

CAE

Standard-Setting Study Final Report

December 12, 2014

Doris Zahner
Principal Measurement Scientist & Director of Test Development



cwra+

Middle Schools

Copyright © 2017 Council for Aid to Education

With the creation and launch of CWRA+ for Middle School, a standard-setting study was conducted to formally establish fair and defensible levels of mastery for this new assessment. CWRA+ for Middle School is an enhancement of the College Work Readiness Assessment, designed to provide valid and reliable information on students' critical-thinking and written-communication skills at the institutional level, as well as at the individual student level. Stakes have now been attached to these individual student levels; hence the need for formally establishing levels of performance. The study was held at CAE headquarters in New York, New York on December 12, 2014.

A. Guiding Principles

The design and execution of the standard-setting study for CWRA+ for Middle School were consistent with procedures adopted in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999). Relevant practices recommended in these documents were applied to study activities relating to the selection and training of the panel of judges, selection and implementation of the standard-setting methods, provision of feedback to the panel, and documentation of the findings.

B. Selection of the Standard-Setting Method

Standard-setting activities, in most standardized assessments, are governed by two general types of approaches: norm-referenced procedures, which yield a **relative** test standard; and criterion-referenced procedures, which yield an **absolute** test standard.

The norm-referenced approach sets the passing point based on test scores and has a set percentage of passing students, thus establishing a “passing rate” (75% as an example). The “relative” nature of this type of standard setting means that the actual passing point (i.e., the number of questions one must answer correctly in order to pass) may shift in relation to the proficiency level of the student cohort taking the exam, while the percentage of successful students will remain constant. CAE views this type of standard used in educational settings where instructors will grade students “on a curve.”

In contrast, criterion-referenced procedures set the passing point based on a predetermined or “absolute” level of mastery required to achieve a passing score. Thus, the level of proficiency associated with the passing point remains constant across different forms, while the passing rate varies in relation to the ability level of the student cohort.

A serious shortcoming associated with norm-referenced techniques is that, due to shifts in student ability from one form of an examination to the next, there is a significant probability of granting a level of mastery to an individual whose proficiency level is below the minimum standard. Since the percentage of students awarded the different levels of mastery under this approach remains constant over time, a decline in the proficiency level of students would fail to produce a decrease in the number of students being granted the level of mastery.

Criterion-referenced standard-setting methods, by contrast, are not subject to the shortcoming described above. For this reason, as well as in consideration of the public-protection aspect of examinations, criterion-referenced methods are favored in the field of standardized testing. Of the available criterion-referenced standard-setting techniques, the Bookmark (Mitzel, Lewis, Patz, & Green, 2001) method is the most popular in educational settings. Bookmarking is popular because it is applicable to selected-response as well as open-ended, constructed-response items. However, in order to utilize this method, item statistics, either in the form of *p*-values for selected-response items or rubric scores for constructed-response items, are

required. Since the standard-setting study occurred after the first administration of CWRA+, CAE will have data for the Bookmark method. The items (or responses) are arranged in order of difficulty, and expert judges are asked to pick the point at which, as an example, Basic, Proficient, and Advanced students would answer correctly (or respond well to) all items (or responses) occurring before this point.

Jaeger (1991) and other experts suggest that the panel of judges in standard-setting studies be content experts who either supervise or train entry-level candidates, or have taken the assessment themselves within the past five years. Since CWRA+ is a new assessment, CAE did not have any eligible judges who previously took the test. CAE followed these recommendations in the identification of qualified individuals to serve as judges in this study. The CWRA+ for Middle School panel had 11 members representing various institutional demographics. Table 1 has the names of the participants and their institutional affiliation.

Table 1: CWRA+ for Middle School Standard-Setting Study Participant List and Institutional Affiliation.

