Fall 2013 CLA+ Cross-Sectional Results | **Institutional Report**

California State University, Fresno

cla+

TABLE OF CONTENTS

Your Results

Appendices

SECTION 1: SUMMARY RESULTS, BY CLASS

Number of Students Tested, by Class

Freshmen: 102 Sophomores: N/A Juniors: N/A Seniors: .N/A

Summary CLA+ Results, by Class

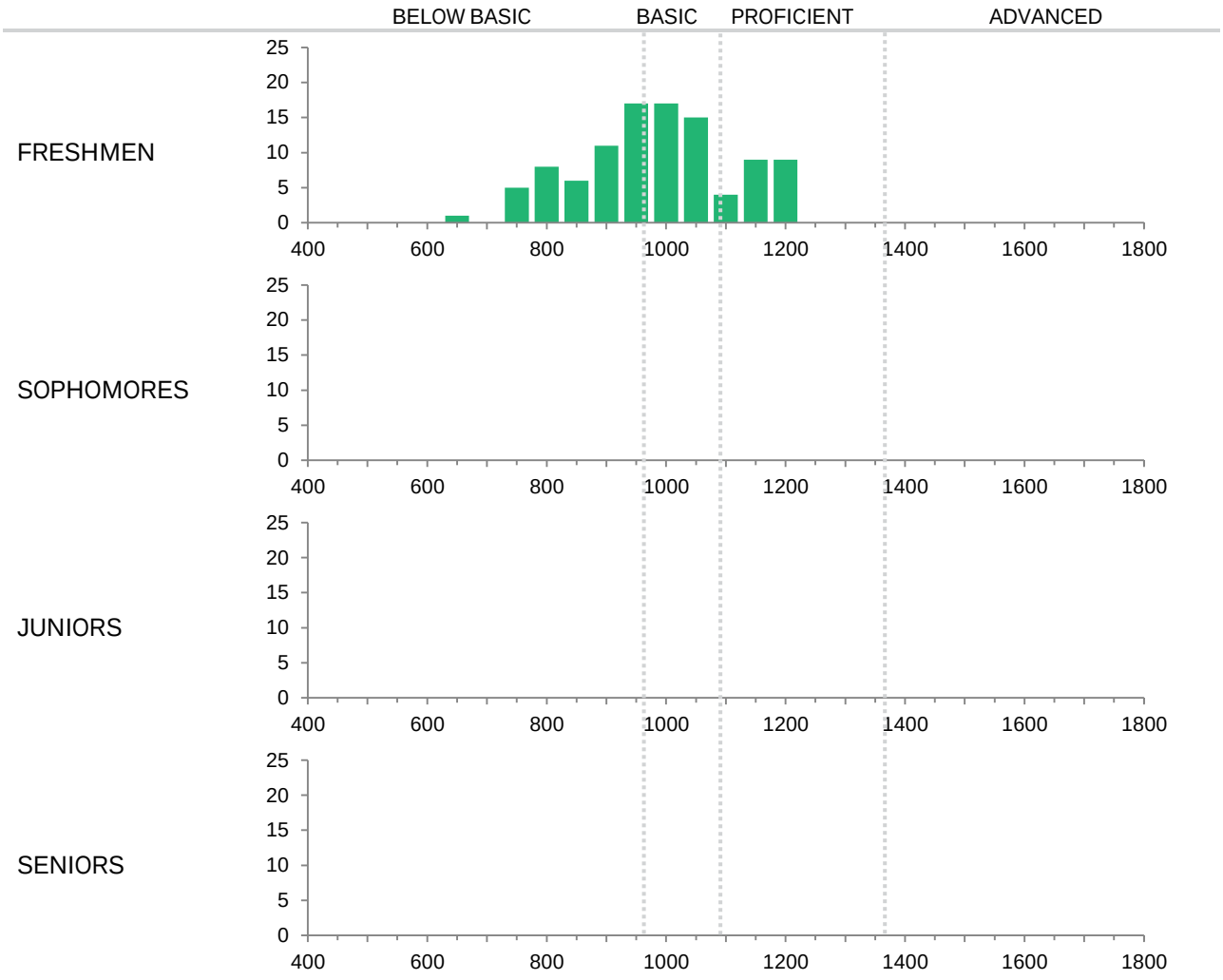
		MEAN SCORE	25TH PERCENTILE SCORE	75TH PERCENTILE SCORE	MEAN SCORE PERCENTILE RANK	EFFECT SIZE V. FRESHMEN
TOTAL CLA+ SCORE	Freshmen	1006	925	1081	47	--
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
PERFORMANCE TASK	Freshmen	1011	893	1125	45	--
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	1002	876	1136	42	--
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
ENTERING ACADEMIC ABILITY	Freshmen	891	770	1010	16	--
	Sophomores	N/A	N/A	N/A	N/A	--
	Juniors	N/A	N/A	N/A	N/A	--
	Seniors	N/A	N/A	N/A	N/A	--

California State University, Fresno has a senior Total CLA+ score of N/A and percentile rank of N/A.

The corresponding Mastery Level for this score is N/A.

SECTION 2: DISTRIBUTION OF MASTERY LEVELS

Distribution of CLA+ Scores, by Mastery Level



Mastery Levels, by Class

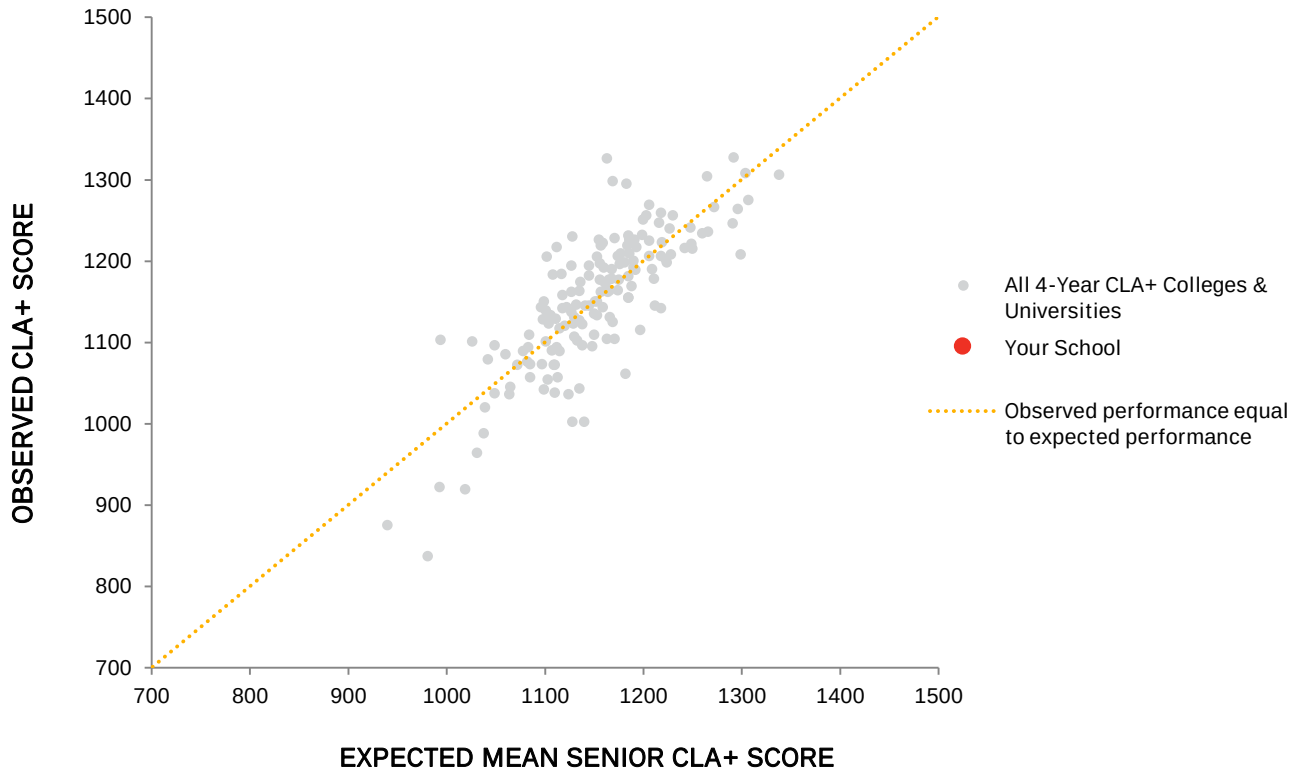
	MEAN TOTAL CLA+ SCORE	MEAN MASTERY LEVEL	PERCENT BELOW BASIC	PERCENT BASIC	PERCENT PROFICIENT	PERCENT ADVANCED
FRESHMEN	1006	Basic	35	42	23	0
SOPHOMORES	N/A	N/A	N/A	N/A	N/A	N/A
JUNIORS	N/A	N/A	N/A	N/A	N/A	N/A
SENIORS	N/A	N/A	N/A	N/A	N/A	N/A

SECTION 3: VALUE-ADDED ESTIMATES

	EXPECTED SENIOR MEAN CLA+ SCORE	ACTUAL SENIOR MEAN CLA+ SCORE
Total CLA+ Score	N/A	N/A
Performance Task	N/A	N/A
Selected-Response Questions	N/A	N/A

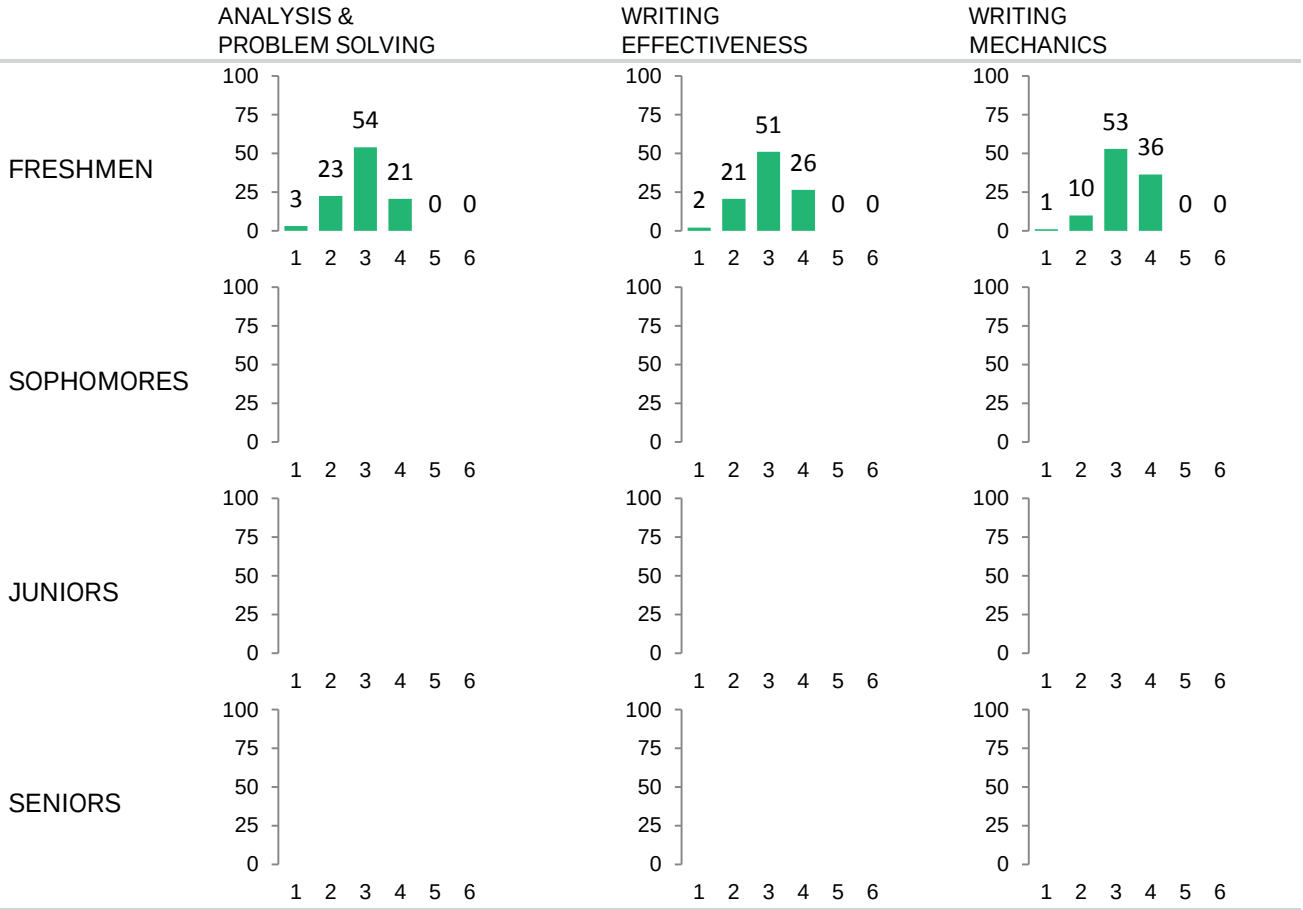
	VALUE-ADDED SCORE	PERFORMANCE LEVEL	PERCENTILE RANK	CONFIDENCE INTERVAL BOUNDS	
				LOWER	UPPER
Total CLA+ Score	N/A	N/A	N/A	N/A	N/A
Performance Task	N/A	N/A	N/A	N/A	N/A
Selected-Response Questions	N/A	N/A	N/A	N/A	N/A

Expected vs. Observed CLA+ Scores



SECTION 4: CLA+ SUBSCORES

Performance Task: Distribution of Subscores (in percentages)



NOTE: The Performance Task subscore categories are scored on a scale of 1 through 6.

