review preceded the part of the process where panelists review KSAs for CR items, only CR items were used in this discussion for the first operational meeting. For this activity, the content facilitators were allowed to select any constructed response item from the form of the NAEP that the panelists took as part of their orientation. The descriptions for these items that were provided to the content facilitators are in Figures 14 and 15. For this part of the process, the KSAs identified as a group for the items that were discussed were not enter in CAB for two reasons:

1. The items were not part of the rating pool and were not in the OIB. The Design Document stipulated that released items would not be used for standard setting, per se.

2. This discussion was a combined-panel activity; thus, only half of the panelists would have had access to their computers.

Figure 14: Item Descriptors for Mathematics Released Items

Multiply a whole number by a decimal
Determine length of side of 30-60-90 triangle given one side
Identify equation of line of best fit for scatterplot
Recognize that extrapolation beyond the data is dangerous
Recognize equivalent expressions using commutativity and distributivity
Determine whether a relation is a function given a table of values
Interpret slope, intersection, x-intercept, verbally and numerically, for a pair of distance time graphs with negative slopes (2 out of 3, anyway)
Recognize an interval estimate that satisfies a complex verbal description
Simplify difference of rational expressions
Validate statement about a figure using coordinate geometry and properties of parallelograms
Solve a system of linear equations for both values (integer solutions)
Find ranges for multiple pairs of data
Solve a word problem involving area
Evaluate a function that requires interpretation of a verbal description to find the value of the independent variable (exponential)
Make an inference about a quadrilateral inscribed in a circle
Estimate solution to a multi-step word problem; rate, proportional reasoning
Build an algebraic expression from a verbal description
Solve for parameter given equation and initial conditions
Solve for and use parameter given equation and initial conditions
Compute length of a side of a triangle using right triangle trigonometry
Compute compound probability of independent events in context
Perform reflection of a line segment in the plane (transformation in the plane)
Perform sequential transformations in the plane

Figure 15: Item Descriptors for Reading Released items

Recognize information explicitly stated in a document
Locate information stated directly in the text
Provide partial information about textually explicit details in a document
Provide complete information about textually explicitly details in a document
Integrate information in a paragraph to recognize specific terms of an agreement in a document
Recognize the meaning of a word as used in the context of a document
Identify information required by a document and make a simple inference to provide a limited explanation for the reason this information is required
Identify information required by a document and make a simple inference to provide an explanation for the reason this information is required
Locate and provide relevant information from a section in a document to support a generalization
Recognize the meaning of a word as used in the context of a document
Provide information about how a document is organized
Analyze the effectiveness of an organizational feature of a document
Integrate information across the whole story to recognize a description of the plot.
Recognize the meaning of a word as used in the context of a story
Integrate local information in a story to recognize a character motivation
Integrate information from a whole story to provide an explanation of a character's actions
Locate and provide textually explicit character motivation
Provide information about a character's thoughts or behaviors
Provide information about a character's thoughts or behaviors and explain the reasons for them
Provide examples from the story that illustrate an author's portrayal of emotions
Analyze story to provide an explanation of how an author uses elements of two different genres in a story
Recognize the meaning of a word as used in the context of a story
Provide text-based description about character in a story based on given character trait
Analyze story to provide text-based evidence to support a given description of character
Integrate details from a story to make a text-based generalization
Synthesize information in a story to recognize a generalization about the author's craft

Similar to the pilot study, the process facilitator led a discussion to identify the KSAs for four CR items – two polytomous and two dichotomous items. It was clarified with them that they were to identify the KSAs for each level of partial or full credit for each item – something that some process facilitators misunderstood during the pilot study. Another improvement from the pilot study was that the items were pre-selected for them and they were provided anchor descriptions that they may use for the discussions. Process facilitators were told that their having the descriptions was to help them in getting the discussion started and not to just tell the panelists what the KSAs were for the selected items. The panelists were expected to participate in the thought process for identifying the KSAs for those items. Participating in the group discussions was expected to give them an experience for identifying KSAs that will help them when they independently do the activity for the rest of the items in the CROIB and then the OIB. There was

no change in the OIB KSA review. The panelists independently reviewed OIB items on their respective item lists. This was followed by a table group discussion of the KSAs that they identified so that each panelist would have had a KSA-specific interaction with each item in their OIB.

Examining KSA entries from the panelists revealed that there was significant improvement relative to the pilot study. Although there were still comments regarding "irrelevant" items, they were not the predominant annotations for any of the panelists. There is no definite evidence that panelists were prematurely placing their bookmark either. That is, none of the comments were similar to the ones in Figure 13 above. However, there are some panelists whose KSA comments are followed by a "YES," a "NOPE," or a "Maybe" (see Figure 16). It is not clear whether it was their way of indicating that that particular item was "too difficult" or "not too difficult" for a student at the borderline of preparedness. It is possible that we have not totally eradicated premature bookmarking during operational 1. Other than the "yes," "no," and "maybe" notes on the items which were made by a few panelists, the KSA comments provided by the panelist in Figure 16 are typical for automotive technician panelists in reading.

