CAE

Standard-Setting Study Final Report

December 30, 2013

Update: December 4, 2014

Doris Zahner
Principal Measurement Scientist & Director of Test Development

cla+

With the creation and launch of CLA+, a standard-setting study was conducted to formally establish fair and defensible levels of mastery for this new and improved assessment. CLA+ is an enhancement of the Collegiate Learning 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., employers can use CLA+ scores as a screening tool for interviewing candidates); hence the need for formally establishing levels of performance. The study was held at CAE headquarters in New York, New York on December 12, 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 CLA+.

A. Guiding Principles

The design and execution of the standard-setting study for CLA+ 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, 2014). 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 professors 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 CLA+, 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 should be content experts who either supervise or train entry-level candidates, or have taken the assessment themselves within the past five years. Since CLA+ 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 CLA+ panel had 12 members representing various sectors of the industry. Table 1 has the names of the participants and their institutional affiliation.

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

Participant	Institution
Aviva Altman	Johnson & Johnson
Jon Basden	Federal Reserve
Mark Battersby	Capilano University (Canada)
Paul Carney	Minnesota State Technical and Community College
Eric Daffron	Ramapo College
Anne Dueweke	Kalamazoo College
Terry Grimes	Council of Independent Colleges
Sonia Gugga	Columbia University
Marsha Hirano-Nakanishi	California State University System
Rachel L. Kay	McKinsey & Company
Michael Poliakoff	American Council of Trustees and Alumni
Elizabeth Quinn	Fayetteville State University
Paul Thayer	Colorado State University

C. Implementation of the Standard-Setting Method

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

Step 1. Judges first reviewed the content of CLA+ 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 procedures and processes.

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 CLA+ rubric and the knowledge, skills, and abilities needed in order to perform well on CLA+. 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 four 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 14, 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 the CLA+, 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 do not meet the minimum requirements to merit a basic level of mastery.
Basic	Students at the basic level should be able to demonstrate that they at least read the documents, made a reasonable attempt at an analysis of the details, and are able to communicate in a manner that is understandable to the reader. Students should also show some judgment about the quality of the evidence. Students at the basic level should also know the difference between correlation and causality. They should be able to read and interpret a bar graph but not necessarily a scatter plot or comprehend a regression analysis. Tables may be out of reach for basic students as well.
Proficient	Students at the proficient level should be able to extract the major relevant pieces of evidence provided in the documents and provide a cohesive argument and analysis of the task. Proficient students should be able to distinguish the quality of the evidence in these documents and express the appropriate level of conviction in their conclusion given the provided evidence. Additionally, students should be able to suggest additional research and/or consider the counterarguments. Minor errors in writing need to be defined rigorously. Proficient students have the ability to correctly identify logical fallacies, accurately interpret quantitative evidence, and distinguish the validity of evidence and its purpose. They should have the ability to determine the truth and validity of an argument. Finally, students should be able to know when a graph or table is applicable to an argument.
Accomplished	Students at the accomplished level of mastery should be able to analyze the information provided in the documents, extract relevant pieces of evidence, and make correct inferences about this information. Accomplished students should be able to identify bias, evaluate the credibility of the sources, and craft an original and independent argument. When appropriate, students will identify the need for additional research or further investigation. They will refute some but not all of the counterarguments within the documents and use this information to advance their argument. Accomplished students also have the ability to correctly identify logical fallacies, accurately interpret and analyze qualitative and quantitative evidence (e.g., graphs and charts), and incorporate this information into their argument. Students will be able to correctly identify false claims and other sources of invalid information and integrate this information in their responses. Student responses are presented in a cohesive and organized fashion. There may be infrequent or minor errors in writing fluency and mechanics, but they will not detract from the reader's comprehension of the text.
Advanced	Students at the advanced level demonstrate consistency, completeness, and show a command of the English language in their response. They have a level of sophistication that is not seen in the proficient or basic levels. Advanced students create and synthesize the provided evidence, are comfortable with ambiguity, are able to structure their thoughts, understand causality, add new ideas, and introduce new concepts in order to create or seek new evidence. They think about conditions and nuances and express finer points and caveats by proposing a conditional conclusion. The students at this level display creativity and synthesis, while understanding the finer points in the documents. For example, advanced students will be able to synthesize the information across multiple documents and address the ambiguities in the data that are presented, such as outliers and knowing how sample size affects outcomes. Advanced students will also be able to identify and highlight gaps in logic and reasoning.

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 CLA+ 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 CLA+ 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 PT and SRQs from the Proficient level of mastery and asked to use the Bookmark methodology 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 28 student responses from the Lionfish 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 CLA+. All materials were collected by CAE.

D. Survey Results

The survey results of the CLA+ post-standard-setting study (Table 4) indicated that all judges agree (8.3% agree, 91.7% strongly agree) that CLA+ measures skills that are essential for students' success in college and beyond.

Meanwhile, 63.6% of the judges do not believe that the CLA+ tasks represent assignments that are in alignment with those given in college. Additionally, 66.7% of the judges think that the CLA+ tasks are not necessarily representative of the tasks (i.e. college entrance essays) encountered during college admission procedures. A possible explanation is the fact that many assessments in college consist of only selected-response items. It has long been recognized that selected-response items may not necessarily elicit higher-order thinking skills in the same manner as performance tasks. Therefore, CLA+ comprises both performance tasks and selected-response items in an attempt to assess 21st-century skills. However, about 90% of the respondents believe that the CLA+ tasks portray scenarios that may be representative of situations encountered in careers after college graduation.

Finally, the majority of judges (70%-90%) indicated that successful performance on CLA+ prepares students to become successful contributors to the workforce. Factors that were investigated included the possibility of being able to compete in a global market, and individual growth and development (students getting ahead in life and being able to make appropriate life choices).

Table 4: *CLA+ Post-Standard-Setting Survey Results*

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
CLA+ measures important skills that college graduates should possess.	0%	0%	0%	8.3%	91.7%
Importance for students to prepare to well on tasks like CLA+.	0%	8.3%	8.3%	33.3%	50%
College courses should include tasks like this as part of their curriculum.	0%	0%	16.7%	25%	58.3%
CLA+ tasks are similar to the type of assessments students are given in college.	0%	63.6%	18.2%	18.2%	0%
Students need good writing skills to perform well on CLA+.	0%	8.3%	8.3%	50%	33.3%
Students need good analytic reasoning and problem-solving skills to perform well on CLA+.	0%	0%	0%	25%	75%
Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Students need good quantitative-reasoning skills to perform well on CLA+.	0%	0%	16.7%	75%	8.3%
CLA+ prompts are similar to situations that students will face in careers after college.	0%	10%	0%	60%	30%
CLA+ prompts are similar to assessment prompts that students will face during college admission processes.	0%	66.7%	22.7%	0%	11.1%
College admissions offices should place more value on the types of skills assessed by CLA+.	0%	12.5%	25%	12.5%	50%
Students who do well on CLA+ would also perform well in a future job requiring analytic-reasoning and problem-solving skills.	0%	0%	0%	40%	60%

Students who do well on CLA+ would also perform well in a future job requiring good written-communication skills.	0%	0%	20%	40%	40%
Learning how to perform well on a task like the CLA+ would help students compete in a global market.	0%	0%	10%	40%	50%
Learning how to perform well on a task like the CLA+ would help students get ahead in life.	0%	10%	0%	50%	40%
Learning how to perform well on a task like the CLA+ would help college graduates make better life decisions.	0%	30%	0%	50%	20%

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