Selected-Response Questions: Mean Subscores

	SCIENTIFIC & QUANTITATIVE REASONING			CRITICAL READING & EVALUATION			CRITIQUE AN ARGUMENT		
	Mean Score	25th Percentile Score	75th Percentile Score	Mean Score	25th Percentile Score	75th Percentile Score	Mean Score	25th Percentile Score	75th Percentile Score
FRESHMEN	477	411	535	489	413	558	487	388	596
SOPHOMORES	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JUNIORS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SENIORS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

NOTE: The selected-response section subscores are reported on a scale ranging approximately from 200 to 800.

SECTION 5: STUDENT EFFORT AND ENGAGEMENT

Student Effort and Engagement Survey Responses

How much effort did you put into the written-response task/ selected-response questions?

		NO EFFORT AT ALL	A LITTLE EFFORT	A MODERATE AMOUNT OF EFFORT	A LOT OF EFFORT	MY BEST EFFORT
PERFORMANCE TASK	Freshmen	0%	4%	37%	29%	29%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	0%	14%	38%	24%	25%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A

How engaging did you find the written-response task/ selected-response questions?

		NOT AT ALL ENGAGING	SLIGHTLY ENGAGING	MODERATELY ENGAGING	VERY ENGAGING	EXTREMELY ENGAGING
PERFORMANCE TASK	Freshmen	4%	15%	38%	35%	8%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	13%	20%	43%	22%	3%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A

SECTION 6: STUDENT SAMPLE SUMMARY

Student Sample Summary

DEMOGRAPHIC CHARACTERISTIC		FRESHMEN N	FRESHMEN %	SOPHOMORES N	SOPHOMORES %	JUNIORS N	JUNIORS %	SENIORS N	SENIORS %
TRANSFER	Transfer Students	--	--	N/A	N/A	N/A	N/A	N/A	N/A
	Non-Transfer Students	--	--	N/A	N/A	N/A	N/A	N/A	N/A
GENDER	Male	23	23%	N/A	N/A	N/A	N/A	N/A	N/A
	Female	78	76%	N/A	N/A	N/A	N/A	N/A	N/A
	Decline to State	1	1%	N/A	N/A	N/A	N/A	N/A	N/A
PRIMARY LANGUAGE	English	39	38%	N/A	N/A	N/A	N/A	N/A	N/A
	Other	63	62%	N/A	N/A	N/A	N/A	N/A	N/A
FIELD OF STUDY	Sciences & Engineering	24	24%	N/A	N/A	N/A	N/A	N/A	N/A
	Social Sciences	9	9%	N/A	N/A	N/A	N/A	N/A	N/A
	Humanities & Languages	12	12%	N/A	N/A	N/A	N/A	N/A	N/A
	Business	11	11%	N/A	N/A	N/A	N/A	N/A	N/A
	Helping / Services	28	27%	N/A	N/A	N/A	N/A	N/A	N/A
	Undecided / Other / N/A	18	18%	N/A	N/A	N/A	N/A	N/A	N/A
RACE/ ETHNICITY	American Indian / Alaska Native / Indigenous	1	1%	N/A	N/A	N/A	N/A	N/A	N/A
	Asian (including Indian subcontinent and Philippines)	16	16%	N/A	N/A	N/A	N/A	N/A	N/A
	Native Hawaiian or other Pacific Islander	0	0%	N/A	N/A	N/A	N/A	N/A	N/A
	African-American / Black (including African and Caribbean), non-Hispanic	7	7%	N/A	N/A	N/A	N/A	N/A	N/A
	Hispanic or Latino	63	62%	N/A	N/A	N/A	N/A	N/A	N/A
	White (including Middle Eastern), non-Hispanic	9	9%	N/A	N/A	N/A	N/A	N/A	N/A
	Other	3	3%	N/A	N/A	N/A	N/A	N/A	N/A
	Decline to State	3	3%	N/A	N/A	N/A	N/A	N/A	N/A
PARENT EDUCATION	Less than High School	44	43%	N/A	N/A	N/A	N/A	N/A	N/A
	High School	25	25%	N/A	N/A	N/A	N/A	N/A	N/A
	Some College	20	20%	N/A	N/A	N/A	N/A	N/A	N/A
	Bachelor's Degree	5	5%	N/A	N/A	N/A	N/A	N/A	N/A
	Graduate or Post-Graduate Degree	8	8%	N/A	N/A	N/A	N/A	N/A	N/A

APPENDIX A: INTRODUCTION TO CLA+

INTRODUCTION TO CLA+

The Collegiate Learning Assessment (CLA) was introduced in 2002 as a major initiative of the Council for Aid to Education (CAE). In the decade since its launch, the CLA has offered a value-added, constructed-response approach to the assessment of higher-order skills, such as critical thinking and written communication. Hundreds of institutions and hundreds of thousands of students have participated in the CLA to date.

Initially, the CLA focused primarily on providing institutions with estimates of their contributions, or value added, to their students' development of higher-order skills. As such, the institution—not the student—was the primary unit of analysis.

In 2013, CAE introduced an enhanced version of the CLA—CLA+—that provides utility and reliability at the individual student, as well as at the institutional, level. CLA+ also includes new subscores for quantitative and scientific reasoning, critical reading and evaluation, and critiquing an argument. New Mastery Levels provide criterion-referenced results that indicate the level of proficiency attained by an institution's students on the higher-order skills measured by CLA+.

When taking CLA+, students complete both a Performance Task (PT) and a series of Selected-Response Questions (SRQs).

The **Performance Task** presents a real-world situation in which an issue, problem, or conflict is identified. Students are asked to assume a relevant role to address the issue, suggest a solution, or recommend a course of action based on the information provided in a Document Library. A full CLA+ Performance Task contains four to nine documents in the library, and students have 60 minutes to complete the task. The Document Library contains a variety of reference sources such as technical reports, data tables, newspaper articles, office memoranda, and emails.

In the **Selected-Response Questions** section, students respond to 25 questions: 10 assess scientific and quantitative reasoning; 10 assess critical reading and evaluation; and five assess the students' ability to critique an argument by detecting logical flaws and questionable assumptions in a given argument. Students have 30 minutes to complete this section. Much like the Performance Task, each set of questions is document-based and requires that students draw information from the accompanying documents.

CLA+ is intended to assist faculty, school administrators, and others interested in programmatic change to improve teaching and learning, particularly with respect to strengthening higher-order skills.

Additionally, CLA+ results allow for direct, formative feedback to students. Faculty may also decide to use students' CLA+ results to make individualized decisions about grading, scholarships, admission, or placement, and students may choose to share their results with potential employers or graduate schools as evidence of the skills they have acquired at their college or university. Institutions may also wish to use CLA+ results to provide independent corroboration of competency-based learning, or to recognize individual students who exhibit the higher-order skills required for twenty-first century careers.

CLA+ helps institutions follow a continuous improvement model that positions faculty as central actors in the link between assessment and the teaching and learning process. While no single test can serve as the benchmark for all student learning in higher education, there are certain skills deemed important by most faculty and administrators across virtually all institutions; indeed the higher-order skills that CLA+ measures fall into this category.

CLA+ is significant because institutions need to have a frame of reference for where they stand and how much progress their students have made relative to the progress of students at other colleges. Yet, CLA+ is not about ranking institutions. Rather, it is about highlighting differences between them that can lead to improvements. Similarly, CLA+ is not about ranking students, but highlighting areas where individual students have excelled or may need to focus more efforts. CLA+ is an instrument designed to contribute directly to the improvement of teaching and learning. In this respect, it is in a league of its own.

APPENDIX B: METHODS

CLA+ METHODOLOGY

CLA+ uses innovative tasks and question sets to evaluate students' performance reflecting the following higher-order skills: analysis and problem solving, writing effectiveness, and writing mechanics on the PTs; and scientific and quantitative reasoning, critical reading and evaluation, and detecting logical flaws and questionable assumptions to critique arguments on the SRQs.

CLA+ measures these skills by giving students one PT and a set of 25 SRQs. Students have 90 minutes to complete the assessment—60 minutes for the PT and 30 minutes for the SRQs.

Results are provided to institutions after they have completed testing in each window. Your institutional report presents information on each section of CLA+ and total CLA+ performance for all freshmen that test in the fall window, and all sophomores, juniors, or seniors that test in the spring window. This includes a PT score, a SRQ score, and a Total CLA+ score. The PT and SRQ scores represent the average performance of your students that completed the respective sections. Total CLA+ scores are equal to the average of the PT and SRQ scores.

Performance Task scores are equal to the sum of the three PT subscore categories—Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics—converted to a common scale. Selected-Response Question scores are equal to the sum of the three SRQ raw subscores—Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument—also converted to a common scale. For more information about the scaling process, please see the *Scaling Procedures* section of this report (Appendix J).

The information presented in your results includes means (averages) and 25th and 75th percentile scores (the score values between which half of your students scored on CLA+), and a percentile ranking for your mean score. Note that percentile rankings are compared to other institutions testing the same class level in the same window; these statistics may not be available, depending on the sample of institutions that have tested accordingly.

CAE reports also include growth estimates for those class levels tested. These growth estimates are provided in two forms: effect sizes and value-added scores.

Effect sizes represent the amount of growth seen from freshman year, in standard deviation units. They are calculated by subtracting the mean freshman performance at your school from the mean of your sophomore, junior, or senior performance, and dividing by the standard deviation of your freshman scores. Effect sizes do not take into account the performance of students at other CLA+ institutions.

Value-added scores, on the other hand, are used to estimate growth from freshman to senior year, relative to that seen across institutions. Value-added modeling is often viewed as an equitable way of estimating an institution's contribution to learning. Simply comparing average achievement of all schools tends to paint selective institutions in a favorable light and discount the educational efficacy of schools admitting students from weaker academic backgrounds. Value-added modeling addresses this issue by providing scores that can be interpreted as relative to institutions testing students of similar entering academic ability. This allows all schools, not just selective ones, to demonstrate their relative educational efficiency.

CLA+ value-added estimation approach employs a statistical technique known as hierarchical linear modeling (HLM). Under this methodology, a school's value-added score indicates the degree to which the observed senior mean CLA+ score meets, exceeds, or falls below expectations established by (1) seniors' Entering Academic Ability (EAA)[1] scores, and (2) the mean CLA+ performance of freshmen at that school, which serves as a control for selection effects not covered by EAA. Only students with EAA scores are included in institutional analyses.

When the average performance of seniors at a school is substantially better than expected, this school is said to have high "value added." To illustrate, consider several schools admitting students with similar average performance on general academic ability tests (e.g., the SAT or ACT) and on tests of higher-order skills (e.g., CLA+). If, after four years of college education, the seniors at one school perform

[1] Combined SAT Math and Critical Reading, ACT Composite, or Scholastic Level Exam (SLE) scores on the SAT Math + Critical Reading scale. Hereinafter referred to as Entering Academic Ability (EAA).

better on CLA+ than is typical for schools admitting similar students, one can infer that greater gains in critical thinking and writing skills occurred at the highest performing school. Note that a low (negative) value-added score does not necessarily indicate that no gain occurred between freshman and senior year; however, it does suggest that the gain was lower than would typically be observed at schools testing students of similar EAA. Value-added scores are placed on a standardized (*z*-score) scale and assigned performance levels. Schools that fall between -1.00 and +1.00 are classified as "near expected," between +1.00 and +2.00 are "above expected", between -1.00 and -2.00 are "below expected," above +2.00 are "well above expected," and below -2.00 are "well below expected." Value-added estimates are also accompanied by confidence intervals, which provide information on the precision of the estimates; narrow confidence intervals indicate that the estimate is more precise, while wider intervals indicate less precision.

In the past, CLA+ value-added models were recalculated after each academic year, allowing for the potential of fluctuation in results due to the sample of participating institutions rather than changes in actual growth within a college or university. The introduction of CLA+ also marks the first time that the value-added equation parameters will be fixed, which will facilitate reliable year-to-year comparisons of value-added scores.

Our analyses include results from all CLA+ institutions, regardless of sample size and sampling strategy. Therefore, we encourage you to apply due caution when interpreting your results if you tested a very small sample of students or believe that the students in your institution's sample are not representative of the larger student body.

APPENDIX C: EXPLANATION OF YOUR RESULTS

The following section provides guidance on interpreting institutional results. For all tables provided in your cross-sectional results, the sample of students reported here include freshmen who have tested in the fall window, and sophomores, juniors, and seniors who have tested in the spring window. To ensure that the results provided across the tables in your report use a consistent sample—in addition to testing in the appropriate window for a given class level—students also need to have (1) completed all sections of the assessment (the Performance Task, Selected-Response Questions, and the accompanying survey), they must (2) have a SAT, ACT, or SLE score submitted to CAE, and (3) not have otherwise been designated for exclusion from institutional analyses during the registrar data submission process.