Participant	Institution
Darra Belle	Virginia Beach City Public Schools
Tammy Boeckman	Fort Dodge Middle School
Colleen Broderick	Donnell Kay Foundation
Will Daley	Readington Township Public Schools
Sarah Field	Envision
Dani Goldstein	Colorado Academy
Patricia Hilliard	NC New Schools
Michael Maloy	The Chapin School
Barbara Sargent	Readington Township Public Schools
Cody Whitesell	New Tech Network

C. Implementation of the Standard-Setting Method

A step-by-step description of the procedures used at the standard-setting meetings is presented below.

Step 1. Judges first reviewed the content of CWRA+ for Middle School by taking the exam in its entirety themselves at the beginning of the meeting. This exercise allowed judges to become familiar with the range of test content, types of test items, and the correct or best answer to each item. All judges were required to sign a security and confidentiality agreement prior to viewing the examinations.

Step 2. CAE provided an overview of the standard-setting study procedure and process.

Step 3. The judges discussed and defined the profile for the four different levels of mastery (Basic, Proficient, Accomplished, and Advanced). This discussion was based on the CWRA+ rubric and the knowledge, skills, and abilities needed in order to perform well on CWRA+ for Middle School. The purpose of this activity was to develop a consensus among judges for each level of mastery. During the subsequent rating activities, judges relied on these consensus profiles to make item-performance estimates. Judges broke into small groups (four groups of two to three judges), and each group discussed the characteristics of one level of mastery. The groups then reconvened and reported their findings to the large group and formed a consensus on all four levels of mastery. Table 2 has the summary for each level of mastery and a description of the students who are below the Basic level of mastery.

Table 2: *Student Levels of Mastery Profiles*

Level of Mastery	Profile
Below Basic	<p>Students who are below basic demonstrate limited control of grammatical conventions and create responses which contain several distracting errors. Student struggles to vary sentence structure and vocabulary. Below basic students may attempt to articulate an argument; however, the information remains largely disorganized.</p> <p>Below basic students ignore or misinterpret information in the Document Library and never reach the point of analysis. Accessibility or comprehension is problematic at this level.</p>
Basic	<p>Students at the basic level may be able to access limited idea(s) (e.g., using information from only one document), stating their position, but may neglect opposing points of view and may inconsistently detect logical or questionable assumptions in supporting documents. In general, basic students provide some evidence in their argument (e.g., fragmented, general, or vague), but that evidence may be insufficient or incorrect to support the claim or position in detail. They may also misrepresent critical information or demonstrate minimal ability to analyze and understand quantitative information. Claims or positions may be largely based in opinion or may not recognize limitations in documents used to frame the task. The basic student may lack an organized approach to their claim or a cohesive structure because of borrowed language from documents. Sentence structure may be rudimentary, overly repetitive, and/or difficult to understand. Additionally, grammatical errors are prevalent in the response and will likely detract from the meaning of their argument.</p>
Proficient	<p>Students at the proficient level can make a claim and are able to use some evidence to support that claim. The majority of evidence to support the claim is based on information from the documents, although some evidence may be based on opinion, or information not presented in documents. Proficient students understand the purpose of the data through a literal interpretation of that data. Some analyses may occur with inconsistencies. Their arguments are presented in a logical manner but may lack thoroughness. Little attempt is made at countering opposing positions. The responses will have a basic organizational structure that is apparent, but a few elements of arguments might be disordered or lacking in detail. The proficient student generates a responses is readable throughout, but their writing contains some mechanical issues. These errors do not distract from comprehension of the text, though.</p> <p>Students do not consistently identify the reliability of information. They may identify areas where data is conflicting, and detect bias in arguments. However, students may have difficulty evaluating alternate conclusions and making predictions based on data presented.</p>