Integrate, interpret text.

Locate and recall information

Locate or recall setting.

State an opinion, did not need to read the story.

Re-read and review

Re-read to find answer.

State an opinion, did not need to read the story.

State an opinion, did not need to read the story.

Compare or connect ideas, different perspectives and situations

State an opinion, did not need to read the story.

Explain or define creativity.

Re-read to find answer

State an opinion, did not need to read the story.

Fact finding, not challenging. YES

The student would have to read and absorb both passages to get to this level. Recall, Interpret and visualize details. Maybe

The student would have to visualize, follow along and recall and have a better than average vocabulary. Maybe

The student would have to calculate the odds of seeing something amazing and realize they are very low and interpret that from the author's passage. Maybe

The student would have interpret the characters feelings and read the entire passage to get this level. YES

Draw a conclusion based on support information. YES

The student would have to recall a setting(Dream), compare analysis and connect ideas, describe a life setting, find information and draw an opinion or conclusion. Yes

Locate information and draw an opinion or conclusion. Yes

The students would have to read the whole passage, recognize some detail and have some comprehension. YES

The student would have to recall and somewhat understand to get this level. YES

The student will have to visualize and have a better than average vocabulary. Take the time to think through this poem. Maybe

The student could answer this question without reading the passage and have the understanding of distress. Interpret the character's mood. YES

Explain and compare perspectives. Maybe

The student could answer this question without reading the passage and have the understanding of decipher. locate and recall. YES

The student will have to interpret and understand the character's views and restate them. Maybe

The student would have to read and comprehrend the passage. Interpret the characters view and ideas at a basic level. Expressing some if any detail. YES

The student could read the passage with little understanding and just recall some detail. YES

The student would have to read and comprehrend the passage. Interpret the characters view and ideas. Expressing detail. Maybe

Determining the character's perspective. Maybe

The students have to skim through the passage but not fully understand to get this level. YES

Determining the meaning of shoddily by context. Maybe

"The student at this level will explain and provide an

opinion and explain. NOPE"

The example in Figure 17, although atypical, is illustrative of comments written by auto tech panelists for mathematics. This example provides a mixture of what we would consider KSAs necessary to respond correctly to items (e.g., "understand slope and [intercept]"), some comments that might address what it would take to receive credit (e.g., "knowing [all parts] and [give] the right [formula]"), as well as some comments that one might consider extraneous (e.g., "mary was the way"). One thing is clear from a visual examination of the KSA annotations from panelists: everyone followed the directions somehow. Panelists were answering the question, "What does it take to receive credit on an item?" (e.g., "knowing geometry" or "knowing algebra" in Figure 17) although they are not necessarily answering the question, "What knowledge, skills, and abilities does one need to demonstrate to receive credit for an item?" The example in Figure 18 clearly shows this. Several times, the panelists for Figure 18 noted "Correct equation and/or correct answer has to be given."

Our take-home message from the snapshots in Figures 16—18 is that the panelists had an understanding of what they were supposed to do, but they did not really understand what we mean by "knowledge, skills, and abilities." Another indication of this assertion is in Figure 19, where a panelist entered mostly "Locate and recall textual information correctly." Incidentally, "Locate and recall" is one of the three cognitive targets specified in Exhibit 8 in the Reading Framework for the 2011 NAEP (NAGB, 2010). According to the Framework, "[t]he term *cognitive targets* is used to refer to the mental processes or kinds of thinking that underlie reading comprehension; the cognitive targets serve to guide the test development process in the items writers "target" these processes or kinds of thinking as they write items." Some panelists did, indeed, enter text that was from a document that was distributed to them during their training on the framework and discussion of the BPDs: an outlined organization of the framework (Figure 10). A couple of panelists went as far as including the outline number (e.g., "2.1c Determine unstated assumptions"). Though these entries are not necessarily incorrect, it is an indication of the panelists' struggle on what we are asking them to do. It is a very difficult task for them to come up with the item KSAs on their own. As a matter of procedural validity, it is hard to imagine how a panel can set a reasonable and acceptable cut score based on a series of steps when one step is not performed according to intended procedures.

An important observation from operational 2 is that if you provide panelists source materials for the KSAs they will use them. Additionally, a conjecture based on the example in Figure 17 is that some panelists may not be able to identify or develop the right KSAs on their own, as exemplified by the first line of most item comments (e.g., "Correct equation and correct answer needs to be given), but during the table group discussion they are able to identify the KSAs with the help of their colleagues (e.g., "Knowledge of statistics and using data to determine Mean, Mode, and Range"). Thus, having small group discussions is very helpful for panelists when identifying KSAs.

