



[collegiate**learning***assessment*]

FREQUENTLY ASKED TECHNICAL QUESTIONS

2007 – 2008

Council for Aid to Education
215 Lexington Ave., Floor 21
New York, NY 10016
Phone: 212-217-0700
Fax: 212-661-9766
Email: cla@cae.org
Web: www.cae.org/cla

TABLE OF CONTENTS

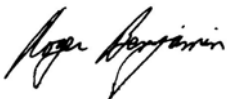
TASKS	1-2
SCORING	2-3
CONTROLLING FOR INCOMING ACADEMIC ABILITY	3-4
CORRELATIONS WITH OTHER MEASURES	4-5
SCALING PROCESS	5-6
ANALYSIS	6
RELIABILITY	7
VALIDITY	7-8
STUDENT EFFORT	8
OTHER FACTORS OF INTEREST	8-9

INTRODUCTION

Performance-based assessments are anchored in a number of psychometric assumptions different from those employed by common multiple-choice exams. As such, initiatives like the Collegiate Learning Assessment (CLA) represent a paradigm shift, the technical underpinnings of which remain unfamiliar to many faculty and institutional researchers. The CLA protocol is novel because it calls for 1) the institution, not the student, to be the initial unit of analysis, 2) a matrix sampling approach, and 3) a value added method which, in turn, requires evidence of the competencies students bring to college. Rightly, colleagues at campuses using or contemplating using the CLA often have a number of important questions. With this in mind, we intend to publish a technical manual that deals with, as comprehensively as possible, the many technical questions associated with the CLA. In the meantime, I hope you will be pleased with this document that addresses several of the major questions our colleagues frequently ask.

We could not continue to improve the CLA without your constructive advice. We would appreciate receiving any new questions that these responses suggest to you.

Sincerely,



Roger Benjamin, Ph.D.
President, Council for Aid to Education

CLA TASKS

How are CLA tasks developed?

The CLA is comprised of Performance Tasks and Analytic Writing Tasks. Performance Tasks ask students to engage in a “real-life” activity (such as preparing a memo or policy recommendation) that requires reviewing and evaluating several documents. There are two types of Analytic Writing Tasks: Make-an-Argument and Critique-an-Argument. The Make-an-Argument prompt asks students to explain why they would agree or disagree with a statement. The Critique-an-Argument prompt asks students to describe the shortcomings in an argument presented by someone else. All CLA tasks evaluate students’ ability to articulate complex ideas, examine claims and evidence, support ideas with relevant reasons and examples, sustain a coherent discussion, and use standard written English. Task development occurs through an iterative process. A team of researchers and writers generate ideas for Make-an-Argument and Critique-an-Argument prompts, and Performance Task storylines, and then contribute to the development and revision of prompts and Performance Task documents. For Analytic Writing Tasks, multiple prompts are generated, revised and pre-piloted, and those prompts that elicit good critical thinking and writing responses in pre-piloting are further revised and submitted to more extensive piloting.

Performance Task development is a much more involved process. During the development of Performance Tasks, care is taken to ensure that sufficient information is provided to permit multiple reasonable solutions to the issues developed in the Performance Task. Documents are crafted such that information is presented in multiple formats (e.g., tables, figures, news articles, editorials, letters, etc.). While working on the Performance Task, a list of the intended content from each document is established and revised. This list is used to ensure that each piece of information is clearly reflected in the document and/or across documents, and to ensure that no additional pieces of information are embedded in the document that were not intended. This list serves as a draft starting point for the analytic scoring items used in the Performance Task scoring rubrics. During revision, information is either added to documents or removed from documents to ensure that students could arrive at approximately three or four different conclusions based on a variety of evidence to back up each conclusion. Typically, some conclusions are designed to be better supported than others. Questions for the Performance Task are also drafted and revised during the development of the documents. The questions are designed such that the first questions prompt the student to read and attend to multiple sources of information in the documents, and later questions require the student to evaluate the documents and then use their analysis to draw conclusions and justify those conclusions using information from the documents.

After several rounds of revision, the most promising of the Performance Tasks and the Make-an-Argument and Critique-an-Argument prompts are selected for pre-piloting. Student responses from the pilot test are examined to identify what pieces of information are unintentionally ambiguous, what pieces of information in the documents should be removed, etc. After revision and additional pre-piloting, the best functioning tasks (i.e., those that elicit the intended types and ranges of student responses) are selected for full piloting. During piloting, approximately 60 students complete both an operational task and one of the new tasks. At this point, draft scoring rubrics are revised and tested in grading the pilot responses, and final revisions are made to the tasks to ensure that the task is eliciting the types of responses intended.