Cross-CLA+ summary data are provided in the following section, *Results Across CLA+ Institutions* (Appendix D), for comparative purposes. The institutions included in that section—also used to determine your percentile rankings, and set the value-added model parameters—are described in *Institutional Sample* section of this report (Appendix E).

In addition to the details presented here, CAE also offers a series of results overview videos to guide institutions through interpreting and making use of their results. These videos will be available for CLA+ in March 2014, on our website at www.cae.org/cla-institutional-reporting.

SUMMARY RESULTS, BY CLASS (Section 1, page 2)

The first table in Section 1 of this report provides the **Number of Students Tested, by Class**. This includes the number of freshmen that were tested in the fall window and the number of sophomores, juniors, and seniors that were tested in the spring CLA+ window this academic year. These numbers indicate the sample size for each ensuing table or figure in your report. Please note that very small samples (e.g., fewer than 100 for any given class) should be interpreted with caution, as smaller sample sizes are less likely to provide reliable or representative results.

This table is followed by summary statistics for the students in your sample. For any class levels not tested or where results are not applicable, values of "N/A" are reported.

The **Summary CLA+ Results, by Class** table provides mean scores, quartiles, percentile ranks, and effect sizes for each class level tested and for each section of the test, as well as summary of your sample's EAA.

The Mean Score column represents the average score of the students included in the sample. This is also considered your institutional CLA+ score.

The 25th Percentile Score indicates the score value at or below which 25 percent of your students scored, and the 75th Percentile Score indicates the score value at or below which 75 percent of your students scored. Accordingly, half (50%) of the students in your sample scored between the 25th and 75th percentile scores shown in the table.

The Mean Score Percentile Rank indicates how well your institution performed relative to other institutions across CLA+. The values in this column represent the percentage of institutions whose mean scores were lower than yours. If there is an insufficient sample of institutions testing at a corresponding class level, you will see the value "N/A" in the relevant cell of the table.

The final—Effect Size v. Freshmen—column in this table presents growth estimates in the form of school-specific effect sizes. Effect sizes indicate the standardized difference in CLA+ scores between entering students and those at each subsequent class level, using your school's standard deviation of entering students. An effect size of 0 indicates no difference between entering and exiting students, while positive effect sizes indicate scores that are higher than those of entering students, with larger effect sizes corresponding to larger score differences.

For a summary of institutional performance across CLA+, please refer to the *Results Across CLA+ Institutions* section of this report (Appendix D).

DISTRIBUTION OF MASTERY LEVELS (Section 2, page 3)

Section 2 of your institutional report focuses on Mastery Levels, which are new, criterion-referenced indicators of performance on CLA+. Mastery Levels are determined by an individual's Total CLA+ score on the student level, and by a sample's mean Total CLA+ score on the institutional level.

There are four Mastery Level categories for CLA+: Below Basic, Basic, Proficient, and Advanced. These categories, and the process through which the Mastery Levels were derived, are described in detail in the *Mastery Levels* section of your report (Appendix H).

There are two tables in your results that address your students' performance in terms of Mastery Levels. The first, **Distribution of CLA+ Scores, by Mastery Level**, includes a histogram of Total CLA+ scores for each class level that you tested, overlaid with the Mastery Level score cut points to show how the distribution of CLA+ scores within your sample(s) corresponds to students' Mastery of the skills measured by CLA+.

The second table presents a summary of **Mastery Levels, by Class**. The first column of data lists the mean Total CLA+ score for each class level tested, followed by the corresponding Mastery Level—the level at which the average student within your sample performed. The next four columns present the percentage of students that performed at each Mastery Level within each class your institution tested.

VALUE-ADDED ESTIMATES (Section 3, page 4)

Section 3 of your institutional report presents estimates of the growth shown by your students from freshman to senior year, in the form of **Value-Added Estimates**. Note that all tables in this section will read "N/A" in the fall 2013 CLA+ administration—at which point only freshmen have been tested—and in cases where schools test classes other than freshmen and seniors.

The first table provides your students' Expected Senior Mean CLA+ Score alongside their Actual Senior Mean CLA+ Score. Expected scores are determined by the typical performance of seniors at institutions testing similar samples of students, given their seniors' EAA and their mean freshman performance on CLA+.

The following table presents your value-added results. Your Value-Added Score represents the difference between an institution's Actual Senior Mean CLA+ Score and its Expected Senior Mean CLA+ score, converted to standard deviation units.

The value-added score for each section of CLA+ is accompanied by a Performance Level, which is determined by the specific value-added score received. Schools that fall between -1.00 and +1.00 are classified as "near expected," between +1.00 and +2.00 are "above expected", between -1.00 and -2.00 are "below expected," above +2.00 are "well above expected," and below -2.00 are "well below expected."

In addition to Performance Levels, each value-added score is assigned a percentile rank. The percentile rank tells an institution the percentage of other institutions whose value-added scores would fall below its own value-added scores, if all the scores were ranked in order of their values.

Value-added estimates are also accompanied by confidence intervals, which provide information on the precision of the estimates; narrow confidence intervals indicate that the estimate is more precise, while wider intervals indicate less precision. Given the inherent uncertainty of value-added estimates, value-added scores should be interpreted in light of available information about their precision. HLM estimation—the method used by CAE for calculating value-added scores—provides standard errors for value-added scores, which can be used to compute a unique 95% confidence interval for each school. These standard errors reflect within- and between-school variation in CLA+ and EAA scores, and they are most strongly related to senior sample size. Schools testing larger samples of seniors obtain more precise estimates of value added and therefore have smaller standard errors and corresponding 95% confidence intervals.

The final component of your value-added results is the scatterplot of Expected vs. Observed CLA+ scores. This figure shows the performance of all four-year colleges and universities, relative to their expected performance as predicted by the value-added model. The vertical distance from the diagonal line indicates the value added by the institution; institutions falling above the diagonal line are those that add more value than expected based on the model. The gold diagonal line represents the points at which observed and expected senior scores are equal. After testing seniors in spring 2014, your institution will appear in red.

More details about CLA+ value-added methodology—including model parameters, guidance on interpreting confidence intervals, and instructions for using your data file to calculate value-added estimates for subgroups of students—are included in the *Modeling Details* section of this report (Appendix K).

CLA+ SUBSCORES (Section 4, page 5)

Each section of CLA+ is scored according to multiple skill-based categories. The three subscores for the PT are: Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. The three subscores for the SRQs are: Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument.

The first table in Section 4, **Performance Task: Distribution of Subscores**, presents the distribution of subscores for the three subscore categories. Subscore categories are scored values ranging from 1 through 6, which each score value corresponding specific response characteristics (see Appendix G: *Scoring CLA+* more information about the scoring rubric). The values in the graphs represent the percentage of students at your institution that performed at each score level.

The second table in Section 4, **Selected-Response Questions: Mean Subscores**, presents summary statistics for the three SRQ subscore categories. Scores in this section of CLA+ are determined by the number of correct responses in the skill set, adjusted for the difficulty of the group of questions asked. Each section subscore is reported in a subscale of approximately 200 to 800.

Mean Scores in this table show the average score received for each class level in the given subscore category. The 25th Percentile Scores indicate the score values at or below which 25 percent of your students scored, and the 75th Percentile Scores indicate the score values at or below which 75 percent of your students scored. Accordingly, half (50%) of the students in your sample scored between the 25th and 75th percentile scores shown in the table.

STUDENT EFFORT AND ENGAGEMENT (Section 5, page 6)

To allow institutions to determine the role of students' effort and engagement in their performance on CLA+, CAE has introduced a set of survey questions to the end of the assessment. These questions ask students how much effort they have put into the written-response (PT) and selected-response (SRQ) sections of CLA+, as well as how engaging they found each section of the assessment.

Answer options are provided on a likert scale, ranging from "No effort at all" to "My best effort" for the effort questions, and from "Not at all engaging" to "Extremely engaging" for the engagement questions.

The **Student Effort and Engagement Survey Responses** table provides the percentage of students at each class level who gave each answer option in the survey.

In addition to providing insight into the effort and motivation levels of an institution's students, these results can help identify cases in which an institution might want to enhance its recruitment efforts to boost motivation. Comparisons to the distribution of survey responses across all schools (see Appendix D: Results Across CLA+ Institutions) allow schools to see the degree to which their students are motivated and engaged relative to others.

STUDENT SAMPLE SUMMARY (Section 6, page 7)

The final section of your CLA+ results is the **Student Sample Summary**, which provides the count and percentage of students within your sample who meet various characteristics. The characteristics reported include: transfer status (reported by participating institutions during the registrar data collection process), gender, primary language, field of study, race or ethnicity, and parental education level. All demographic characteristics are provided by students in the post-assessment survey.

APPENDIX D: RESULTS ACROSS CLA+ INSTITUTIONS

SECTION D1: SUMMARY RESULTS, BY CLASS

Number of Participating Institutions, by Class

Freshmen: 169 Sophomores: N/A Juniors: N/A Seniors: N/A

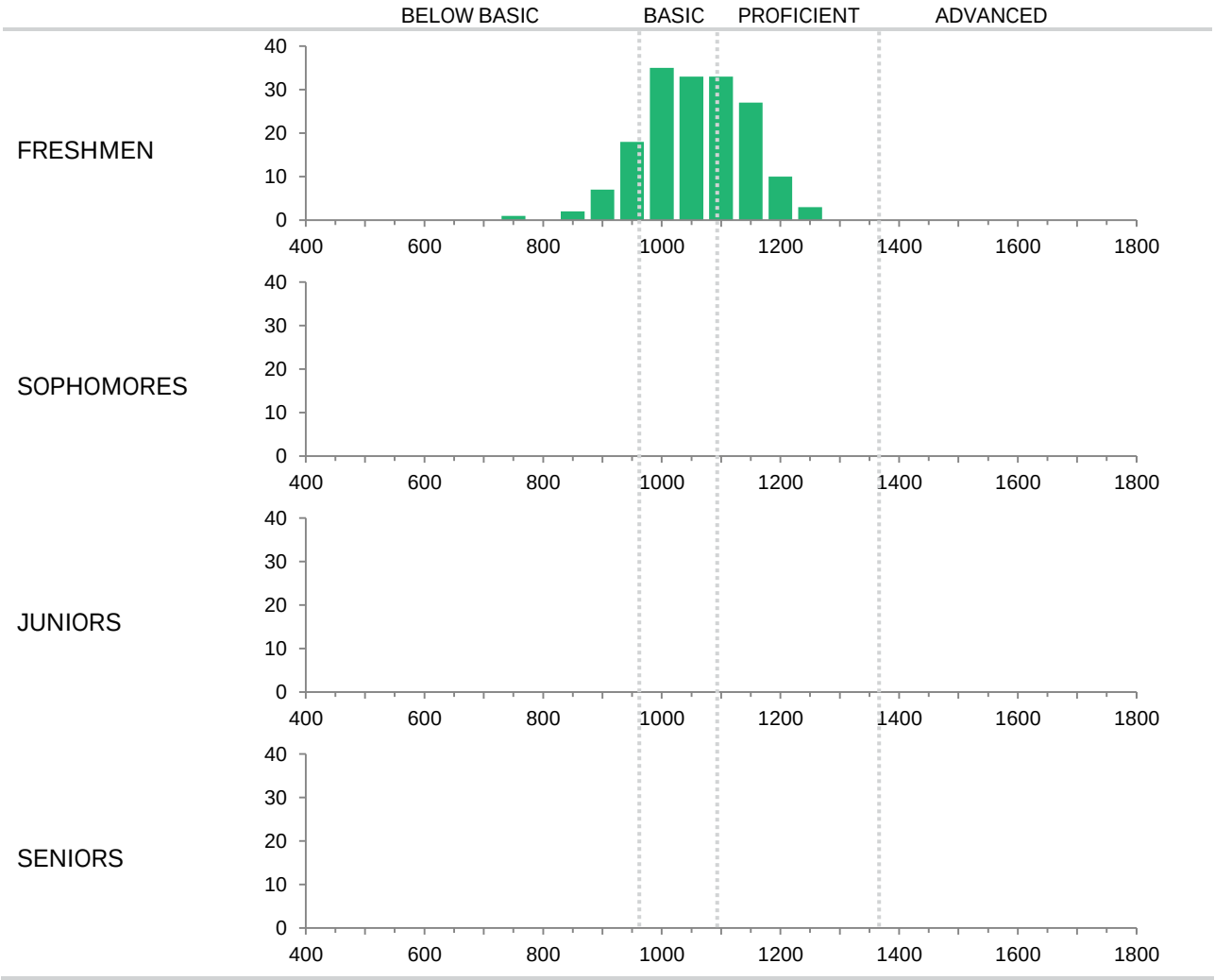
Summary of CLA+ Results Across Institutions, by Class

		MEAN SCORE	25TH PERCENTILE SCORE	75TH PERCENTILE SCORE	MEAN EFFECT SIZE V. FRESHMEN
TOTAL CLA+ SCORE	Freshmen	1032	974	1096	--
	Sophomores	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A
PERFORMANCE TASK	Freshmen	1028	967	1089	--
	Sophomores	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	1036	974	1089	--
	Sophomores	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A
ENTERING ACADEMIC ABILITY	Freshmen	1022	948	1106	--
	Sophomores	N/A	N/A	N/A	--
	Juniors	N/A	N/A	N/A	--
	Seniors	N/A	N/A	N/A	--

The average CLA+ institution has a senior Total CLA+ score of N/A, and a corresponding Mastery Level of N/A.