Figure 17: Mathematics Example 1 from Operational 1

knowing that point on line and justify you answer bar understand the possible of real number

able to do part but not explain

able to do algebraand justify the answer

geo and formula

fail the answer the all parts

algebr one right answer

able to answer rith and not complete the grafh to showw answer

under standing that x can only be one point

been able to plot a graft and look a data for the answer

geometry only one right answeer

only one answer

wrong answer

need to know basic algeriba sub 6 for x ad sub

one right geo

understand slope and intersep

one answer algebra

mary was the way

knowing allparts and given the right formal

goe and knowing the susction of proof

number only one right answer

need to know or understand a ex sheet

knowingthe formal but not proff it all the way

answer part right and not sure of the rest justify

knowing geometry

algebra and grafts

did not circle

knowing algeba only one answer

able to polot graft and not sure of the answer number

knowing abgebra

only one right answer knowing now to use a ruler

knwing how to graft only one right answer

able complete the both part

number properties

knowing sq. mush be vol

number properties one answer

kwoning algebra formal

Figure 18: Mathematics Example 2 from Operational 1*

Student needs to understand what realumbers are and the inclusive and exclusive use of or". Student must know how to graph them on a number line"
Drew the correct traingle but didn't label it
Student must understand what a Function is and how to graph it.
Only one response is correct. Knowledge of geometric formulas involving the use of pie.
correct values given but didn't draw the triangle
Only one correct response is possible Knowledge of basic algebra and ability to perfom factoring and soilving for x
student failed to answer the whole question
Student must understand what real numbers are understand inclusive and exclusive use of or""
Correct formula and answer to be given Knoledge of graphs and ability to use them to predict probabilities.
Corr3ect equation and answer to be given Ability to read a map using scales and formulate an equation to come up with the right compuitation
Correct equation and answer has to be given Ability to read a map using scales.
partially correct response
Student must be able to add, subtract, multiply and understand exponets.
Correct formula and answer Knowledge of geometric formulas
Student understands slope/intercept values.
only one correct response is possible. Knowledge and ability to perform operationws involving exponets
Correct equation and/or correct answer needs to be given Knowledge of statistics and using data to determine Mean, Mode and Range
Student obviously is a whiz at geometry and understand how to determine line segments using square root formulas. Student understand how to determine area and perimeter. Student must get all four parts to the question correct
Student needs to have a firm grasp of the Pythagorean Theorem and have the mathamatical ability to perform geometric calculations.
Correct equation and/or answer given Ability to dicsern dievistion between numbers in a set and figure averages

*This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

*Figure 19: Reading Example 2 from Operational 1**

Locate and recall setting.

Compare, connect ideas, perspectives, problems and solutions

State an opinion, did not need to read the story.

"Simple verbage knowledge required only.

+"

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

need many details with coupious skills of concepts in text."

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Advanced analytical skilles"

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Advanced technical skills."

Unnecessary complex vocublary memorization skills.

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Learner should be able to locate and answer this question."

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Most should be at this level."

"Locate and recall texual information correctly.

identify and interpert texual information partially.

Show beginning understanding of analayzing and critiqing information."

"Locate and recall textual information correctly.

"Locate and recall textual information.

Analyze and critique texual information.

Interpret texual information."

"Locate and recall key text points.

Decide which is more important key point.

Compare and contrast textual key concepts.

Analyze, evaluate and critique textual information.

Advanced concepts"

"Locate/recall texual information.

Evaluate/ analyze and describe textual information.

Intrepert text."

"Locate and recall specific text point.

indentify/i nterpert specific text point meaning."

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Should be able to understand meaning in context."

"locate and recall specific text point correctly.

Interpert, analyze/critique specific text point correctly.

deciper true meaning of text.

Learner should have partial skill and during classes hone skill."

"locate and recall two specific text points correctly.

Interpert, analyze two specific text points correctly.

Compare and contrast textual point mostly correctly.

Should be able to partially complete here."

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Correctly used supplied text. inference is more advanced skill set we teach."

"Locate and recall textual information correctly.

Interpert, analyze/critique textual iformation correctly based on supplied information.