Why are both Performance Tasks and Analytic Writing Tasks necessary?

CLA scores reflect a holistic assessment of the higher order skills of critical thinking, analytic reasoning, written communication, and problem solving. All Performance Tasks and Analytic Writing Tasks require the use of all of these skills, but in different proportions. For example, Analytic Writing Tasks strongly emphasize written communication while Performance Tasks elicit greater use of problem solving skills, particularly with complex and sometimes contradictory materials. Similarly, prompts within each task type vary in which skills they draw upon most. Students are randomly assigned to a task type (Performance Task or Analytic Writing Task) and then to a prompt within that task, so each student only answers a small portion of the full complement of CLA prompts. By using this “matrix sampling” strategy, institutions reduce the testing burden on individual students and benefit from the full breadth of the task types.

SCORING

Can you describe the CLA scoring rubrics?

There are two types of items that appear on CLA scoring rubrics: holistic and analytic. Holistic scoring refers to general dimensions, such as evaluation of evidence, drawing conclusions, acknowledging alternative explanations and viewpoints, and overall writing, whereas analytic scoring refers to items particular to each prompt.

Scoring rubrics for the Performance Tasks are tailored to each specific prompt and include a combination of both holistic and analytic scoring items. Though there are many types of analytic items on the Performance Task scoring rubrics, the most common represent a list of the possible pieces of information a student could or should raise in their response. These cover the information presented in the Performance Task documents as well as information that can be deduced from comparing information across documents. The analytic items are generally given a score of 0 if the student did not use the information in their response, or 1 if they did. The number of analytic items varies by prompt.

Performance Task holistic items are scored on Likert scales. There are multiple holistic items per Performance Task that require graders to provide an evaluation of different aspects of critical thinking and reasoning in the student responses. These holistic items include areas such as the student’s use of the most relevant information in the Performance Task, their recognition of strengths and weakness of various pieces of information, overall critical thinking, and overall writing.

Critique-an-Argument rubrics also include a combination of analytic and holistic scores. Critique-an-Argument analytic items are a list of possible critiques of the argument presented in the prompt. In addition, a few holistic items are used to rate the overall tone, critical thinking and writing over the entire response. The method for computing a Critique-an-Argument or Performance Task raw score is essentially a unit-weighted sum of every analytic item and every holistic item.

For all task types, blank responses or responses that are entirely unrelated to the task (e.g., writing about what they had for breakfast) are assigned a 0 and are flagged for removal from the school-level results.

Make-an-Argument scoring rubrics include only holistic scores, scored on Likert scales. Because the scoring rubric consists only of holistic items, the rubric is the same for all Make-an-Argument prompts.

The holistic items include ratings for various aspects of writing (e.g., organization, mechanics) and critical thinking (e.g., reasoning and logic, sophistication and depth of treatment of the issues raised in the prompt) as well as two overall assessments of writing and critical thinking. The method for computing a Make-an-Argument raw score is essentially a unit-weighted sum of every holistic item.

How are CLA tasks scored?

During the 2007–08 CLA administration, all scoring was conducted by trained graders. In the future, a combination of machine and human scoring may be used.

How are graders trained and evaluated?

All grader candidates undergo rigorous training in order to become certified CLA graders. All training includes an orientation to the task and scoring rubric, instruction on how to use the scoring items, repeated practice grading a wide range of student responses, and extensive feedback and discussion after grading each response. After participating in training, graders complete a reliability check where all graders score the same set of student answers. Scorers with low agreement or reliability (determined by comparisons of raw score means, standard deviations and correlations among the scorers) are either further coached or removed from scoring.

CONTROLLING FOR INCOMING ACADEMIC ABILITY

How do you “crosswalk” between the ACT and the SAT?

If a participating institution collects ACT scores instead of SAT scores, they are converted to the SAT’s scale of measurement using a standard crosswalk. The maximum ACT score of 36 corresponds to the SAT (Math + Verbal/Critical Reading) maximum of 1600, an ACT score of 35 corresponds to 1580, and so forth. The full crosswalk is printed in the sample institutional report, available on our website. The correlation between ACT Composite and SAT Math + SAT Verbal/Critical Reading has been shown to be .92.

Sources: “Concordance Between ACT Assessment and Recentered SAT I Sum Scores” by N.J. Dorans, C.F. Lyu, M. Pommerich, and W.M. Houston (1997), *College and University*, 73, 24-31; “Concordance between SAT I and ACT Scores for Individual Students” by D. Schneider and N.J. Dorans, *Research Notes (RN-07)*, College Entrance Examination Board: 1999; “Correspondences between ACT and SAT I Scores” by N.J. Dorans, *College Board Research Report 99-1*, College Entrance Examination Board: 1999; *ETS Research Report 99-2*, Educational Testing Service: 1999.