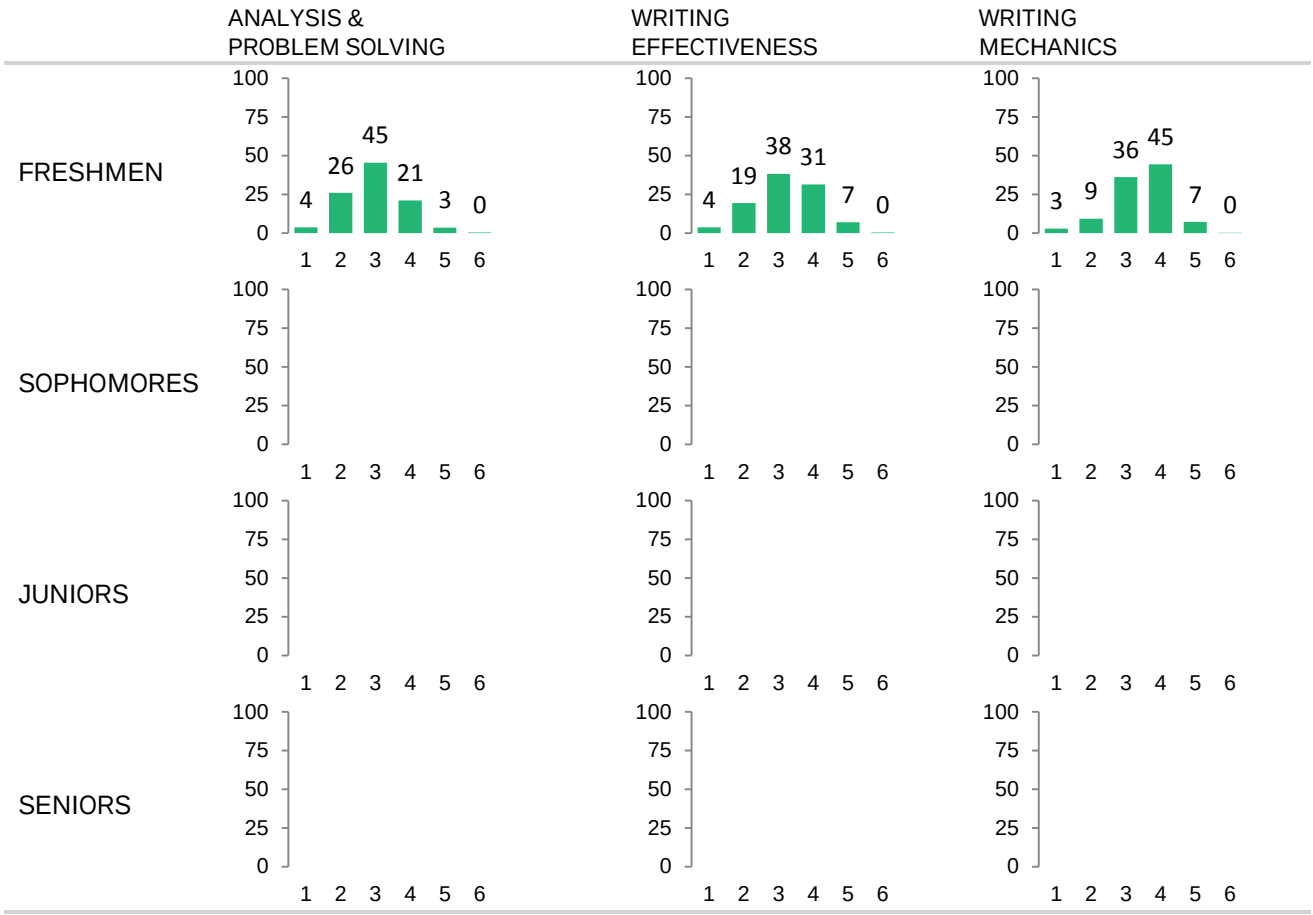
SECTION D.2: DISTRIBUTION OF MASTERY LEVELS ACROSS INSTITUTIONS

Distribution of Mean CLA+ Scores, by Mastery Level



SECTION D4: CLA+ SUBSCORES ACROSS INSTITUTIONS

Performance Task: Mean Distribution of Subscores (in percentages)



NOTE: The Performance Task subscore categories are scored on a scale of 1 through 6.

Selected-Response Questions: Mean Subscores Across Institutions

	SCIENTIFIC & QUANTITATIVE REASONING			CRITICAL READING & EVALUATION			CRITIQUE AN ARGUMENT		
	Mean Score	25th Percentile Score	75th Percentile Score	Mean Score	25th Percentile Score	75th Percentile Score	Mean Score	25th Percentile Score	75th Percentile Score
FRESHMEN	499	473	519	498	476	520	498	471	524
SOPHOMORES	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JUNIORS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SENIORS	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

NOTE: The selected-response section subscores are reported on a scale ranging approximately from 200 to 800.

SECTION D5: STUDENT EFFORT AND ENGAGEMENT ACROSS CLA+ INSTITUTIONS

Mean Student Effort and Engagement Survey Responses

How much effort did you put into the written-response task/ selected-response questions?

		NO EFFORT AT ALL	A LITTLE EFFORT	A MODERATE AMOUNT OF EFFORT	A LOT OF EFFORT	MY BEST EFFORT
PERFORMANCE TASK	Freshmen	1%	5%	35%	35%	24%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	2%	14%	42%	28%	14%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A

How engaging did you find the written-response task/ selected-response questions?

		NOT AT ALL ENGAGING	SLIGHTLY ENGAGING	MODERATELY ENGAGING	VERY ENGAGING	EXTREMELY ENGAGING
PERFORMANCE TASK	Freshmen	7%	17%	42%	28%	6%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A
SELECTED-RESPONSE QUESTIONS	Freshmen	15%	27%	38%	17%	3%
	Sophomores	N/A	N/A	N/A	N/A	N/A
	Juniors	N/A	N/A	N/A	N/A	N/A
	Seniors	N/A	N/A	N/A	N/A	N/A

SECTION D6: STUDENT SAMPLE SUMMARY ACROSS CLA+

Student Sample Summary Across CLA+ Institutions

DEMOGRAPHIC CHARACTERISTIC		FRESHMEN Mean %	SOPHOMORES Mean %	JUNIORS Mean %	SENIORS Mean %
TRANSFER	Transfer Students	--	N/A	N/A	N/A
	Non-Transfer Students	--	N/A	N/A	N/A
GENDER	Male	39%	N/A	N/A	N/A
	Female	60%	N/A	N/A	N/A
	Decline to State	2%	N/A	N/A	N/A
PRIMARY LANGUAGE	English	80%	N/A	N/A	N/A
	Other	20%	N/A	N/A	N/A
FIELD OF STUDY	Sciences & Engineering	26%	N/A	N/A	N/A
	Social Sciences	10%	N/A	N/A	N/A
	Humanities & Languages	11%	N/A	N/A	N/A
	Business	14%	N/A	N/A	N/A
	Helping / Services	26%	N/A	N/A	N/A
	Undecided / Other / N/A	14%	N/A	N/A	N/A
RACE/ ETHNICITY	American Indian / Alaska Native / Indigenous	1%	N/A	N/A	N/A
	Asian (including Indian subcontinent and Philippines)	8%	N/A	N/A	N/A
	Native Hawaiian or other Pacific Islander	1%	N/A	N/A	N/A
	African-American / Black (including African and Caribbean), non-Hispanic	14%	N/A	N/A	N/A
	Hispanic or Latino	19%	N/A	N/A	N/A
	White (including Middle Eastern), non-Hispanic	50%	N/A	N/A	N/A
	Other	4%	N/A	N/A	N/A
	Decline to State	4%	N/A	N/A	N/A
PARENT EDUCATION	Less than High School	8%	N/A	N/A	N/A
	High School	24%	N/A	N/A	N/A
	Some College	24%	N/A	N/A	N/A
	Bachelor's Degree	27%	N/A	N/A	N/A
	Graduate or Post-Graduate Degree	18%	N/A	N/A	N/A

APPENDIX E: INSTITUTIONAL SAMPLE

The CLA+ sample of institutions is comprised of all institutions that have tested freshmen in fall 2013 or sophomores, juniors, or seniors in spring 2014. Because spring 2014 testing is currently underway, data for non-freshmen will not be available until early summer 2014. Unlike with the previous incarnation of the assessment, the CLA+ sample remains fixed from year to year. By using a fixed sample of institutions for national comparisons, institutions can more easily track their own progress from year to year, without questions of whether changes in percentile rankings for an individual institution are due to true changes in performance or simply reflective of differences in the comparative sample.

To ensure national representativeness, CAE will continue to assess the sample of institutions and–if there are significant changes—update the institutional sample as needed.

SAMPLE REPRESENTATIVENESS

CLA+-participating institutions appear to be generally representative of their classmates with respect to entering ability levels as measured by Entering Academic Ability (EAA) scores.

Specifically, across institutions, the average EAA score of CLA+ freshmen was only seven points higher than that of the entire freshman class (1038 versus 1031, over n=123 institutions), and the correlation between the average EAA score of CLA+ freshmen and their classmates was high (r=0.93).

These data suggest that, as a group, students tested as part of the CLA+ institutional sample are similar to all students at the schools that make up the sample of CLA+ institutions. This correspondence increases confidence in the inferences that can be made from the results with the samples of students that were tested at a school to all the students at that institution.

CARNEGIE CLASSIFICATION

The following table shows CLA+ schools grouped by Basic Carnegie Classification. The spread of schools corresponds fairly well with that of the 1,587 four-year not-for-profit institutions across the nation.

Note that counts in this table exclude some institutions that do not fall into these categories, such as Special Focus Institutions and institutions based outside of the United States.

Carnegie Classification of CLA+ Institutional Sample

	NATION (N=1,683)		CLA+ (N=144)	
CARNEGIE CLASSIFICATION	N	%	N	%
DOCTORATE-GRANTING UNIVERSITIES	283	17	22	15
MASTER'S COLLEGES AND UNIVERSITIES	651	39	78	54
BACCALAUREATE COLLEGES	749	45	44	31

Source: Carnegie Foundation for the Advancement of Teaching, Carnegie Classifications Data File, January 16, 2014.

SCHOOL CHARACTERISTICS

The following table provides statistics on some important characteristics of colleges and universities across the nation compared with CLA+ schools.

These statistics suggest that CLA+ schools are fairly representative of four-year, not-for-profit institutions nationally. Percentage public and undergraduate student body size are exceptions.

School Characteristics of CLA+ Institutional Sample

SCHOOL CHARACTERISTIC	NATION	CLA+
PERCENTAGE PUBLIC	30	56
PERCENTAGE HISTORICALLY BLACK COLLEGE OR UNIVERSITY (HBCU)	4	4
MEAN PERCENTAGE OF UNDERGRADUATES RECEIVING PELL GRANTS	31	30
MEAN SIX-YEAR GRADUATION RATE	51	48
MEAN BARRON'S SELECTIVITY RATING	3.6	3.1
MEAN ESTIMATED MEDIAN SAT SCORE	1058	1027
MEAN NUMBER OF FTE UNDERGRADUATE STUDENTS (ROUNDED)	3,869	7,296
MEAN STUDENT-RELATED EXPENDITURES PER FTE STUDENT (ROUNDED)	$12,330	$10,497

Sources: College Results Online dataset, managed by and obtained with permission from the Education Trust, covers most four -year Title IV-eligible higher-education institutions in the United States. Data were constructed from IPEDS and other sources. Because all schools did not report on every measure in the table, the averages and percentages may be based on slightly different denominators. Data also come from the Carnegie Foundation for the Advancement of Teaching, Carnegie Classifications Data File, January 16, 2014.

CLA+ INSTITUTIONS

The institutions listed here, in alphabetical order, comprise the sample of institutions testing freshmen. To view a list of current participating institutions, please visit www.cae.org/claparticipants.