Figure 20: 2009 NAEP Grade 12 Reading Framework: Organization for Alignment Studies

Level	Description	DOK
1	Locate/Recall: Locate or recall textually explicit information within and across texts, which may involve making simple inferences as needed for literal comprehension	
1.1	Locate or recall textually explicit information and make simple inferences within and across *both literary and informational texts*	
1.1.a	Locate or recall specific information such as definitions, facts, and supporting details in text or graphics	
1.2	Locate or recall textually explicit information and make simple inferences within and across *literary texts*	
1.2.a	Locate or recall character traits	
1.2.b	Locate or recall sequence of events or actions	
1.2.c	Locate or recall setting	
1.2.d	Locate or recall figurative language	
1.2.e	Locate or recall organizing structures of literary texts, such as verse or stanza in poetry or description, chronology, comparison, etc. in literary non-fiction	
1.3	Locate or recall textually explicit information and make simple inferences within and across *informational texts*	
1.3.a	Locate or recall the topic sentence or main idea	
1.3.b	Locate or recall the author's purpose	
1.3.c	Locate or recall causal relations	
1.3.d	Locate or recall organizing structures of texts, such as comparison/contrast, problem/solution, enumeration, etc.	
2	Integrate/Interpret: Make complex inferences within and across texts	
2.1	Integrate/Interpret: Make complex inferences within and across *both literary and informational texts*	
2.1.a	Describe problem and solution, or cause and effect	
2.1.b	Compare or connect ideas, perspectives, problems, or situations	
2.1.c	Determine unstated assumptions in an argument	
2.1.d	Describe or analyze how an author uses literary devices or text features to convey meaning	
2.1.e	Describe or analyze how an author uses organizing structures to convey meaning	
2.1.f	Describe or analyze author's purpose	
2.2	Integrate/Interpret: Make complex inferences within and across texts *literary texts*	
2.2.a	Interpret mood, tone, or voice	
2.2.b	Integrate ideas to determine theme	
2.2.c	Interpret a character's conflicts, motivations, and decisions	

Level	Description	DOK
2.2.d	Examine relations between or among theme, setting, plot, or characters	
2.2.e	Explain how rhythm, rhyme, sound, or form in poetry contribute to meaning	
2.3	Integrate/Interpret: Make complex inferences within and across texts *informational texts*	
2.3.a	Summarize major ideas	
2.3.b	Draw conclusions and provide supporting information	
2.3.c	Find evidence in support of an argument	
2.3.d	Distinguish facts from opinions	
2.3.e	Determine the importance of information within and across texts	
2.4	Integrate/Interpret: Apply understanding of vocabulary to comprehension of *both literary and informational texts*	
2.4.a	Determine word meaning as used in context	
3	Critique/Evaluate: Consider text(s) critically	
3.1	Critique/Evaluate: Consider *both literary and informational texts* critically	
3.1.a	Judge the author's craft and technique	
3.1.b	Analyze, critique, or evaluate the author's perspective or point of view	
3.1.c	Take different perspectives in relation to a text	
3.2	Critique/Evaluate: Consider *literary text* critically	
3.2.a	Evaluate the role of literary devices in conveying meaning	
3.2.b	Determine the degree to which literary devices enhance a literary work	
3.2.c	Evaluate a character's conflict, motivations, and decisions	
3.3	Critique/Evaluate: Consider *informational text* critically	
3.3.a	Evaluate the way the author selects language to influence readers	
3.3.b	Evaluate the strength and quality of evidence used by the author to support his or her position	
3.3.c	Determine the quality of counterarguments within and across texts	
3.3.d	Judge the coherence or logic of an argument	

Operational Session 2

Beginning with the second operational meeting, the content facilitator modeled the identification or development of KSAs using both MC and CR items. It was determined from operational 1 that modeling the thought process for identifying KSAs using CR from the released blocks only was not enough. Content facilitators then led the panelists in identifying the KSAs for all the items in the released blocks. Recall that panelists took a form of the assessment that contains the two released blocks, and that the items in the released block were not included in the OIB.

Following this training, panelists reviewed the KSAs for four assigned CR items in their panel groups with the discussion led by the content facilitator. During this activity, process facilitators had access to anchor descriptions for all partial- and full-credit responses for each item, and they were instructed to use these descriptions to help guide panelists in meaningful discussions. KSAs identified as a group were entered in the CAB. Panelists then proceeded to identify KSAs for the rest of the CR items independently, followed by an independent review of items in the OIB, then the table group discussions so that each panelist has had a chance to interact with each item in the OIB. Note that in this operational session, the panelists were instructors in training programs for Licensed Practical Nursing (LPN) and Pharmacy Technicians. They are not necessarily teaching mathematics or reading.

One thing that was very noticeable in the KSAs written by operational 2 panelists is that their notes were more engaged. Figures 21 and 22 are typical KSA snapshots for mathematics and reading respectively. Note that for each item, multiple KSAs were identified by the panelist. Figures 23 and 24 might be considered not as good in the sense that the KSAs noted were simplistic and not very descriptive of what students should know and be able to do to receive credit for the item. For example, "numbers[,] counting[,] understand question" from Figure 23 is not as informative as "Solving algebraic equations[,] understanding number [properties] and order of [operations,] addition [subtraction multiplication] division" for Figure 21. Similarly for reading, "Locate information[,] Interpret information[,] understand vocabulary" from Figure 24 is not as informative as "Locate information[,] Use context clues for word meaning[,] Make inferences regarding character traits" from Figure 22. In light of that, we have to recognize that the KSA annotations are supposed to be notes-to-self and, giving the panelists the benefit of the doubt, short cut descriptions might be meaningful and informative enough to the panelists for when they determine whether a student performing at the borderline will be able to respond to an item correctly.