How strong is the correlation between the SAT/ACT and the CLA?

At the student level of analysis, correlations range from .40 to .53 for the Analytic Writing Tasks, and .55 to .72 for the Performance Tasks. At the institutional level of analysis, correlations range from .73 to .88 for Analytic Writing Tasks and .78 to .92 for Performance Tasks.

If the SAT/ACT and CLA are so closely correlated, why can’t the SAT/ACT be used as a substitute

for freshmen scores? That is, why test freshmen at all?

A student's SAT or ACT score explains only about 20% of the variance in that student's CLA scores and a school's mean SAT or ACT score explains only about 75% of the variance in school means on the CLA. Moreover, if students' SAT/ACT scores were used as a substitute for freshmen scores, the problem would be that SAT/ACT would not account for first-year students' entering level analytic reasoning, critical thinking, problem solving, and written communication skills (as measured by the CLA instrument). SAT/ACT scores simply allow us to control for general cognitive ability.

How can you control for incoming ability if our students do not have SAT or ACT scores?

Institutions at which many students do not have SAT or ACT scores can choose to embed the Scholastic Level Exam (SLE) into CLA testing. The SLE is a short-form (12-minute) cognitive aptitude test produced by Wonderlic (a commercial test provider). Wonderlic reports that SLE scores, as a measure of cognitive ability, are stable over time. SLE scores need not be age-adjusted for students between ages 15 and 29.

What evidence do you have that SLE scores are equivalent to SAT scores to control for incoming academic ability?

In spring 2006, over 1,150 students (seniors at four-year colleges and universities and exiting students at community colleges) took the Scholastic Level Exam (SLE). These students also took either a 90-minute Performance Task or a 75-minute Analytic Writing Task as part of the CLA. Registrar offices supplied ACT and/or SAT scores for these students. Students were given 12 minutes to complete the 50-item SLE online. The mean total score (sum of all items, worth one point apiece) was 38, with a standard deviation of 8.

Student-level correlations between the SLE total score and ACT Composite, SAT Verbal, SAT Math, and SAT Composite Equivalent scores were as follows:

.68	SLE and ACT Composite
.68	SLE and SAT Verbal/Critical Reading
.66	SLE and SAT Math
.70	SLE and SAT Composite Equivalent

There were 24 schools where at least 10 students had an SLE and either an SAT or ACT score. The school-level correlation between the mean SLE total score and mean SAT Composite Equivalent score at these 24 schools was 0.92.

CORRELATIONS WITH OTHER MEASURES

To what degree is the NSSE correlated with the CLA?

Correlations between the NSSE and CLA were explored using data from the CLA feasibility study. This work was published in *Research in Higher Education*. An abstract of this article follows:

This study examines (1) the extent to which student engagement is associated with experimental and traditional measures of academic performance, (2) whether the relationships between engagement and academic performance are conditional, and (3) whether institutions differ in terms of their ability to convert student engagement into academic performance. The sample consisted of 1058 students at 14 four-year colleges and universities that completed several instruments during 2002. Many measures of student engagement were linked positively with such desirable learning outcomes as critical thinking and grades, although most of the relationships were weak in strength. The results suggest that the lowest-ability students benefit more from engagement than classmates, first-year students and seniors convert different forms of engagement into academic achievement, and certain institutions more effectively convert student engagement into higher performance on critical thinking tests.

Source: Carini, R., Kuh, G., Klein, S. STUDENT ENGAGEMENT AND STUDENT LEARNING: Testing the Linkages. *Research in Higher Education*. Vol. 47, No. 1, February 2006.

Are there linkages or relationships between the CLA and any standardized placement test, e.g., a test used to determine what initial math or English course a freshman should take, such that the placement test could serve as a control for the entering ability of students?

To date, we have not conducted research to determine whether any linkages or agreements between the CLA and various standardized placement tests that would determine an initial freshman course exist. That being said, some participating institutions are utilizing the CLA in a pre/post fashion to determine the efficacy of certain programs or courses for entering students.

SCALING PROCESS

What is the procedure for converting raw scores to scale scores?

Each Performance Task and Analytic Writing Task has a unique scoring rubric, and the maximum number of reader assigned raw score points differs across tasks. Consequently, a given reader-assigned raw score, such as 15 points, may be a relatively high score on one task but a low score on another task. To adjust for such differences, reader-assigned "raw" scores on the different tasks are converted to a common scale of measurement. This process results in "scale" scores that reflect comparable levels of proficiency across tasks. For example, a given CLA scale score indicates about the same percentile rank regardless of the task on which it was earned. This feature of the CLA scale scores allows combining scores from different tasks to compute a school's mean scale score for each task type as well as a total scale score across types.