<p>Accomplished</p>	<p>Students at the accomplished level create an organized response that includes a logical beginning, middle, and end, and/or states a clear problem with solution(s) and supporting evidence. A clear thesis is present. The organization at the accomplished level is clear but may be formulaic.</p> <p>In the course of generating a response, there is an attempt to distinguish between relevant and irrelevant information and select evidence that best supports the argument. Accomplished students are able to infer the appropriate qualitative and quantitative information from the documents and analyze that information with respect to their argument with few inconsistencies. The accomplished student synthesizes information from multiple sources, with few inconsistencies. All of the evidence that is provided in the response is drawn from the assigned texts rather than outside knowledge. The response is appropriate to the task, and all elements of the performance task are addressed. The tone of the response is generally appropriate to purpose and audience. Where appropriate, students acknowledge and attempt to refute alternative decisions/conclusions/ positions.</p> <p>The accomplished student writes complete sentences with varied sentence length and structure. The effective use of transition words is present throughout the response. Students may inconsistently take risks with voice or structure. At the accomplished level, student responses contain few grammatical or mechanical errors, and errors do not interfere with meaning.</p>
<p>Advanced</p>	<p>Advanced students demonstrate consistency, thoroughness, and originality by going beyond the expected. Advanced students state an explicit decision/conclusion/position that reflects insight and/or the complexity of the issue or situation. Students at this level integrate information from multiple sources and are able to filter and choose information that is both relevant and credible, demonstrating comprehension of what is presented in the documents. Students can draw conclusions and identify bias, if applicable. Students acknowledge that there are valid points on the opposing side.</p> <p>Advanced students demonstrate a command of organization that reflects logic, which makes it easy to follow their arguments. Students at this level commit to an organizational structure that creates a cohesive stance. Students do this by, for example, making use of topical paragraphs and building pros and cons. Advanced responses reflect the development of ideas with comprehensive use of both qualitative and quantitative evidence within the documents. Students at the Advanced level can successfully evaluate the value of information, not just collecting information to support the argument but also identifying connecting and conflicting information. Students at the advanced level can go beyond what is expected and beyond the explicit to make inferences.</p> <p>Advanced students write with precision and variety, indicating clarity of thought and an intentionality of choice. There is a level of sophistication that is evident in the response (e.g., appropriate voice, vocabulary).</p>

Step 4. The Bookmark (Mitzel, Lewis, Patz, & Green, 2001) method, a commonly used criterion-referenced passing point technique within the education sector, was then introduced. It is based on the judgments of a panel of experts regarding the expected test performance of students at each of the levels of mastery.

The Bookmark method was introduced to judges during an orientation session. Discussion topics included the philosophy of criterion-referenced standard-setting procedures, factors impacting item difficulty, the nature of the Bookmark method, and the use of the rating estimates to determine the passing score. Judges were also cautioned about common rating pitfalls, such as interpreting the rating question as an opportunity to “prescribe” required knowledge for a student rather than estimating likely performance (i.e., thinking that a student should rather than **will** answer the item correctly). Judges had an opportunity to ask questions regarding any aspect of the Bookmark method during this session.

CAE also included a discussion on the consequences of false positive and false negative results—that is, awarding students who do not have the appropriate level of mastery and failing to award qualified students with the appropriate level of mastery. If unqualified students are awarded an inappropriate level of mastery, this undermines the value of the credential. Conversely, if qualified students are not rewarded with the appropriate level of mastery, this calls into question the validity of the credential.

Step 5. Judges applied the Bookmark method to 10 sample CWRA+ selected-response items. The purpose of this activity was to ensure that the judges fully understood the rating task and the consequences of their item estimates. All 10 items had been administered previously, and CAE shared actual difficulty ratings with participants after they made their independent judgments.

Judges provided practice ratings in response to the following question, “As CWRA+ items become more difficult, where would a basic/proficient/accomplished/advanced student not be able to get any more questions correct?”

Each individual judge’s practice ratings were shared with the group; judges with high- and low-end ratings were asked to explain why they rated as they did. This discussion provided a good opportunity to reinforce the fact that item difficulty was not based solely on content, but also on item construction and the quality of the distractors. Some judges acknowledged that they would refocus their thinking after discussing their ratings with their colleagues.