The four examples in Figures 21—24 are, in a lot of ways, typical for operational 2. First, more often than not, there are multiple KSAs identified per item. Second, there seems to be more evidence of cutting and pasting of KSAs than was observed in the pilot study and operational 1. Third, although the author can only speak on mathematics KSAs, there is evidence that even though panelists entered mathematical concepts that students should know and be able to do at the borderline of preparedness, the panelists might not necessarily understand the concept. Examples of these are:

- "computing the volume of a rectangle"
- "know that the area of a circle is 360"
- "use path theory" or "pyga theorem"

- "compute mean and medium"
- "square routes"

Additionally, one of the process facilitators for a mathematics panel observed panelists in his group who were entering the wrong KSAs. This is consistent with panelists' difficulty with the academic aspect of NAEP reported in Loomis (2012).

Even though the panelists had been trained not to base all their comments on the relevance of items to their training program during the KSA review of OIB, they made it a point to do just that. As a matter of fact, we have been informed of whole subcontent areas that are irrelevant to specific training programs. Entries in Figure 21 are from a Pharmacy Technician panelist. Note the indication that Geometry is not relevant to their training program. This sentiment was expressed in other ways during the standard-setting meeting. The LPN panelists similarly expressed that Measurement is not relevant to their training program.

Figure 21: Typical Mathematics Example from Operational 2*

Measure Ratio and porportions Understanding maps and scaling solve word problems using rates addition, subtraction, mulpiplication division converting units
Understading greater than and less than Understanding and/or Real numbers/number line Substituting variables Express data using words, symbols, and graphs
Solving algebraic equations understanding number rproperties and order of opoerations addition subraction multiplicatioin division
Use ruler and protractor Knowledge of tiangles and angles ratios and porportions multiplicationa and division addition and subtratction Knowledge of laws of sine
Understading greater than and less than Understanding and/or Real numbers/number line Substituting variables Express data using words, symbols, and graphs
Use ruler and protractor Knowledge of tiangles and angles ratios and porportions multiplicationa and division addition and subtratction Knowledge of laws of sine
understanding number rproperties and order of opoerations addition subraction multiplicatioin division
Understandig geometric figures and measurement
Geometry Understanding and interperating graphs
Big G
Nterpreting map, scaling and measurement
Algebraic formula dealing with inverse function
Coordinate plane, Algebra
bigg G

This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

Figure 22. Typical Reading Example from Operational 2*

Interpret mood Integrate ideas to theme Summarize
Infer feeling, character's conflicts Locate event Determine unstated assumption
Can provide a story theme without clear support from passage. Can identify topic from story but not clear theme.
Provides and substantiates opinion regarding author's ability and success to build suspense Explains author's writing style
Locate sequence of events in literary text Infer action of specific character Locate specific information
Student can explain what the essay shows aboutthe author as a person and support the answer with relevant details from both the essay and the biographical sketch. Student offers generalizations about White's character that could be inferred from the essay, and uses details from both passages that support the generalization. Must be able to generalize and summarize. Must understand vocabulary.
Student demonstrates understanding of how the city setting contributes to the effectiveness of the essay. Can explain that the events in the essay are more significant and unusual because they are not expected in a city environment.
Can provide adequate interpretation of statement with supporting details.
Student can provide an appropriate feeling inferred from the essay. May make specific reference to White's experience at the zoo or may provide a more general statement that reflects White's experience.
Locate information Use context clues for word meaning Make inferences regarding character traits
Assess text critically Infer meaning and draw conclusions
Makes incorrect evaluation Unclear analysis Incorrect assumptions Makes incomplete argument
Analyze author's purpose Determine main idea Summarize information
Determine assumptions Use context clues to find word meaning
Students should be able to identify sections of text: beginning, middle and end. Students can create a persuasive essay using support from the textual elements. Students are able to evaluate comparative assessments presented in informational text.

This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

*Figure 23: Another Mathematics Example for Operational 2**

numbers counting understand question	understand graphs placing coordinates on the graph
number properties addition most common denomatator	understanding scatter graphs interpreting data location of midpoint to form a line
algebra what operation	algebra formulate equation
interpred data mean, medium, mode	number properties what opperations
know the metric measurements reading scales formulate the operation	probablity read a graph
geometry pie calculate area of circle metric measurments foumulate operations	calculate area of triangle formulate operation geometry multistep pythagorean theorem
calculate area of triangel geometry formulate operations	pythaagorean theorem geometry relationships prooving equations calculate area of triangle
formua for interest concept of interest interpret graph without values	working linear equations with multiple ...
understand graphs placing the points on the graph no need to show solution (equation)	math operations in order use exponets
functions of planes verbal descrecion of figure and interpret	setting up formulas (comp. spreadsheet) formulating operations
computer spreadsheets formulas in cells formulas functions for the calculations	relationships calculate area of triangel
understand graphs formulate operation order of operations function	foumulating operations graphs substitute values in an equation
measurement calculate perimeter	missing information - area of towns population v. density

