Standard Setting for Performance Assessment

This document describes the procedures, recommendations, and expected outcomes for standard-setting studies for the CLA+ and CWRA+. The studies will be conducted following the first administration and scoring of the CLA+/CWRA+ in the spring/summer of 2013.

The CLA+ will be used for a variety of purposes. In addition to the current primary uses at the institutional level for value-added assessment and institutional ranking, CLA+ will be used to obtain individual student results. In order to attach meaning to individual scores, an individual's performance may be judged relative to others that have taken the assessment (norm-referenced), or relative to an absolute standard or criterion representing an agreed-upon benchmark for competency or proficiency (criterion referenced). In order to provide meaningful judgments about student proficiency or competency, from a criterion-referenced standpoint, the CLA+ development team is proposing that a standard-setting study be conducted in order to formally establish performance levels for the CLA+. The CLA+ is an enhancement of the CLA, designed to provide valid and reliable information on students' critical thinking and writing skills at the institutional level as well as at the individual student level. There is the potential for stakes to be attached to these individual-level results; hence the need for formally establishing levels of performance.

STANDARD SETTING FOR THE CLA

Traditionally, the CLA has been a low-stakes, institution-level assessment of the value added by schools to their students' education. In 2007, CAE conducted a standard-setting study of the CLA to establish a set of benchmarks. This study was done in addition to the value-added and percentile information that is provided in the institutional reports (Hardison & Vilamovska, 2008). The results yielded a set of benchmark standards for entering college freshmen and a set for exiting college seniors.

Although this original study (Hardison & Vilamovska, 2008) was conducted in accordance with the procedures adopted in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999), the methodology used is not appropriate for the proposed CLA+ standard-setting study since the assessment itself has changed in a number of ways.

SELECTION OF THE STANDARD SETTING METHOD

Standard-setting activities in examination programs 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 establishes standards based upon students' test scores and compares student performance to one another. The relative nature of this type of standard setting means that the actual benchmark score that represents, for example, the 75th percentile, may shift in relation to the proficiency level of the cohort for a particular administration. However, the percentage of students falling within a certain proficiency range remains the same from administration to administration.

In contrast, criterion-referenced procedures establish a benchmark based on a predetermined or absolute level of mastery. Thus, the level of proficiency associated with the benchmark remains constant across different test forms and administrations, while the number of students who fall within each established proficiency range varies in relation to the ability level of the cohort.

Although less serious for low-stakes assessments, a shortcoming associated with norm-referenced techniques in high-stakes testing is that, due to shifts in student ability from one form of an administration to the next, there is a probability of incorrectly granting a certificate of proficiency to a student who does not merit the certificate. This should be a serious consideration if CAE is going to be granting certificates to CLA+ students as part of a high-stakes testing program.
Criterion-referenced methods, by contrast, are not subject to the shortcoming described above. For this reason, criterion-referenced methods are favored. Of the available criterion-referenced standard-setting techniques, the procedure attributed to William Angoff (1971) is the most widely used in the standardized assessment industry for selected-response items. Angoff suggested that expert judges review each item on a test, and then make an estimate of the probability that a “borderline” candidate would answer the item correctly. The borderline refers to the border between a minimally competent and an incompetent candidate. The sum of these probabilities, aggregated across judges, is the established standard for the test. The Angoff technique has been modified since its inception to improve the technical quality of the resulting standard.

Another popular technique used in standard-setting studies is the Bookmark (Lewis, Mitzel, Green, & Patz, 1999) method. Bookmarking is popular because it is applicable to both selected-response and 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 will occur after the administration of the beta-test of the CLA+, CAE will have data for the Bookmarking 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, a highly proficient student would correctly answer all items occurring before this point. The scores aggregated across judges, is the established standard for the test.

RECOMMENDATION

CAE is proposing the following for the CLA+ standard-setting study:

- Use the modified-Angoff (1971) or the Bookmark (Lewis, et al., 1999) method for the selected-response items.
- Use the Bookmark (Lewis, et al., 1999) method for the PT responses. Use total raw scores for this part of the study.
- Equate the forms.
- Recruit expert judges for two separate committees, one for high school and the other for higher education.
- The committees should consist of a panel of expert judges who are familiar with the knowledge, skills, and abilities of the target cohorts; demographically diverse; institution type (e.g., small/large; private/public); industry type (e.g., small/large; private/NFP; Field); race/ethnicity; geography; years of experience (oftentimes, those with fewer years of work experience work more closely with students entering the workforce).

ANTICIPATED OUTCOMES

The anticipated result of this study would be the establishment of three levels of performance standards: basic, proficient, and advanced. Of course, the language used for these benchmarks can be finessed or changed.

SAMPLE AGENDA (1 DAY)

- Judges complete the actual examination as if they are a CLA+ student.
- Presentation on the standard-setting process
- Review of the profiles for the basic, proficient, and advanced CLA+ students
- Modified-Angoff standard-setting procedure: Practice application
- Modified-Angoff standard-setting procedure
- Bookmarking procedure: Practice application
- Bookmarking procedure
- Group discussion of results

BUDGET

CAE anticipates that each study will cost between $15,000 and $20,000 (CLA+/CWRA+). The costs include a per diem for each participant (averaging $750 a day), travel expenses, and food. The study will require an overnight stay for participants who are traveling from out of town.

CAE is a non-profit organization best known for its work measuring and improving the quality of secondary and postsecondary education. CAE features two nationwide assessment programs—the Collegiate Learning Assessment (CLA+) and the College & Work Readiness Assessment (CWRA+). These instruments provide member institutions with tools to improve critical thinking and writing skills through the connection of effective teaching, learning, and assessment efforts framed through authentic performance-based practices.
REFERENCES