Step 6. Following the workshop, judges applied the Bookmark method to 35 student responses from the PT, arranged from the lowest score to the highest score, and 25 selected-response items arranged from least to most difficult as measured by p-value. Item statistics were not supplied for this task and judges worked independently during the process to avoid rater bias.

Step 7. Finally, judges were asked to complete a survey that asked questions about the face and content validity of CWRA+ for Middle School. All materials were collected by CAE.

D. Survey Results

The survey results of the CWRA+ for Middle School post-standard-setting study (Table 4) indicates that most respondents either agree or strongly agree (90%) that CWRA+ for Middle School measures skills that high school students should possess upon graduation. Furthermore, a large majority of judges (80%) believe that middle school curriculum should include tasks like the CWRA+ for Middle School as part of their curriculum.

A large majority of judges indicated that successful performance on CWRA+ for Middle School is contingent upon factors such as good written-communication skills (90%), strong analytic-reasoning and problem-solving skills (90%), and strong quantitative-reasoning skills (90%).

Finally, only 50% of judges do not believe that the CWRA+ for Middle School tasks are similar to tasks given in high school, indicating a divide between what is taught and assessed inside the classroom and what is expected of students post-graduation.

Table 4: CWRA+ for Middle School Post-Standard-Setting Survey Results

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
CWRA+ measures important skills that middle schoolers should possess	10%	0%	0%	10%	80%
Important for students to prepare to do well on tasks like CWRA+ for Middle School	10%	0%	10%	40%	40%
Middle school courses should include tasks like this as part of their curriculum	10%	0%	10%	0%	80%
CWRA+ for Middle School prompts are similar to the type of assessments students are given in middle school	0%	20%	40%	20%	20%
Students need good writing skills to perform well on CWRA+ for Middle School	0%	0%	10%	20%	70%
Students need good analytic-reasoning and problem-solving skills to perform well on CWRA+ for Middle School	10%	0%	0%	20%	70%
Students need good quantitative-reasoning skills to perform well on CWRA+ for Middle School	0%	0%	10%	30%	60%
CWRA+ for Middle School prompts are similar to assessment prompts that students will face in high school	0%	10%	40%	30%	20%
CWRA+ for Middle School prompts are similar to assessment prompts that students will face during high school admission processes (if applicable)	0%	37.5%	25%	12.5%	25%
CWRA+ for Middle School prompts are similar to assessment prompts that students will face during college school admission processes	0%	33.3%	22.2%	33.3%	11.1%
College admissions offices should place more value on the types of skills assessed by the CWRA+ for Middle School	0%	0%	0%	22.2%	77.7%
Students who do well on the CWRA+ for Middle School would also perform well in a future job requiring analytic-reasoning and problem-solving skills	10%	0%	30%	30%	30%
Students who do well on the CWRA+ for Middle School would also perform well in a future job requiring good written-communication skills	10%	0%	20%	30%	40%
Learning how to perform well on a task like the CWRA+ for Middle School would help students compete in a global market	10%	0%	30%	40%	20%
Learning how to perform well on a task like the CWRA+ for Middle School would help students get ahead in life	10%	0%	40%	30%	20%
Learning how to perform well on a task like the CWRA+ for Middle School would help high school graduates make better life decisions	10%	20%	30%	40%	0%
Skills measured by the CWRA+ for Middle School are also important in middle school	0%	10%	20%	40%	30%

References

- ACT. (2013). *The reality of college readiness 2013*. Retrieved from <http://www.act.org/readinessreality/13/pdf/Reality-of-College-Readiness-2013.pdf>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Jaeger, R. M. (1991). Selection of judges for standard-setting. *Educational Measurement: Issues and Practice*, 10(2), 3–14. doi: 10.1111/j.1745-3992.1991.tb00185.x
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-282). Mahwah, NJ: Lawrence Erlbaum Associates.
- National Assessment of Educational Progress, National Center for Education Statistics. (2011). *Writing 2011: Summary of major findings*. Retrieved from http://nationsreportcard.gov/writing_2011/summary.aspx