This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

Figure 24: Another Reading Example for Operational 2*

Interpret analyze explain	interpret locate and recall suportive documentation
analyze interpret	vocabulary
interpret	locate and recall
interpret	identify main topic
analzye and explain from both documents supporting documentation locate and recall	locate and recall supporting documentation from text locate main idea
interpret and expalin setting	recognizes one or the other
interpret explain draw a conclusion	analyze compare and contrast interpret explain draw conclusions
locate and recall interpret tone	comparing and contrasting analzye and support
locate and recall	locate and recall
locate and recall	vocabulary interpret
vocabulary interpret	locate and recall interpret
locate and recall	locate and recall explain
unsupportive judgement locate and recall	interpet what he author says describe the main idea paraphase
draw conclusions and interpet the information	interpet local and recall main topic
locate and recall	vocabulary interpret
score 4- determination,compare and contrast, locate and describe	draw a conclusion
local and recall	interpret and describe locate and recall
interpret the tone and voice draw conclusion summarize	interpret theme draw conclusions

*This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

Operational Session 3

An enhancement to the KSA review that was implemented only for the third operational JSS session was the provision of item descriptions to the panelists. Observations from the first three JSS meetings indicated that job training panelists lack the academic background to describe KSAs for grade 12 NAEP items in mathematics and reading. Identifying KSAs is a necessary part of the Bookmarking method and it is time-consuming and challenging even for panelists with the right academic background. Providing panelists a descriptor for each item was a sound modification to the process that seems to have yielded desirable results.

A description was provided for nearly every item in the OIBs. The descriptions were developed by the Frameworks committees at the request of the Governing Board for another project. With this modification was another slight change to the process: unlike in the earlier workshops, the initial KSA review was not performed individually, but with a partner. For the initial review using the item descriptions, the following instructions were given to the panelists:

> A description for each item will be provided to you to help you with your task. You may use the provided description as the KSAs, you may paraphrase them, or you may decide to develop your own KSAs. You and your partner should share your thoughts and suggestions regarding the KSAs for each item, but you do not have to agree on the KSA you record for each item. (p. 18 of the Facilitator Handbook, Appendix K of the Process Report)

A visual examination of KSAs for operational 3 revealed more meaningful and informative KSAs entered by panelists in the OIB. The expectation was that panelists would either have copied the item descriptors as KSAs or at least modified them. It was surprising that panelists hardly copied the item descriptors. It seems like the descriptors were instead used as a starting point in developing their own KSAs. Figures 26 and 27 are two examples for mathematics comparing what the panelists entered and what the item descriptors were. Some panelists used a simplified version of the item descriptors for their KSAs (e.g., Figures 26, generally) while some panelists added to the item descriptors (e.g., Figure 27, generally). There are examples of more involved elaborations, but those are not in either of Figures 26 and 27. For each of Figures 26 and 27, there were KSAs that were related to the descriptors and those that seemed totally different from the descriptors. In Figure 26, the KSA "Knowledge of the Pythagorean theorem; problem solving in 3 dimension situations." is and elaboration of the item descriptor "Compute length of hypotenuse on face of a 3D solid." In Figure 27, the KSA "Proper use of ruler and protractor. Underestand triangle properties" do not really match "Recognize and apply the fact that standard deviation is a measure of spread." Figure 27 also has examples of item descriptors that were copied almost verbatim such as "Interpret information on a spreadsheet and determine correct formula" versus "Interpret information in a spreadsheet and identify correct formula." The findings described for mathematics are very similar to those in Reading. Unfortunately, direct comparisons between the KSAs that panelists identified and the item descriptors are not available at this time.

One surprising finding in operational 3 is that relevance of items to the training program was again a comment that made an appearance in KSAs for Computer Support Specialists. The difference between this time and the pilot study is that their comments of "Not relevant to

computer support specialist" appear after the KSA. Lastly, with the provision of the item descriptors, which are one-liners, a lot of panelists wrote one-liner KSAs. There were still a number of panelists who wrote multiple KSAs for each item. As an example, a panelist wrote the following as the KSAs for one item.

"calculator skills
volume
conversions both minutes to hours and gallons/cubic feet
rates
approximation
rounding
fractions and decimals
multi-step problem
order of operations within multi-step equation
combine rates for net flow
volume = rate times time
answer must be in terms of hours ie reading directions"

No descriptor was provided for that item.

