

CAE

Standard-Setting Study Final Report

December 30, 2013

Update: December 4, 2014

Doris Zahner
Principal Measurement Scientist & Director of Test Development



cwra+

Copyright © 2017 Council for Aid to Education

With the creation and launch of CWRA+, a standard-setting study was conducted to formally establish fair and defensible levels of mastery for this new and improved assessment. CWRA+ 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 level results (e.g., college admissions officers can use CWRA+ scores for student applicants); hence the need for formally establishing levels of performance. The study was held at CAE headquarters in New York, New York on December 13, 2013. A two-step follow up study was conducted in November 2014 to establish the cut score for the Accomplished student, an additional level of mastery for the CWRA+.

A. Guiding Principles

The design and execution of the standard-setting study for CWRA+ 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 has seen this type of standard used in educational settings where instructors 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+ panel had 15 members representing various sectors of the industry. Table 1 has the names of the participants and their institutional affiliation.

Table 1: CWRA+ Standard-Setting Study Participant List and Institutional Affiliation

Participant	Institution
Mark Battersby	Capilano University (Canada)
Ray Bryant	Warwick Valley School District
Paul Carney	Minnesota State Technical and Community College
Lisa Gioe	Millennium Brooklyn High School
Peter Gow	Beaver Country Day School
John Gulla	E.E. Ford Foundation
Bonnie Hain	Baltimore County School District
Michael Maloy	The Chapin School
Jonathan Martin	Independent Consultant
Syna Morgan	Douglas County School District
Andrew Niblock	Greenwich Country Day School
Dominic Randolph	Riverdale Country School
Drew Schrader	New Tech Network
Tyler Thigpen	Mount Vernon Presbyterian School
Amada Torres	National Association of Independent Schools
Todd Wirt	Wake County School District
Doug Wren	Virginia Beach City Public Schools

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+ by taking the exam in its entirety 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, and Advanced). This discussion was based on the CWRA+ rubric and the knowledge, skills, and abilities needed in order to perform well on CWRA+. 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 (three groups of five 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 three levels of mastery. On November 13, 2014, a fifth panel convened to develop the profile for the Accomplished level of mastery. Based on the distribution of the data from the first full administration of CWRA+, CAE decided to further delineate the Proficient students into either Proficient or Accomplished, a higher level 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	Students who are below basic frequently make severe errors and often interfere with meaning. Students write simple sentences and some non-sentences.
Basic	<p>Students at the basic level create responses that state or imply a decision, conclusion or position and provide some analysis that may be minimal, inaccurate or irrelevant. A basic student would provide an argument with some supporting information from sources and an attempt to cohesively organize that argument. Yet, the elaboration is limited and the organization lacks sufficient cohesion and clarity. For the basic student, severe errors are infrequent, but there are minor errors that sometimes interfere with meaning. The basic student also writes sentences that are similar in structure in length with an overreliance on sentences with simple structure. They draw obvious inferences from sources, rarely recognizes relevant information and takes all information at face value.</p> <p>Analysis and Problem Solving and Writing Effectiveness are more important than Writing Mechanics in making the decision on for the cut score decision.</p>
Proficient	<p>Students at the proficient level have the ability to make inferences from the document and provide some support for position but may omit some evidence. They address most elements of the task although sometimes tangentially. Students make a few accurate claims about the quality of evidence while citing the evidence provided in the documents. However, there may have few misinterpretations of the information and evidence provided in the documents.</p> <p>The students at this level are writing generally understandable sentences with minor errors and use conventions of standard written English. The student responses are communicated in a way that is readily comprehensible.</p> <p>There is an evaluation the relative value of common logical strategies (e.g., bad cause and effect). They extract meaningful information and recognize utility from basic graphs and be able to draw conclusions from them. There is an understanding of correlation versus causality as well as a basic understand of the design of the experiment. Proficient students will know what makes a credible scientific claim and provide an appropriate critical evaluation of sources.</p>

<p>Accomplished</p>	<p>Students at the accomplished level of mastery have the ability to make inferences from the document and provide sufficient evidence (based on multiple sources) to support their claim. This would include generating accurate interpretations of the document library, developing coherent arguments using much of the information provided in the documents, and potentially identifying, but not fully developing potential future steps and the need for additional research. They are also able to identify and address bias when making inferences or drawing conclusions, assess the relevancy of the qualitative and quantitative data (e.g., read and understand a graph and identify limitations and shortcomings; demonstrate an understanding that correlation does not necessarily imply causality), distinguish credible versus non-credible sources of information, and generate counter-claims. Accomplished students state a decision/recommendation/position and develop their argument based upon the identified information; however, they fall short of using evidence to fully support and leverage their argument. They have the ability to identify and extend the impact of the supporting versus counter- evidence and their broader implications.</p> <p>The accomplished students write responses that are cohesive, organized, and elaborated effectively. The student recognizes the correct audience and writes in a way that demonstrates understanding of the intended audience. The sources (documents) of evidence in support of students' claims can be identified. The student's intent is clear and the organization or the argument and understanding it represents is accurate and logical. There may be some minor spelling and syntax errors, but the sentences are generally well-constructed with varying and sometimes advanced vocabulary and structure, communicating a level of sophistication in the response.</p>
<p>Advanced</p>	<p>Students at the advanced level discern the merit of information and evaluate the strength of arguments, including identifying bias. They demonstrate a thorough evaluation of the evidence by making connections between the information found in the documents, potentially identifying patterns, and if applicable, refutes false or weak claims, that ultimately informs one's response. They clarify potential further steps, either a next step moving forward or additional research that is needed or would be helpful. They also address counter-arguments and demonstrate the weaknesses of them and/or the ways in which they are less compelling in order to strengthen one's own arguments.</p> <p>Advanced students provide a decision/recommendation with thorough support of the argument articulated in an effective way. The evidence is thoroughly examined, including addressing and navigating contradictory responses, and the interpretation of the documents is comprehensive. They fully respond to the prompt.</p> <p>Student writing is precise, purposeful, uses a varied vocabulary, sentence structure and length, and is free or almost entirely free from mechanical error. Their responses are organized in a fluid, coherent, and engaging way. It is easy to follow the student's argument, which also has the correct audience in mind and appropriately addresses them. They use the correct genre to deliver the response, whether it is a blog response, report, memo, speech, etc.</p> <p>Students should be able to consistently reason analytically and solve problems and be able to understand the nuances when integrating information across multiple sources.</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 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 questions (SRQs). 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.