To convert the reader assigned raw scores to scale scores, the raw scores on a measure were transformed to a score distribution that had the same mean and standard deviation as the SAT scores of the freshmen who took that measure. This type of scaling maintains the relative standing of a student on a task relative to other students who took that task. For example, the student with the highest raw score on a task will also have the highest scale score on that task, the student with the next highest raw score will be assigned the next highest scale score, and so on.

This type of scaling generally results in the highest raw score earned on a task receiving a scale score of approximately the same value as the maximum SAT score of any freshman who took that task. Similarly, the lowest raw score earned on a task would be assigned a scale score value that is approximately the

same as the lowest SAT score of any freshman who took that task. On very rare occasions, a student may achieve an exceptionally high or low raw score (i.e., well above or below the other students taking that task). When this occurs, it results in assigning a student a scale score that is outside of the normal SAT range. Prior to the spring of 2007, scores were capped at 1600 (the maximum allowable on the SAT). Capping was discontinued starting in fall 2007.

Do scaling equations change with each administration?

In the past, CAE revised its scaling equations each fall. However, many institutions would like to make year-to-year comparisons (i.e., as opposed to just fall to spring). To facilitate this activity, beginning in the fall of 2007, CAE has used the same scaling equations it developed for the fall 2006 administration. As a result of this policy, a given raw score on a task receives the same scale score regardless of when the student took the task.

ANALYSIS

What is the process for averaging students' scores for comparison and reporting?

To be included in the calculation of a mean score for a school, students must be in the correct class year (verified by the registrar); have either an ACT Composite, or SAT Math and SAT Verbal/Critical Reading score, or SLE score; and have a complete CLA task score (either Performance Task or Analytic Writing Task). The total scale score is the mean of the Performance Task and Analytic Writing Task scale scores for those students with ACT/SAT/SLE scores and who are in the correct class year.

Does CLA analysis account for ceiling effects?

No school's averages achieve the theoretical maximum of scaled CLA scores. There are however, *individual* students who have, in the past, achieved a maximum scale score on the CLA, as a function of exceptional performance. Historically, we capped, or "dog-legged" the distribution at 1600 (the maximum of the SAT Verbal/Critical Reading + SAT Math). This did impact the scores of a very small percentage of students. After researching this further, we opted to lift the cap, starting in fall 2007.

Does the CLA correct for range restriction?

Range restriction is not necessary here because the institution is the unit of analysis and the CLA does not use a range-restricted population of institutions. CLA analysis is not concerned with the full range of SAT scores that exist; rather, it involves looking at the full range of SAT means of institutions. Summary statistics on the mean SAT for students sampled in the CLA are similar to national figures. Specifically, across 1,325 four-year institutions in the U.S., and looking at the estimated median SAT or ACT equivalent of freshman class, there is a minimum of 730, mean of 1069, maximum of 1510, and standard deviation of 126. Across CLA schools (fall 2006, n=118) for the same variable there is a minimum of 765, mean of 1067, maximum of 1486, and standard deviation of 133.

Source: College Results Online dataset, managed by the Education Trust, covers most 4-year Title IV-eligible higher-education institutions in the United States. Data were obtained with permission from the

Education Trust and constructed from IPEDS and other sources. For details see www.collegeresults.org/aboutthedata.aspx.

RELIABILITY

How reliable are CLA scores?

The reliability of CLA scores is assessed from multiple perspectives during each administration.

To evaluate the consistency of scorers, a random sample of approximately 10 percent of answers to each prompt are rescored by the grading team. Scorers are blinded to the fact that the answers were previously scored. Over a recent two-year period, the correlations between the first and second readings across the Performance Tasks have averaged .81, ranging from .76 to .87. These results are consistent with all previous CLA administrations. During the fall 2007 administration, new Analytic Writing Tasks were introduced: four Make-an-Argument and four Critique-an-Argument prompts. The average inter-reader correlations were .65 for the Make-an-Argument essays (ranging from .57 to .70), and .80 for the Critique-an-Argument essays (ranging from .77 to .84).