Figure 25: Panelist KSA Comments Compared to Item Descriptors Example 1*

Panelist6 Comment	ITEM DESCRIPTOR
Graphical solution of compound problem	Identify graphical solution of a compound linear inequality in one variable
Inequalities.	Identify graphical solution of a compound linear inequality in one variable
converwsion between degrees	Recognize a correct (equivalent) proportion for conversion between degrees and radians
Basic geometry	Identify the type of quadrilateral embedded in a 3D figure
Knowledge of thePythagorean theorem; problem solving in 3 dimension situations.	Compute length of hypotenuse on a 3D solid in some cases
Knowledge of geometry, recognition of triangles within a quadrilateral, and pythagorean theorem,algebraic expressin, wrong answer.	Compute length of hypotenuse on a 3D solid in some cases
Knowledge of geometry, recognition of triangles within a quadrilateral, pythagorean theorem,wrong answer.	Compute length of hypotenuse on face of a 3D solid
Using graphs to find rate of change	Read information from two graphs to make inference about rate of change
Identify missing information. Need surface area in square mile	Recognize that information is missing in a rate problem
Missing information Computation of total population based on given information.	Identify missing information in a rate problem
Knowledge of graphical functions.	Recognize the general form of an exponential function given graphically
Multiplication of decimal numbers, convrsion of indices to decimals.	Convert from scientific notation (negative exponent) to decimal representation
Knowledege of formulas and functions in excel.	Interpret information in a spreadsheet and identify correct formula
Use of functions and formulas in excel.	Interpret information in a spreadsheet and create formula
graphs and scatter charts	Estimate relative correlations from a scatterplot
Knowledge of bivariate data.	Estimate relative correlations from a scatterplot
linear function and recurrsion	Identify the recursive definition for a linear function given in tabular form
use calculator to determine probability	Find a compound probability
Knoledge of probability, dependedent and independent events.	Find a compound probability
inequalities and cube roots	Estimate cube root between consecutive whole numbers

This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

Figure 26: Panelist KSA Comments Compared to Item Descriptors Example 2*

Panelist4 Comment	Item Descriptor
had to solve the compound inequality for x id the solution from the given graphs	Identify graphical solution of a compound linear inequality in one variable
Differentiate geometric shapes.	Identify the type of quadrilateral embedded in a 3D figure
find links of sides of cube and area of trapezoid. 2 correct	Compute length of hypotenuse on a 3D solid in some cases
Understand area calculation. Partial credit if certain parts are correct and others are not.	Compute length of hypotenuse on a 3D solid in some cases
Understand area calculation. Partial credit if certain parts are correct and others are not.	Compute length of hypotenuse on face of a 3D solid
Interpret the question, and know the difference between density and area. /and that Area is needed to know which town has the greater population.	Recognize that information is missing in a rate problem
Partial credit if explanation describes missing information	Identify missing information in a rate problem
Understand counter examples related to algebraic equations. Partial credit if conclusion or hypothesis is correct.	Use hypothesis, conclusion, and definition to identify counterexamples in an algebraic setting in some cases
use hypothesis conclusion and definition to identify counter examples in an algebraic setting in two cases	Use hypothesis, conclusion, and definition to identify counterexamples in an algebraic setting
determine the inverse of a function	Find the inverse of a linear function defined symbolically
Interpret information on a spreadsheet and determine correct formula	Interpret information in a spreadsheet and identify correct formula
need to create formulas based on a spreadsheet information	Interpret information in a spreadsheet and create formula
Must be able to compare and interpret slopes in linear equastions. Proper use of calculator. Partial credit if answers are correct but no explanation, one of two correct answers.	Interpret/compare slope xor intercept of two linear functions, given as equations in literal slope-intercept form and graphically
Ability to compare slope properties of linear equations. Correct use of calculator. Partial credit if explanation is correct but answer is incorrect.One of two answers correct. Correct anwer but no explanation.	Interpret/compare slope and intercept of two linear functions, given as equations in literal slope-intercept form and graphically
use slope and xy plane coordinates	Interpret/compare slope and intercept of two linear functions, given as equations in literal slope-intercept form and graphically
identify charicteristics of a minor word problem	Identify characteristics of a well-designed survey

*This is an excerpt. The rest of the KSAs entered by this panelist may be found in the Appendix.

Concluding Remarks

There was no intention to examine KSAs noted by panelists on their OIBs. This is not a customary part of Bookmarking implementations in NAEP or other assessment programs. The examination of KSAs was prompted by the unusual results from the pilot study (Loomis, 2012). The CAB software afforded us the ability to conveniently examine what panelists had written, realizing that for the most part, they were not writing the types of comments or annotations that we expected and intended them to write based on the instructions we provided. This led us to improvements in our procedures. We would recommend some of these procedural improvements in the implementation of Bookmarking standard setting, and some for standard setting in general.