CLA+ Schools

Alaska Pacific University
Antelope Valley College
Appalachian State University
Augsburg College
Augustana College (SD)
Aurora University
Barton College
Bellarmine University
Bob Jones University
Bowling Green State University
Brigham Young University - Idaho
California Maritime Academy
California Polytechnic State University San Luis Obispo
California State Polytechnic University, Pomona
California State University, Bakersfield
California State University, Channel Islands
California State University, Chico
California State University, Dominguez Hills
California State University, East Bay
California State University, Fresno
California State University, Fullerton
California State University, Long Beach
California State University, Los Angeles
California State University, Monterey Bay
California State University, Monterey Bay, Computer Science and Information Technology
California State University, Northridge
California State University, Sacramento
California State University, San Bernardino
California State University, San Marcos
California State University, Stanislaus
Centenary College of Louisiana

Clarke University
College of Saint Benedict/St. John's University
Collin College
Colorado Christian University
Concord University
Concordia College
Culver-Stockton College
CUNY - Baruch College
CUNY - Borough of Manhattan Community College
CUNY - Bronx Community College
CUNY - Brooklyn College
CUNY - College of Staten Island
CUNY - Hostos Community College
CUNY - Hunter College
CUNY - John Jay College of Criminal Justice
CUNY - Kingsborough Community College
CUNY - LaGuardia Community College
CUNY - Lehman College
CUNY - Medgar Evers College
CUNY - New York City College of Technology
CUNY - Queens College
CUNY - Queensborough Community College
CUNY - The City College of New York
CUNY - York College
Dillard University
Drexel University, Department of Architecture and Interiors
Earlham College
East Carolina University
Eastern Connecticut State University
Emory & Henry College
Fayetteville State University
Flagler College
Florida International University Honors College
Frostburg State University
Georgia College & State University
Great Basin College
Hardin-Simmons University
Hastings College
Hong Kong Polytechnic University
Howard Community College
Humboldt State University
Illinois College
Indiana University of Pennsylvania
Jacksonville State University
Keene State College
Kent State University
Kepler Kigali
Kepler Kigali, Control
Keuka College
LaGrange College
Lewis University
Lynchburg College
Marshall University
Miami University - Oxford
Miles College
Minneapolis College of Art and Design
Minnesota State Community & Technical College
Mississippi University for Women
Monmouth University
Montclair State University
Morgan State University
National Louis University
Nevada State College
New York University Abu Dhabi
Newberry College
Nicholls State University
North Dakota State University
Nyack College
Ohio Wesleyan University
Our Lady of the Lake
Pittsburg State University
Plymouth State University
Presbyterian College
Purchase College
Queen's University
Quest University
Ramapo College of New Jersey
Robert Morris University
Roger Williams University
Saginaw Valley State University
San Diego State University
San Francisco State University
San Jose State University
Schreiner University
Shepherd University
Sonoma State University
Southern Connecticut State University
Southern Virginia University
Southwestern University
St. Ambrose University
St. John Fisher College
Stetson University
Stonehill College
SUNY Cortland
Texas A&M International University
Texas A&M University-Texarkana
Texas State University - San Marcos
Texas Tech University
The Citadel
The College of Idaho
The Ohio State University
The Sage Colleges
Truckee Meadows Community College
Truman State University
University of Bridgeport
University of Evansville
University of Great Falls
University of Hawaii at Hilo, College of Business and Economics
University of Houston
University of Jamestown
University of Louisiana - Lafayette
University of Missouri - St. Louis
University of New Mexico
University of North Carolina Pembroke
University of North Dakota
University of Saint Mary

University of Texas - Pan American
University of Texas at Arlington
University of Texas at Austin
University of Texas at Dallas
University of Texas at El Paso
University of Texas at San Antonio
University of Texas at Tyler
University of Texas of the Permian Basin
Ursuline College
Warner University
Weber State University
West Chester University
Western Carolina University
Western Governors University
Western Kentucky University
Western Michigan University
Western Nevada College
Westminster College (MO)
Westminster College (UT)
Wichita State University
Wichita State University, School of Engineering
Wiley College
William Peace University
William Woods University
Winston-Salem State University
Wisconsin Lutheran College
Yakima Valley Community College

APPENDIX F: CLA+ TASKS

INTRODUCTION TO CLA+ TASKS AND SELECTED-RESPONSE QUESTIONS

CLA+ consists of a Performance Task (PT) and a set of Selected-Response Questions (SRQs). All CLA+ exams are administered online.

The PTs consist of open-ended prompts that require constructed responses. SRQs are presented in three sets, each focusing on a different skill area. Students choose one response—out of four provided—to each question asked.

CLA+ requires that students use critical-thinking and written-communication skills to perform cognitively demanding tasks. The integration of these skills mirrors the requirements of serious thinking and writing tasks faced in life outside of the classroom.

OVERVIEW OF THE CLA+ PERFORMANCE TASK (PT)

Each PT requires students to use an integrated set of analytic reasoning, problem solving, and written-communication skills to answer an open-ended question about a hypothetical but realistic situation. In addition to directions and questions, each PT also has its own Document Library that includes a range of informational sources, such as: letters, memos, summaries of research reports, newspaper articles, maps, photographs, diagrams, tables, charts, and interview note or transcripts. Each PT is typically accompanied by between four and eight documents. Students are instructed to use these materials in preparing their answers to the Performance Task's question within the allotted 60 minutes.

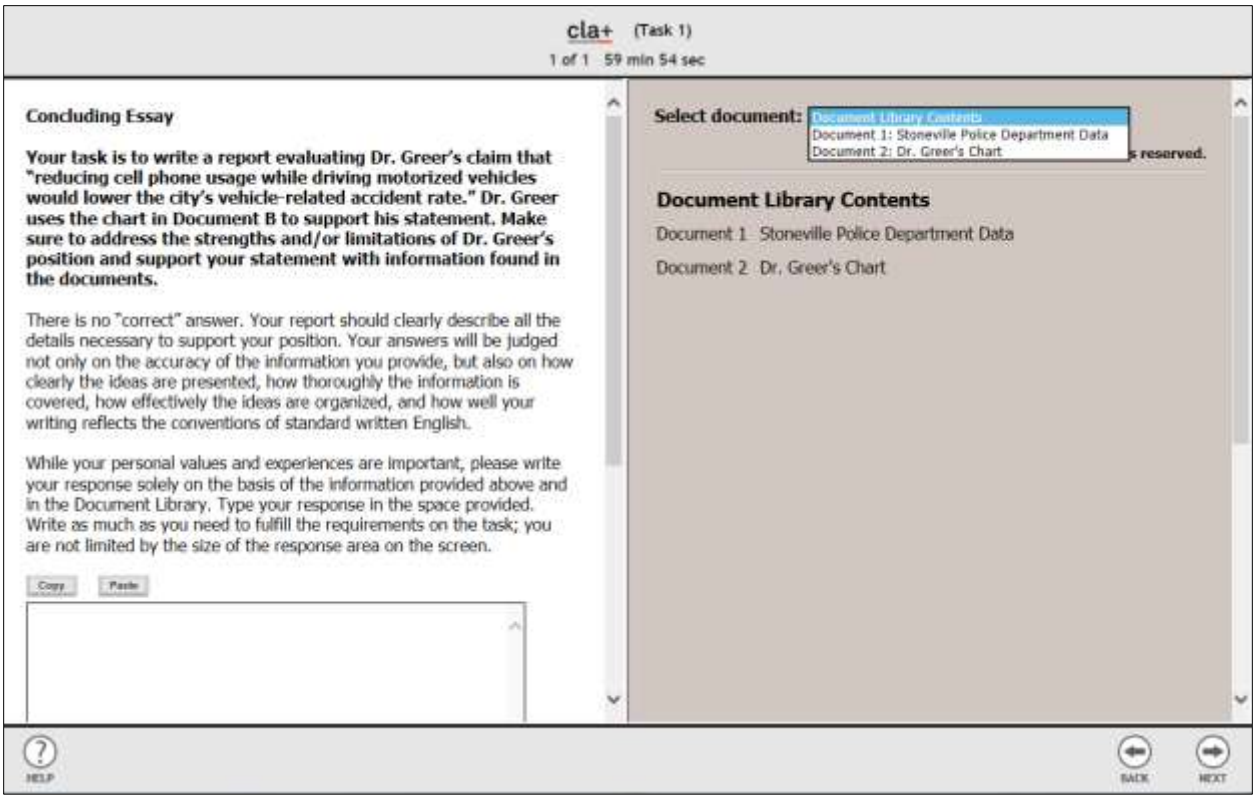
The first portion of each Performance Task contains general instructions and introductory material. The student is then presented with a split screen. On the right side of the screen is a list of the materials in the Document Library. The student selects a particular document to view by using a pull-down menu. A question and a response box are on the left side of the screen. An example is shown on the following page. There is no limit on how much a student can type.

No two PTs assess the exact same combination of skills. Some ask students to identify and compare and contrast the strengths and limitations of alternative hypotheses, points of view, courses of action, etc. To perform these and other tasks, students may have to weigh different types of evidence, evaluate the credibility of various documents, spot possible bias, and identify questionable or critical assumptions.

Performance Tasks my also ask students to suggest or select a course of action to resolve conflicting or competing strategies and then provide a rationale for that decision, including why it is likely to be better than one or more other approaches. For example, students may be asked to anticipate potential difficulties or hazards that are associated with different ways of dealing with a problem, including the likely short- and long-term consequences and implications of these strategies. Students may then be asked to suggest and defend one or more of these approaches.

Alternatively, students may be asked to review a collection of materials, and then choose amongst a set of options to solve a problem or propose a new solution to the problem. PTs often require students to marshal evidence from different sources; distinguish rational arguments from emotional ones and fact from opinion; understand data in tables and figures; deal with inadequate, ambiguous, or conflicting information; spot deception and holes in the arguments made by others; recognize information that is and is not relevant to the task at hand; identify additional information that would help to resolve issues; and weigh, organize, and synthesize information from several sources.

To view a sample CLA+ PT, please visit the Sample Tasks section of CAE's website at www.cae.org/cla.



Preview of the Performance Task Document Library

OVERVIEW OF CLA+ SELECTED-RESPONSE QUESTIONS (SRQs)

Like the PT, CLA+ SRQs require students to use an integrated set of critical-thinking skills across three question sets: the first assesses scientific and quantitative reasoning, the second assesses critical reading and evaluation, and the final set requires students to detect logical flaws and questionable assumptions to critique an argument. Also like the PT, each question set is accompanied by one to three documents of varying natures. Students are instructed to use these materials in preparing their answers to the questions within the allotted 30 minutes.

The **Scientific & Quantitative Reasoning** section contains ten questions that require students to use information and arguments provided in (an) accompanying document(s) to apply critical-thinking skills. Some of the questions may require students to: make inferences and hypotheses based on given results; support or refute a position; identify information or quantitative data that is connected and conflicting; detect questionable assumptions (such as implications of causation based on correlation); evaluate the reliability of the information provided (such as the experimental design or data collection methodology); draw a conclusion or decide on a course of action to solve the problem; evaluate alternate conclusions; or recognize that the text leaves some matters uncertain and propose additional research to address these matters. The supporting documents in this section present and discuss real-life research results.

The **Critical Reading & Evaluation** section also contains 10 questions that require students to use information and arguments from (an) accompanying document(s) to apply critical-thinking skills. Some of the questions may require students to: support or refute a position; identify connected and conflicting information; analyze logic; identify assumptions in arguments; make justifiable inferences; or evaluate the reliability of the information provided. The supporting documents in this section may present debates, conversations, or multiple literary or historical texts with opposing views on an authentic issue.

The **Critique an Argument section** contains five questions. Students are presented with a brief argument about an authentic issue, and must use their critical-thinking skills to critique the argument.

Some of the questions may require students to: evaluate alternate conclusions; address additional information that could strengthen or weaken the argument; detect logical flaws and questionable assumptions in the argument; and evaluate the reliability of information, including recognizing potential biases or conflicts of interest.

To view sample CLA+ SRQs, please visit the Sample Tasks section of CAE's website at www.cae.org/cla.

ASSESSMENT DEVELOPMENT

CAE has a team of experienced writers who—with researchers and editorial reviewers—generate ideas for tasks, question sets, and supporting documents. Each group then contributes to the development and revision of the tasks, questions, and accompanying documents.

Performance Task Development

During the development of PTs, care is taken to ensure that sufficient information is provided to permit multiple reasonable solutions to the issues present in the PT. Documents are crafted such that information is presented in multiple formats (e.g., tables, figures, news articles, editorials, emails, etc.).

While developing a PT, 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 documents, and that no unintended additional pieces of information are embedded. This list serves as a draft starting point for scorer trainings, and is used in alignment with the analytic scoring items used in the PT scoring rubrics.

During the editorial and revision process, 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 supported better than others.

The question for the PT is also drafted and revised during the development of the documents. The question is designed such that students are prompted to read and attend to multiple sources of information in the documents, then evaluate the documents and use their analyses to draw conclusions and justify those conclusions.

After several rounds of revisions, the most promising of the PTs and SRQ sets are selected for piloting. Student responses from the pilot test are examined to identify what pieces of information are unintentionally ambiguous, and what pieces of information in the documents should be removed. After revisions, the tasks that elicit the intended types and ranges of student responses are made operational.