In addition to inter-reader consistency, the internal consistency reliability of the component items related to each of the scoring rubrics is calculated along with overall task scores making up the CLA. Calculations are carried out at both student and school levels (i.e., average across students within schools). Cronbach alpha coefficients, a measure of internal consistency, are estimated for each of the Performance Tasks and Analytic Writing Tasks, with the average of these coefficients used as an estimate for the overall Performance Task and Analytic Writing Task. The reliability of the school's total CLA score is estimated using the formula for estimating a test battery, here consisting of two tests (i.e., the school's average Performance Task and Analytic Writing Task scores). For the fall 2007 test administration, the average alpha coefficients for the Analytic Writing Tasks were .82 and .91 for individual and school-level means, respectively. For the Performance Tasks, these coefficients were .84 and .92 for individual and institution scores respectively. The estimated reliability of the school's total test score was .97. These results were also consistent with those found in past administrations.

VALIDITY

Do you have any evidence of construct validity?

CAE (CLA) is currently participating in a construct validity study in concert with ACT (CAAP) and ETS (MAPP) to investigate construct validity of the three measures.

What about the face validity of your measures?

We encourage you to see for yourself. To access a retired CLA Performance Task, please visit: <http://cla.cae.org>. When prompted, enter Session Number 27958-9319708.

You will see the same interface as do students; thus, you will be prompted to answer identifying information prior to beginning the assessment. Using the demonstration will allow you to view one Performance Task. Students typically take one of approximately seven different prompts from the available Performance Tasks.

Sample Analytic Writing Tasks (Make-an-Argument and Critique-an-Argument) are available on our website: http://www.cae.org/content/pro_collegiate_sample_measures.htm.

STUDENT EFFORT

We are concerned that students won't devote sufficient effort to the CLA and that our CLA institutional results will suffer as a result. Do you control for student effort?

The CLA does not control for self-reported student effort. Our investigations reveal that students differ in how hard they try and the reasons why they participate. However, self-reported level of effort and motivating factors are largely unrelated to CLA scores and they have an even a weaker relationship with CLA scores after incoming academic ability (i.e., ACT/SAT/SLE) is taken into account. This is especially true when we examine these relationships at the institutional level. When using the questionnaire based self-report by students of their level of effort, motivation accounts for only 2 to 4 percent of the variance in CLA results between schools using the cross-sectional design.

There are two larger, related points that need to be made regarding motivation. First, campuses need to think about how they integrate assessment into their institutional fabric such that participation in assessments like the CLA is part of their education. We believe that campuses should have mechanisms to feedback information to students, faculty, programs, and the institutions themselves. Short of this, motivation will be a problem no matter the incentives. In addition, higher education institutions do and should instill in their students the habit of devoting due effort to any task that they agree to undertake, whether that be taking a course, engaging in volunteer work, or participating in an assessment because their college asked them to do so. In other words, producing students that strive to perform well in all endeavors, academic and non-academic, is itself a goal every higher education institution has or should have. The degree of motivation that students bring to the CLA is inextricably linked to the norms of performance that their college expects – indeed, demands – of them. In that sense, student willingness to perform well on the CLA is one indicator of institutional impact.

OTHER FACTORS OF INTEREST

Are there differences in scores by sex? By racial/ethnic group?

Whether test-takers in one demographic subgroup (e.g., female or African American students) perform as well as other test-takers with similar ability levels in another subgroup (e.g., male or Hispanic/Latino students) is an important question because the presence of such differences might suggest the presence of subgroup bias. To answer this question, CLA researchers conducted a series of regression analyses on over 17,000 first-year students and over 10,900 seniors taking the CLA in fall 2005 and spring 2006 to determine whether student ability (as measured by SAT or ACT scores), race/ethnicity, gender or primary language spoken (other than English) were related to student scores after controlling for these students' SAT scores.

Results of 22 separate OLS regression models conclusively demonstrate that student demographics were unrelated to CLA scores after controlling for ability level. Further, the findings were markedly consistent for first-year students and seniors. A similar investigation was conducted with data from fall 2007, which confirmed that demographic differences account for less than one percent of the variance between CLA scores after controlling for incoming academic ability through SAT/ACT/SLE scores.

Is there an interaction between performance task “topic” and a student’s major?

We have looked for but have not found any interaction between task topic and student major. We plan to investigate this question again using data from spring 2008.

What is the relationship between CLA scores and time spent on CLA tasks?

There is a moderate positive correlation between CLA scores and time spent on CLA tasks as well as between SAT scores and time spent on CLA tasks. Most students need up to 90 minutes to fully address the Performance Task and up to 75 minutes to fully address the combination of the two Analytic Writing Tasks.

Where can we find additional technical information about the CLA?

For additional information, please read *The Collegiate Learning Assessment: Facts and Fantasies (2007)* and *An Approach to Measuring Cognitive Outcomes Among Higher Education Institutions (2005)*.

These articles are located on the CLA website at:

http://www.cae.org/content/pro_collegiate_reports_publications.htm.