- Standardize instructions through General Sessions and the same PowerPoint Slides for providing instructions. These are in addition to a Facilitator Handbook that serves as the script for process facilitators, and facilitator training prior to the standard setting meeting as well as a daily debriefing session.
- Provide KSAs for the items as a starting point for the panelists. They may agree with them and take them as-is. They may disagree with them and create their own, or modify them to what they think is appropriate. Anecdotal information from experienced facilitators of Bookmarking standard setting says that it is always a struggle for panelists to come up with KSAs. This phenomenon is not isolated to a situation like JSS where there is no "strong alignment between criteria… and the assessment for which the criteria are applied" (p. 22, Loomis, 2012). So, even if it is just for efficiency in process implementation, providing them with starting KSAs, such as the anchor descriptions, will be helpful. At an extreme, one might consider having those anchor descriptions already in the CAB OIB and CROIB when panelists start the KSA review. This author will not go that far as the thought process and internalization that is realized by agreeing or disagreeing with the initial entry and the action of entering the KSAs is essential to their understanding the items and what it takes to respond correctly in terms of KSAs. Further, findings from this study indicate that panelists seem to use the anchor description for starting the discussion of the item, but they do not use the anchor descriptions are their comments.

A limitation of the current paper is that the evidences provided here are snapshots, and do not necessarily generalize to the whole panel, the whole operational implementation, or the JSS standard-setting effort. Thus, evidence here may be considered just a notch or two above anecdotal. For more definitive findings one might consider coding the KSAs to indicate whether they are really KSAs and not extraneous comments (e.g., "this item is not required for our program;" "I sometimes have problems with geometry items like this myself"); and if they are KSAs, are they actually identified correctly (not saying "able to add fractions," when it should be "recognize what right angles are")? This approach might allow us to quantify the improvements based on each level of support provided to panelist when they identify KSAs.

Despite the limitation of the study, the finding regarding KSAs is important in terms of procedural validity of cut scores resulting from Bookmarking standard setting. The examination of the KSAs is important during the process and after the process. The best indications whether panelists have a good grasp of the assessment for which standards are being set are the KSAs that they identify for each item. Without understanding what students should know and be able to

do to receive credit for each item, how can panelists gauge whether a student matching the description of performance at the borderline will be able to respond correctly to an item with a 0.67 probability? If the KSAs indicate that the panelists are not ready, standard setters might decide to provide additional training and instruction before proceeding to bookmarking. After standard setting, the KSAs provided by panelists might provide information if concerns are raised with the cut scores that were set.

On paper-based Bookmarking, an onsite examination of KSAs written by panelists will mean flipping through pages of OIBs and reading panelists' handwriting. For an examination of KSAs after standard setting, it will make sense to have the KSAs entered in a database. This is a time-consuming and labor-intensive endeavor. Our ability to engage in an investigation of KSAs written by panelists was a collateral benefit of the computerization of Bookmarking. As this paper has illustrated, the use of technology will not only increase process efficiency (as intended), it is also bound to increase the effectiveness of the Bookmarking standard setting method.

References

ACT, Inc. (2005a, April). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade twelve mathematics*: Process report. Iowa City, IA: Author.

ACT, Inc. (2005b, April). *Developing achievement levels on the 2005 National Assessment of Educational Progress in grade twelve mathematics: Technical report*. Iowa City, IA: Author.

ACT, Inc. (2007). *Developing achievement levels on the 2006 National Assessment of Educational Progress in grade twelve economics*: Process report. Iowa City, IA: Author.

ACT, Inc. (2010). *Developing achievement levels on the 2009 National Assessment of Educational Progress in science for grades four, eight, and twelve: Process report*. Iowa City, IA: Author.

Loomis, S.C. (2012). *A study of "irrelevant" items: Impact on bookmark placement and implications for college and career readiness.* Paper presented in the *Setting Academic Prepareness Standards for Job Training Programs: Are we prepared?* Session at the annual meeting of the National Council in Measurement in Education, April 14, 2012, Vancouver, British Columbia, Canada.

Measured Progress & WestEd. (2012). *National Assessment of Educational Progress Grade 12 preparedness research project judgmental standard setting (JSS) studies: Technical report*. Dover, NH: Authors.

National Assessment Governing Board. (2009). *Making new links, 12th grade and beyond: Technical panel on 12th grade preparedness research final report*. Washington, D.C.: U.S. Department of Education.

National Assessment Governing Board. (2010). *Design document for 12th grade NAEP preparedness research judgmental standard setting studies*. Washington, D.C.: U.S. Department of Education.

National Assessment Governing Board. (2011). *Reading Framework for the2011 National Assessment of Educational Progress*. Washington, D.C.: U.S. Department of Education.

WestEd & Measured Progress. (2011). *National Assessment of Educational Progress Grade 12 preparedness research project judgmental standard setting (JSS) studies: Process report*. San Francisco, CA: Authors.