Selected-Response Questions Development

The process for developing SRQs is similar to that of PTs. Writers develop documents—based on real-life data and issues—that might make use of flawed arguments, present multiple possibly valid (or invalid) assumptions or conclusions, and potentially leave open alternative conclusions or hypotheses. These characteristics serve as the foundation for the selected-response questions that accompany the documents.

During review, question editors work with writers to confirm that the correct answer options are in fact correct based on the information provided in the documents, and that incorrect answers are not potentially plausible. Likewise, reviewers take care to ensure that the questions are measuring the intended critical-thinking skills.

After several rounds of revision, the most promising of the SRQ passages and questions are selected for piloting. Student responses from the pilot test are examined to identify what pieces of information, questions, or response options are unintentionally ambiguous, and what pieces of information in the documents should be removed. After revision, the best-functioning question sets (i.e., those that elicit the intended types and ranges of student responses) are selected for the operational test.

APPENDIX G: SCORING CLA+

SCORING CRITERIA

Performance Task responses are scored in three skill areas: Analysis & Problem Solving, Writing Effectiveness, and Writing Mechanics. Each of these skill areas represents a subscore.

Subscores are assigned criterion-referenced scores, with each score value corresponding to characteristics of a student's response. The Performance Task rubric is available on our website at www.cae.org/claptrubric.

Selected-Response Question section scores are determined by the number of correct responses provided in each of the three question sets. Each question set represents its own subscore category: Scientific & Quantitative Reasoning (10 questions), Critical Reading & Evaluation (10 questions), and Critique an Argument (5 questions). Because each question set is not necessarily of equal difficulty, the scores are adjusted to account for differences between them, and then reported on a common scale. Details about this scaling process are provided in the *Scaling Procedures* section of this report (Appendix J).

THE SCORING PROCESS

During piloting of any new **Performance Tasks**, all responses are double-scored by human scorers. These scoring responses are then used to train human scorers and the machine-scoring engine for all operational test administrations.

Once tasks are fully operational, CLA+ uses a combination of automated and human scoring for its PTs. CAE uses Intelligent Essay Assessor (IEA) for its automated scoring. IEA is the automated scoring engine developed by Pearson Knowledge Technologies to evaluate the meaning of text, not just writing mechanics. Pearson has trained IEA for CLA+ using a broad range of real CLA+ responses and scores to ensure its consistency with scores generated by human scorers.

Unless a school chooses lower-stakes testing (no calculation of student Mastery Levels), each PT response is double-scored: once by IEA, and once by a trained human scorer. Responses from schools that choose low-stakes testing are scored only once, by IEA.

All scorer candidates undergo rigorous training in order to become certified CLA+ scorers. Training includes an orientation to the prompts and scoring rubrics/guides, repeated practice grading a wide range of student responses, and extensive feedback and discussion after scoring each response.

To ensure continuous human scorer calibration, CAE developed the E-Verification system for the online Scoring Interface. The E-Verification system was developed to improve and streamline scoring. Calibration of scorers through the E-Verification system requires scorers to score previously-scored results, or "Verification Papers", when they first start scoring, as well as throughout the scoring window. The system will periodically present Verification Papers to scorers in lieu of student responses, though they are not flagged to the scorers as Verification Papers. The system does not indicate when a scorer has successfully scored a Verification Paper, but if the scorer fails to accurately score a series of Verification Papers, he or she will be removed from scoring and must participate in a remediation process. At this point, scorers are either further coached or removed from scoring.

Each response receives subscores in the categories of Analysis & Problem Solving, Writing Effectiveness, and Writing Mechanics. Subscores are assigned on a scale of 1 (lowest) to 6 (highest). Blank responses or responses that are entirely unrelated to the task (e.g., writing about what they had for breakfast) are flagged for removal from results.

For the **Selected-Response Questions** section of CLA+, each question set receives a corresponding subscore. Subscores are determined by the number of questions that were correctly answered, with that score adjusted for the difficulty of the specific set received and then reported on a common scale. For instance a student might correctly answer seven questions on one Critical Reading & Evaluation question set but would only correctly answer six questions had they received another Critical Reading & Evaluation question set. Scores are equated so that each subscore category has the same mean and

standard deviation of scores, making them comparable to each other.

Unless a student fails to start a section or is unable to finish due to a technical glitch or connection error, any unanswered question is scored as incorrect, though if a student does not attempt at least half of the SRQs, the student will not receive a score for that section. Score values range from approximately 200 to 800 for each section.

APPENDIX H: MASTERY LEVELS

SETTING STANDARDS FOR 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. The study was held at CAE headquarters in New York, New York on December 12, 2013. Twelve distinguished panelists, representing varying sectors of college and employers were invited to participate in the study.

As part of the standard-setting study, the panelists discussed and defined the profile for the three 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 and also create a narrative profile for the necessary knowledge, skills, and abilities CLA+ students. 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. The table below lists the panelists for the CLA+ standard-setting study.

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
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

CLA+ MASTERY LEVELS

Individual CLA+ Mastery Levels are determined by the Total CLA+ score received by a given student. On the institution level, Mastery Levels are determined by the school's average performance for a given cohort of students.

The results of the CLA+ standard-setting study are used by CAE to distinguish between CLA+ students who have varying knowledge, skills, and abilities as measured by the assessment. Individual institutions should not use these results for purposes other than this (e.g., basing graduation or employment decisions on individual CLA+ levels of mastery). If an institution is interested in using CLA+ results as part of a graduation requirement, a separate standard-setting study should be conducted with this specific purpose in mind.

The following table summarizes each level of mastery and also provides a description of the students who are below the basic level of mastery.

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.
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.

APPENDIX I: DIAGNOSTIC GUIDANCE

INTERPRETING CLA+ RESULTS

CLA+ results can be used as a measure of overall institutional performance on tasks that measure higher-order skills, and can also be a tool for identifying areas of skill or weakness for individual students. Examining performance across sections of CLA+ can serve as an initial diagnostic exercise. The two sections of CLA+—the Performance Task and the Selected-Response Questions—differ in the combination of skills necessary for high performance.

The PT measures Analysis & Problem Solving, Writing Effectiveness, and Writing Mechanics. The SRQs measure Scientific & Quantitative Reasoning, Critical Reading & Evaluation, and Critique an Argument (the ability to detect logical flaws and questionable assumptions).

Selected-Response Question subscores are assigned based on the number of questions correctly answered, with that value adjusted for the difficulty of the particular questions sets received, and then converted to a common scale. These subscores were placed on a scale with a mean of 500 and a standard deviation of 100 based on the performance of freshmen that tested in fall 2013, and so each SRQ subscore score ranges from approximately 200 to 800.

Performance Task subscores are assigned on a scale of 1 (lowest) to 6 (highest). Unlike the SRQ subscores, PT subscores are not adjusted for difficulty. These subscores remain unadjusted because they are intended to facilitate criterion-referenced interpretations. For example, a "4" in Analysis and Problem Solving means that a response had certain qualities (e.g., "Provides valid support that addresses multiple pieces of relevant and credible information..."), and any adjustment to that score would compromise the interpretation.

The ability to make claims like, "Our students seem to be doing better in Writing Effectiveness than in Analysis & Problem Solving on the Performance Task" is clearly desirable. This can be done by comparing each subscore distribution to its corresponding reference distribution displayed in Section 4 (page 5) of your institutional report. Please examine the results presented in the Performance Task subscore distribution table in combination with the Performance Task scoring rubric, available on CAE's website at www.cae.org/claptrubric.

In addition to the subscores, the CLA+ Mastery Level scores allow for results to be interpreted in terms of the qualities that are exhibited by examinees. Each Mastery Level corresponds to certain evidence of critical-thinking and written-communication skills. Details about each Mastery Level are provided in the preceding section of this report (*Mastery Levels*, Appendix H).

COMPARING RESULTS ACROSS ADMINISTRATIONS

One way to assess performance is to track changes in CLA+ scores over time. This can be done either by testing a cohort of students longitudinally, or participating regularly in cross-sectional administrations of CLA+.

Because the assessment format for CLA+ differs from that of its predecessor, the CLA, direct score comparisons will not be feasible across data from before and after fall 2013. However, scaling equations can be used to adjust CLA scores for making comparisons with CLA+.

Schools wishing to compare CLA+ results to prior year results can use the following equation, which was derived by comparing CLA and CLA+ total scores from 134 institutions that tested students on both the new and old forms of the assessment (r=0.822):

CLA scores from fall 2010 – spring 2013:

$$score_{CLA+} = 184.188 + (0.812 \cdot score_{CLA})$$

CLA scores from before fall 2010:

$$score_{CLA+} = 268.066 + (0.6897 \cdot score_{CLA})$$

Aside from making direct score comparisons across earlier administrations, schools can also use their percentile rankings to determine changes in performance relative to other CLA+ institutions.

All test administrations after the fall 2013 CLA+ window, however, will be easily comparable to each other. Because the institutional and student sample used for setting norms (percentile rankings, value-added parameters, etc.) are fixed based on the institutional sample from the 2013-14 academic year, any changes in value-added score or ranking can be attributed to the school's CLA+ results, rather than potential shifts in the norming sample.

APPENDIX J: SCALING PROCEDURES

CONVERTING CLA+ SCORES TO A COMMON SCALE

To provide CLA+ scores, CAE converts each section score to a common scale of measurement. This allows for the combining of scores from different tasks to compute a school's mean scale score for each section of CLA+, as well as a total average scale score across the two CLA+ sections.

For each **Performance Task**, raw subscores are summed to produce a raw total score. Because not all tasks have the exact same level of difficulty, raw total scores from 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 approximately the same percentile rank regardless of the task on which it was earned.

For the PT, a linear transformation is used to convert raw scores to scale scores. This process results in a scale score distribution with the same mean and standard deviation as the SAT Math and SAT Critical Reading combined (or converted ACT) scores of college freshmen taking CLA+; in this case the data used were from college freshmen that took CLA+ in fall 2013. This type of scaling preserves the shape of the raw score distribution and maintains the relative standing of students. 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 makes it such that a very high raw score earned on the task (not necessarily the highest possible score) corresponds approximately to the highest SAT (or converted ACT) score of any freshman who tested in fall 2013. Similarly, a very low raw score earned on a task would be assigned a scale score value that is close to the lowest SAT (or converted ACT) score of any freshman who took CLA+ in fall 2013. On rare occasions that students achieve exceptionally high or low raw scores, this scaling procedure may produce scale scores that fall outside the normal SAT (Math + Critical Reading) score range of 400 to 1600.

For each of the subscores in the **Selected-Response Questions** section, raw section scores—determined by the number of correct responses—are first equated, and then converted to a common scale. The equating process simply takes the scores from each set of questions in a subscore area, and then converts those scores to have the same mean and standard deviation across all forms of the question set. This process adjusts scores according to the difficulty of a given item set, so that comparisons can be made across test forms.

These equated section scores are then converted to a more interpretable scale using a linear transformation, with a mean of 500 and standard deviation of 100, based on the fall 2013 freshmen taking CLA+. This scale results in selected-response section subscores ranging from approximately 200 to 800, similar to the subsections of the SAT.

The weighted average of the section subscores are then transformed again, using the same scaling parameters as the PT—using the distribution of SAT scores of college freshmen that took CLA in fall 2012—to place both sections of CLA+ on the same scale.

CLA+ **Total Scores** are calculated by taking the average of the two sections of CLA+ completed by a given student. Students that did not complete or give scorable responses to both sections of the assessment will not receive total scores.

SCALING EAA SCORES

To facilitate reporting results across schools, ACT scores are converted (using the ACT-SAT crosswalk that follows) to the scale of measurement used to report combined SAT Math and Critical Reading scores.

For institutions where a majority of students did not have ACT or SAT scores (e.g., two-year institutions and open-admission schools), we make available the Scholastic Level Exam (SLE), a short-form cognitive ability measure, as part of CLA+. The SLE is produced by Wonderlic, Inc. SLE scores are

converted to SAT scores using data from 1,148 students that participated in the spring 2006 administration of the CLA that had both SAT and SLE scores.

These converted scores (both ACT to SAT and SLE to SAT) are referred to simply as entering academic ability (EAA) scores.

Standard ACT to SAT Crosswalk

ACT	SAT
36	1600
35	1560
34	1510
33	1460
32	1420
31	1380
30	1340
29	1300
28	1260
27	1220
26	1190
25	1150
24	1110
23	1070
22	1030
21	990
20	950
19	910
18	870
17	830
16	790
15	740
14	690
13	640
12	590
11	530

Source: ACT (2008). *ACT/College Board Joint Statement.* Retrieved from http://www.act.org/aap/concordance/pdf/report.pdf

APPENDIX K: MODELING DETAILS

MODELING STUDENT-LEVEL SCORES

Within each school, an equation like the following is used to model the relationship between senior students' EAA scores and their CLA+ scores:

$$CLA_{ij} = \overline{CLA}_j + 0.43\left(EAA_{ij} - \overline{EAA}_j\right) + r_{ij}$$

(Note that the coefficients used here are for illustrative purposes only; the coefficients used for CLA+ value-added modeling will be available following the spring 2014 administration of CLA+.)