During the follow-up study to establish the cut score for the Accomplished level of mastery, panelists were given student responses for the Performance Task (PT) and SRQs from the Proficient level of mastery and asked to use the Bookmark method to identify an additional cut score. This new cut score differentiates Proficient students from Accomplished students and the cut score for the Accomplished level of mastery was inserted between the existing Proficient and Advanced cut scores. The original cut scores for the Below Basic, Basic, Proficient, and Advanced levels of mastery were not altered.

Step 6. Following the workshop, judges applied the Bookmark method to 35 student responses from the Life Expectancy PT, arranged from the lowest score to the highest score, and 25 SRQs (from Video Games, Cigarette Tax, and Psychology and Ethics), 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+. All materials were collected by CAE.

D. Survey Results

The survey results of the CWRA+ post-standard-setting study (Table 4) indicated that all respondents strongly agree (100%) that CWRA+ measures skills that high school students should possess upon graduation. Furthermore, all of the judges believed that it is imperative to prepare to do well on tasks as reflected on CWRA+.

A large majority of judges indicated that successful performance on CWRA+ is contingent upon factors such as good written-communication skills (83.3%), strong analytic-reasoning and problem-solving skills (100%), and strong quantitative reasoning skills (86.6%). Meanwhile, all participants agree that successful performance on CWRA+ would be an indicator of successful performance on similar future job-related tasks. Most judges also agree that the learning process of how to become a successful performer on CWRA+ may also prepare individuals to become successful students at the college level.

Finally, 47.6% of judges did not believe that the CWRA+ 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+ Post-Standard-Setting Survey Results

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
CWRA+ measures important skills that high school graduates should possess	0%	0%	0%	0%	100%
Importance for students to prepare to do well on tasks like CWRA+	0%	0%	0%	6.7%	93.3%
High school courses should include tasks like this as part of their curriculum	0%	0%	0%	21.4%	78.6%
CWRA+ tasks are similar to the type of assessments students are given in high school	6.7%	40%	33.3%	6.7%	13.3%
Students need good writing skills to perform well on CWRA+	0%	0%	6.7%	66.7%	26.7%
Students need good analytic reasoning and problem solving skills to perform well on CWRA+	0%	0%	0%	6.7%	93.3%
Students need good quantitative reasoning skills to perform well on CWRA+	0%	0%	13.3%	53.3%	33.3%
Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
CWRA+ prompts are similar to assessment prompts that students will face in college	20%	13.3%	46.7%	6.7%	13.3%
CWRA+ prompts are similar to assessment prompts that students will face during high school admission processes	0%	27.3%	45.5%	27.3%	0%
CWRA+ prompts are similar to assessment prompts that students will face during college school admission processes	0%	36.4%	45.5%	18.2%	0%
College admissions offices should place more value on the types of skills assessed by the CWRA+	0%	0%	0%	36.4%	63.6%
Students who do well on the CWRA+ would also perform well in a future job requiring analytic reasoning and problem solving skills	0%	0%	0%	33.3%	66.7%
Students who do well on the CWRA+ would also perform well in a future job requiring good written communication skills	0%	0%	0%	33.3%	66.7%
Learning how to perform well on a task like the CWRA+ would help students compete in a global market	0%	0%	0%	41.7%	58.3%
Learning how to perform well on a task like the CWRA+ would help students get ahead in life	0%	0%	16.7%	41.7%	41.7%
Learning how to perform well on a task like the CWRA+ would help high school graduates make better life decisions	0%	8.3%	25%	16.7%	50%
Skills measured by the CWRA+ are also important in middle school	0%	0%	8.3%	33.3%	58.3%
Skills measured by the CWRA+ are also important in elementary school	0%	16.7	16.7%	33.3%	33.3%

References

- 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
- National Center for Public Policy and Higher Education, & Southern Regional Education Board (2010, June). *Beyond the rhetoric: Improving college readiness through coherent state policy*. Retrieved from http://www.highereducation.org/reports/college_readiness/index.shtml