In this equation, CLA_{ij} is student i in school j's CLA+ score, and this is modeled as a function of school j's average senior CLA+ score ($\overline{CLA}_j$) and student i's EAA score (EAA_{ij}) minus the average EAA score of participating seniors at school j. Specifically, a student's CLA+ score equals (a) the school's average senior CLA+ score plus (b) an adjustment based on the student's EAA score relative to the average among senior participants in school j and (c) a residual term r_{ij} equal to the difference between a student's observed and expected CLA+ performance, with positive numbers "better than expected." Here, the student-level slope coefficient for EAA is 0.43, which indicates that for every 1 point difference in EAA, one would expect a 0.43 point difference in CLA+ performance.

To illustrate the use of this equation for computing a student's expected CLA+ score, consider a school with an average senior CLA+ score of 1200 and an average EAA score of 1130. A senior student in this school with an EAA score of 1080 would be expected to have a CLA+ score of $1200 + 0.43(1080 - 1130) = 1179$. If this student actually scored a 1210 on CLA+, the residual term r_{ij} would be +31 because this student scored 31 points higher than one would expect given his or her EAA. Using the equation described here would produce student-level deviation scores that differ slightly from those that inform the performance levels reported in your Student Data File.

MODELING SCHOOL-LEVEL SCORES

Institutional value-added scores are derived from the school-level equation[2] of the HLM, which takes the form

$$\overline{CLA}_j = 355 + 0.32\left(EAA_j\right) + 0.45\left(\overline{CLA}_{fr,j}\right) + u_j$$

Where $\overline{CLA}_{fr,j}$ is the average CLA+ score of participating freshmen at school j, and u_j is that school's value-added score estimate ($\overline{CLA}_j$ and EAA_j are defined the same as in the student-level equation). Specifically, u_j is the difference between a school's observed and expected average senior CLA+ performance. In this equation, 355 is the school-level intercept, 0.32 is the school-level slope coefficient for average EAA, and 0.45 is the school-level slope coefficient for average freshman CLA+. Combined with average EAA and average freshman CLA+ scores, these coefficients allow for computing expected senior average CLA+ scores.

It may seem unconventional to use the average freshman CLA+ score from a different group of students as a predictor of the average senior CLA+ score, but analyses of CLA+ data consistently indicate that average freshman CLA+ performance adds significantly to the model. That is, average EAA and average freshman CLA+ account for different but nevertheless important characteristics of students as they enter college. Moreover, this model would not be credible as a value-added model for CLA+ scores if there were no control for CLA+ performance at the start of college.

As a conceptual illustration of this approach, consider several schools administering CLA+ to groups of seniors that had similar academic skills upon entering college—as indicated by average SAT or ACT scores and average freshman CLA+ scores. If, at the time of graduation, average CLA+ performance at one school is greater than average performance at the other schools testing groups of students with similar entering characteristics, one can infer that greater gains in critical-thinking and written-communication skills occurred at this school. That is,

[2] As with the example provided for student-level CLA+ scores, the coefficients provided here are for illustrative purposes only. CLA+ value-added modeling coefficients will be provided to participating schools following the close of the spring 2014 CLA+ administration.

this school has greater value added than the other schools.

To illustrate the use of the school—level equation for estimating value-added scores, consider a school with an average freshman CLA+ score of 1050, an average senior CLA+ score of 1200, and an average senior EAA score of 1130. According to the school-level equation, one would expect the senior average CLA+ performance at this school to be $355 + 0.32(1130) + 0.45(1050) = 1189$. The observed senior average CLA+ performance was 1200, which is 11 points higher than the typical school testing students with similar EAA and freshman CLA+ scores. Converted to a standard scale, the value-added score would be 0.28, which would place the school in the "Near Expected" performance category of value added.

Value-added scores are properly interpreted as senior average CLA+ performance relative to the typical school testing students with similar academic skills upon entering college. The proper conditional interpretation of value-added skills is essential.

First, it underscores the major goal of value-added modeling: obtaining a benchmark for performance based on schools admitting similar students. Secondly, a high value-added score does not necessarily indicate high absolute performance on CLA+. Schools with low absolute CLA+ performance may obtain high value-added scores by performing well relative to expected (i.e., relative to the typical school testing students with similar academic skills upon entering college). Likewise, schools with high absolute CLA+ performance may obtain low value-added scores by performing poorly relative to expected. Though it is technically acceptable to interpret value-added scores as relative to all other schools participating in CLA+ after controlling for entering student characteristics, this is not the preferred interpretation because it encourages comparisons among disparate institutions.

INTERPRETING CONFIDENCE INTERVALS

It is important to keep in mind that value-added scores are estimates of unknown quantities. Put another way, the value-added score each school receives is a "best guess" based on the available information. Given their inherent uncertainty, value-added scores must be interpreted in light of available information about their precision. HLM estimation (described in the *Methods* section of this report, Appendix B) provides standard errors for value-added scores, which can be used to compute a unique 95% confidence interval for each school. These standard errors reflect within- and between-school variation in CLA+ and EAA scores, and they are most strongly related to senior sample size. Schools testing larger samples of seniors obtain more precise estimates of value added and therefore have smaller standard errors and corresponding 95% confidence intervals.

With a senior sample size near 100, our example school has a standard error of 0.35 (on the standardized value-added score scale). This school's 95% confidence interval has a range from -0.41 to 0.97, which was calculated as the value-added estimate plus or minus 1.96 multiplied by the standard error. To provide some perspective, consider that the confidence interval would have been about 30% larger (from -0.60 to 1.16) if this school tested half as many students. If this school tested twice as many students, the confidence interval would have been about 20% smaller (from -0.26 to 0.83).

Unfortunately, inaccurate interpretations of confidence intervals are common. It is *not* correct to say "there is a 95% chance that my school's 'true' value-added score is somewhere between -0.41 and 0.97" because it is either in the interval or it is not in the interval. Unfortunately, we cannot know which. The confidence interval reflects uncertainty in the estimate of the true score (due to sampling variation), not uncertainty in the true score itself. Correctly interpreted, a 95% confidence interval indicates the variation in value-added scores we should expect if testing were repeated with different samples of students a large number of times. It may be stated that, "if testing were repeated 100 times with different samples of students, about 95 out of the 100 resulting confidence intervals would include my school's 'true' value-added score."

Using conventional rules for judging statistical significance, one could draw several inferences from this school's 95% confidence interval. First, it can be said that this school's value-added score is significantly different from value-added scores lower than -0.41 and greater than 0.97. Second, because 0 is within the range of the 95% confidence interval, it may be said that this school's value-added score is not significantly different from 0. Note that a value-added score of 0 does *not* indicate zero learning; it

instead indicates typical (or "near expected") senior average CLA+ performance, which implies learning typical of schools testing students with similar academic skills upon entering college.

STATISTICAL SPECIFICATION OF THE CLA+ VALUE-ADDED MODEL

Level 1 (Student Level): $CLA_{ij} = \beta_{0j} + \beta_{1j}(EAA_{ij} - \overline{EAA}_j) + r_{ij}$

- CLA_{ij} is the CLA+ score of student i at school j.
- EAA_{ij} is the Entering Academic Ability score of student i at school j.
- $\overline{EAA}_j$ is the mean EAA score at school j.
- β_{0j} is the student-level intercept (equal to the mean CLA+ score at school j).
- β_{1j} is the student-level slope coefficient for EAA at school j(assumed to be the same across schools).
- r_{ij} is the residual for student i in school j, where $r_{ij} \sim N(0, \sigma^2)$ and σ^2 is the variance of the student-level residuals (the pooled within-school variance of CLA+ scores after controlling for EAA).

Level 2 (School Level): $\beta_{0j} = \gamma_{00} + \gamma_{01}(\overline{EAA}_j) + \gamma_{02}(\overline{CLA}_{fr,j}) + \mu_{0j}$ and $\beta_{1j} = \gamma_{10}$

- $\overline{CLA}_{fr,j}$ is the mean freshman CLA+ score at school j.
- γ_{00} is the school-level value-added equation intercept.
- γ_{01} is the school-level value-added equation slope coefficient for senior mean EAA.
- γ_{02} is the school-level value-added equation slope coefficient for freshman mean CLA+.
- γ_{10} is the student-level slope coefficient for EAA (assumed to be the same across schools).
- μ_{0j} is the value-added equation residual for school j (i.e., the value-added score), where $\mu_{0j} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\tau_{00} & 0\\0 & 0\end{bmatrix}\right)$ and τ_{00} is the variance of the school-level residuals (the variance in mean CLA+ scores after controlling for mean EAA and mean freshman CLA+ scores).

Mixed Model (combining the school- and student-level equations):

$$CLA_{ij} = \gamma_{00} + \gamma_{01}(\overline{EAA}_j) + \gamma_{02}(\overline{CLA}_{fr,j}) + \gamma_{10}(EAA_{ij} - \overline{EAA}_j) + \mu_{0j} + r_{ij}$$

ESTIMATED PARAMETERS FOR THE VALUE-ADDED MODEL

Estimated Parameters for the Value-Added Model

	γ_{00}	γ_{10}	γ_{01}	γ_{02}	STANDARD DEVIATION
TOTAL CLA+ SCORE	--	--	--	--	--
PERFORMANCE TASK	--	--	--	--	--
SELECTED-RESPONSE QUESTIONS	--	--	--	--	--

Following the completion of the spring administration of CLA+, the table above will show the estimated parameters for the CLA+ value-added model. Using these estimated parameters and the instructions below (also described in the statistical models on the previous page), one can compute the expected senior CLA+ score for a given school. In combination with the observed mean score for seniors at that school, this can be used to compute the school's value-added score. These values can also be used to perform subgroup analyses, or estimate value-added for groups of students that have been tested longitudinally.

HOW TO CALCULATE CLA+ VALUE-ADDED SCORES

To calculate value-added scores for your students, you will need:

- Samples of entering and exiting students with CLA+ and EAA scores (see your CLA+ Student Data File)
- The estimated parameters for the value-added model (see table above)

1. Refer to your CLA+ Student Data File to identify your subgroup sample of interest. The subgroup must contain freshmen and seniors with CLA+ scores and EAA scores.

2. Using your CLA+ Student Data File, compute:
   - The mean EAA score of seniors (exiting students) in the sample
   - The mean CLA+ score of freshmen (entering students) in the sample
   - The mean CLA+ score of seniors (exiting students) in the sample

3. Calculate the senior sample's expected mean CLA+ score, using the parameters from the table above. Please note that the same equation can be used for the individual sections of CLA+, as well as for the total CLA+ score, by placing the appropriate parameter values from the appropriate row in the table into the equation:

$$Expected\ Score = \gamma_{00} + \gamma_{01}(senior\ mean\ EAA) + \gamma_{02}(freshman\ mean\ CLA\ score)$$

4. Use your expected score to calculate your subgroup sample's value-added score:

$$Value\text{-}added\ Score,\ unstandardized = (Observed\ senior\ mean\ score) - (Expected\ senior\ mean\ score)$$

5. Convert that value-added score to standard deviation units:

$$Value\text{-}added\ Score,\ standardized = \frac{Value\text{-}added\ score,\ \ unstandardized}{Standard\ Deviation}$$

APPENDIX L: PERCENTILE LOOK-UP TABLES

PERCENTILE LOOK-UP TABLES FOR CLA+ SCORES

For schools interested in the distribution of performance across CLA+, CAE provides percentile look-up tables for CLA+ scores. The tables list the scores for each section of CLA+ (PT and SRQs), as well as the Total CLA+ and EAA, associated with each percentile value.

These tables are available on CAE's website. Institution-level percentile scores are accessible at www.cae.org/claplusschoolpercentiles, and student-level percentile scores are available at www.cae.org/claplusStudentpercentiles.

The tables currently only contain data for freshmen, but following the spring 2014 administration of CLA+, these tables will be updated to contain percentile values for college seniors, as well as for value-added scores.

APPENDIX M: STUDENT DATA FILE

EXPLORING STUDENT DATA

In tandem with your report, we provide a CLA+ Student Data File, which includes variables across three categories: self-reported information from students in their CLA+ online profile and CLA+ post-assessment survey; CLA+ scores and identifiers; and information provided by the registrar.

We provide student-level information for linking with other data you collect (e.g., from NSSE, CIRP, portfolios, local assessments, course-taking patterns, participation in specialized programs, etc.) to help you hypothesize about factors related to institutional performance.

Student-level scores were historically not designed to be diagnostic at the individual level; however, with the CLA+ student-level scores have greater utility. Student-level results can now be used for formative purposes, to identify individual students' areas of weakness, as well as that of the entire sample.

The student data files now include more information such as subscores for Scientific & Quantitative Reasoning, Critical Reading & Evaluation, and Critique an Argument (the ability to identify logical fallacies and questionable assumptions within an argument).

The data file also includes responses from new survey questions—where students were asked about the level of effort they put into each section of CLA+ and how engaged they were with the assessment—to provide more context to institutional results and individual student scores. These responses can also help schools identify motivation issues within their samples, so that they can make adjustments to their outreach and recruitment methods for future administrations.

If an institution has elected to use the Local Survey feature within the CLA+ testing platform—a tool that allows schools to append up to nine survey questions of their own to the end of the assessment—these questions and students' corresponding responses will also appear in the student data file, allowing schools to create a richer, customized data set to facilitate institutional research with CLA+.

Schools may also choose to analyze the performance of subgroups of students to determine whether certain groups of students might be in need of targeted educational enhancements. Value-added scores can be estimated for these subgroups, as well, and compared to growth seen across the institution as a whole.

Starting with the fall 2013 administration student-level CLA+ data can also be compiled from year to year, yielding a larger and much richer dataset than could be achieved within a single academic year. Likewise, aggregating student data across years allows schools to longitudinally track and identify improvements in critical thinking and written communication made by individual students.

APPENDIX N: MOVING FORWARD

WHAT NEXT?

The information presented in your institutional report is designed to help you better understand the contributions your institution is making toward your students' learning gains. However, the institutional report alone provides a snapshot of student performance.

When combined with the other tools and services CLA+ has to offer, the institutional report can become a powerful tool in helping you and your institution target specific areas of improvement, while effectively and authentically aligning teaching, learning, and assessment practices in ways that may improve institutional performance over time.

We encourage institutions to examine performance across CLA+ and communicate the results across campus, link student-level CLA+ results with other data sources, pursue in-depth sampling, collaborate with their peers, and participate in professional development offerings.

Student-level CLA+ results are provided for you to link to other data sources (e.g., course-taking patterns, grades, portfolios, student surveys, etc.). These results are strengthened by the provision of scores in the areas of Analysis & Problem Solving, Writing Effectiveness, Writing Mechanics, Scientific & Quantitative Reasoning, Critical Reading & Evaluation, and Critique an Argument to help you pinpoint specific areas that may need improvement. Internal analyses, which you can pursue through in-depth sampling, can help you generate hypotheses for additional research.

While peer-group comparisons will be available in the coming year, the true strength of peer learning comes through collaboration. CLA+ facilitates collaborative relationships among our participating schools by encouraging the formation of consortia and hosting periodic web conferences featuring campuses doing promising work using CLA+.

Our professional development services shift the focus from general assessment to the course-level work of faculty members. Performance Task Academies—two-day hands-on training workshops—provide opportunities for faculty to receive guidance in creating their own CLA+-like performance tasks, which can be used as classroom or homework assignments, curriculum devices, or even local-level assessments. More information is available on the Events page of the CAE website (www.cae.org).

CAE staff can also provide institutions with workshop sessions geared toward making use of your student data file. In these sessions, CAE researchers will collaborate with institutional staff to find ways to dig deeper into student results to answer questions about students' performance on CLA+, and identify areas of strength or weakness. To arrange for one of these sessions, please email clateam@cae.org.

Through the steps noted above, we encourage institutions to move toward a continuous system of improvement stimulated by CLA+. Our programs and services—when used in combination—are designed to emphasize the notion that, in order to successfully improve higher-order skills, institutions must genuinely connect their teaching, learning, and assessment practices in authentic and effective ways.

Without your contributions, CLA+ would not be on the exciting path that it is on today. We look forward to your continued involvement!

APPENDIX O: CAE BOARD OF TRUSTEES AND OFFICERS

CAE Board of Trustees and Officers

Council for Aid to Education
215 Lexington Avenue
Floor 16
New York, NY 10016

CAE

cla+

TECHNICAL FAQ

cae | 215 Lexington Avenue, Floor 16 | New York, NY 10016 | cla@cae.org | (212) 217-0700 | @cla_beat

WHAT IS THE DIFFERENCE BETWEEN THE CLA AND CLA+?

When first launched more than a decade ago, the Collegiate Learning Assessment (CLA) pioneered a constructed-response approach to the assessment of higher-order skills. Initially, CLA was designed to measure an institution's contribution, or value added, to the development of these higher-order skills to their student body, and therefore the institution—not the student—was the primary unit of analysis.

CLA employed a matrix sampling approach, under which students were randomly distributed either a Performance Task (PT) or an Analytic Writing Task, for which students were allotted 90 minutes and 75 minutes, respectively. CLA Performance Tasks presented a real-world situation in which an issue, problem, or conflict is identified, and students are asked to assume a relevant role to address the issue, suggest a solution, or recommend a course of action based on the information provided in a document library. Analytic Writing Tasks consisted of two components—one in which students are presented with a statement around which they must construct an argument (Make an Argument), and another in which students are given a logically flawed argument that they must then critique (Critique an Argument).

In its original form, the utility of CLA was limited. Because the assessment consisted of just one or two responses from each student, reliable results were only available at the institutional level, and students' results were not directly comparable. Likewise, reporting for CLA was restricted to the purposes of its value-added measure; institutions were not eligible for summary results unless they had tested specified class levels in the appropriate testing windows.

Now, however, CLA+ has both greater utility and more flexibility. CLA+ comprises both a performance task and a set of selected-response questions, with each student receiving both components of the assessment. The inclusion of different item formats not only allows for the assessment of a broad range of content and skills but it also affords the opportunity to obtain reliable results at both the institution and student levels. As in its former incarnation, CLA+ allows for an institution to compare its student learning results with the learning results at similarly selective institutions, and use that information to improve teaching and learning. Additionally, CLA+ results at the student level allow for a comparison of how well each individual performed relative to their peers at the same class level both within an institution and across CLA+. Unlike the CLA reports, the CLA+ reports will include data on each of the four standard college class levels, allowing schools to receive results for any class level they test, regardless of the window in which the assessment occurs.

The new selected-response questions (SRQs) are also developed with the intent to assess higher-order cognitive skills rather than the recall of factual knowledge. Similar to the Performance Task, students are presented with a set of questions as well as one or two documents to refer to when answering each question. The supporting documents include a range of information sources, such as letters, memos, photographs, charts, and newspaper articles.

In addition to the subscores previously reported for the Performance Task (Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics), CLA+ uses selected-response items to measure students' performance on the following higher-order skills: Scientific and Quantitative

Reasoning, Critical Reading and Evaluation, and Critique an Argument (formerly a constructed-response section of the Analytic Writing Task).

As was the case with the CLA, CLA+ requires 90 minutes of testing time, during which students will have an hour to complete a Performance Task and half an hour to respond to the 25 selected-response questions. Like performance tasks, the selected-response items require students to analyze accompanying documents. Students then answer questions that are designed to measure their ability to apply scientific and quantitative reasoning, critically read and evaluate the texts, and detect logical flaws and questionable assumptions to critique an argument.

Additionally, CLA+ also introduces a new metric in the form of mastery levels. These mastery levels indicate how well students have mastered the higher-order skills measured by CLA+. The mastery level categories are: Below Basic, Basic, Proficient, and Advanced. These levels are determined by specific cut scores established in a standard setting study conducted by CAE with a panel of experts in fall 2013. More information regarding the standard setting study is presented in this document.

CLA+ TASKS

How are CLA+ tasks developed?

CAE item developers follow a rigorous and structured item development plan when creating new PTs and SRQs. The primary goal is to develop assessment items that are authentic and engaging for the students. This is accomplished through a series of checklists, including whether the students can reasonably craft an argument using only the information that is provided and whether there is enough information to support or refute from multiple perspectives. One of the unique features of CLA+ is that no prior knowledge of any specific content area is necessary in order to perform well on the assessment. Students are assessed on their critical-thinking and written-communication skills, not on how much knowledge they have of subjects such as US history or chemistry.

The documents for both the PTs and the SRQs are presented in the most appropriate format for the scenario. This can include, but is not limited to, an abstract from a journal, tables, charts, graphs, memos, blog postings, newspaper articles, maps, and reports. Throughout development, CAE item developers outline, write, and revise the content from each document within a PT or a selected-response set. This process ensures that the documents cover all of the necessary information and that no additional or unintentional content is imbedded in or missing from the documents. CAE editors review initial drafts of the tasks and provide feedback to the developer for revisions.

For the PTs specifically, item developers are instructed to create scenarios where there is more than one possible conclusion, solution, or recommendation. Each possible outcome is supported by at least some evidence provided in the documents. Typically, some of the possible conclusions are designed to be better supported than others. However, if a student chooses one of these less supported conclusions, he or she can still perform well on the PT if the response aligns to the higher range of the scoring rubric.

The SRQs, like the PTs, represent real-world scenarios and problems. Students are expected to answer questions that require them to critically read and evaluate a passage or situation, use scientific and/or quantitative reasoning, and identify logical fallacies in an argument. These types of questions, therefore, require students to think at a deeper level than the traditional recall-and-recognition questions which are seen on many standard multiple-choice assessments.

After several rounds of revision between the developer and one or more of CAE's editors, the most promising tasks are selected for pilot testing. During piloting, students complete one pilot task as if it were an operational task. At this point, draft scoring procedures 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.

For the selected-response items, a classical item analysis is conducted after the pilot testing to determine whether further revisions are necessary before the item becomes operational. Items are examined in terms of item discrimination and item difficulty. A point-biserial correlation is computed to determine the relationship between the item score (correct vs. incorrect) and the total test score. This value is often referred to as the item discrimination index. A high correlation between the item score and the total test score is an indication that the item does well at discriminating between students with low test scores and students with high test scores. The item difficulty, called a "p-value," is the proportion of students answering the item correctly and the total number of students who answered the question. The p-value is examined to ensure that there is sufficient range in terms of item difficulty, meaning there should not be too many items that are either very difficult or very easy. The items that are too difficult or too easy tend to not have satisfactory point-biserial correlations because they do not discriminate well between students with low test scores and students with high test scores. Operational items in the CLA+ bank have p-values between .30 and .80 and a point-biserial correlation of at least .10.

The item developers, editors, and measurement scientists who develop CLA+ tasks have varied backgrounds including history, English, mathematics, psychology, and psychometrics. Over the years of developing the CLA and CLA+, the team now has extensive experience with test development and writing evaluation.

What is the benefit of including different item formats (Performance Tasks and Selected-Response Questions) in the assessment?

As mentioned in the introduction, prior to CLA+, the assessment was only valid and reliable at the institutional level. CLA clients often asked if the student results could be used to make decisions about performance at the individual student level. CAE recommended against using the scores as individual assessments because the reliability at the individual level was not established.

In order to increase the reliability of the CLA scores for individual students, more data needed to be collected from each student. There were several different models that could have been employed, including administering more than one PT to each student. However, due to the desire to limit the amount of time students spend testing, CAE decided to develop CLA+ with the traditional performance-based PT as the anchor and a set of 25 SRQs, which assess the same construct as the PT (analytic reasoning and problem solving). These SRQs boost the reliability at the individual student level significantly while keeping the total testing time the same as the original CLA.

Additionally, both PTs and SRQs are capable of measuring critical-thinking skills. Each section has strengths and weaknesses and the arrangement of the two different format types creates a balance of strengths relative to weaknesses in terms of content coverage, reliability and validity evidence, and scoring objectivity and efficiency.

SCORING

Can you describe the CLA+ scoring rubrics?

The CLA+ scoring rubric for the PTs consists of three subscores: Analysis and Problem Solving (APS), Writing Effectiveness (WE), and Writing Mechanics (WM). Each of these subscores is scored from a range of 1-6, where 1 is the lowest level of performance and 6 is the highest level of performance, with each score pertaining to specific response attributes. For all task types, blank or entirely off-topic responses are flagged for removal from results. Because each prompt may have differing possible arguments or relevant information, scorers receive prompt-specific guidance in addition to the scoring rubrics. Additionally, the reported subscores are not adjusted for difficulty like the overall CLA+ scale scores, and therefore are not directly comparable to each other. These PT scores are intended to facilitate criterion-referenced interpretations, as defined by the rubric.

APS measures a student's ability to make a logical decision or conclusion (or take a position) and support it with accurate and relevant information (facts, ideas, computed values, or salient features) from the Document Library.

WE assesses a student's ability to construct and organize logically cohesive arguments. This is accomplished by strengthening the writer's position by elaborating on facts or ideas (e.g., explaining how evidence bears on the problem, providing examples, and emphasizing especially convincing evidence).

WM evaluates a student's facility with the conventions of standard written English (agreement, tense, capitalization, punctuation, and spelling) and control of the English language, including syntax (sentence structure) and diction (word choice and usage).

The CLA+ rubric is available on our website at http://cae.org/images/uploads/pdf/CLA_Plus_Scoring_Rubric.pdf.

The selected-response section of CLA+ consists of three subsections, each of which has a corresponding subscore category: Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument. Subscores in these sections are scored according to the number of questions correctly answered, with scores adjusted for the difficulty of the particular question set received. Scores for scientific and quantitative reasoning and critical reading and evaluation can range from 0 to 10, and scores for critique an argument range from 0 to 5.

How are CLA+ tasks scored?

During piloting of any new Performance Tasks, all responses are double-scored by human scorers. These double-scored responses are then used to train human scorers and a machine-scoring engine for all operational test administrations.

CAE uses Intelligent Essay Assessor (IEA) for its machine scoring. IEA is the automated scoring engine developed by Pearson Knowledge Technologies to evaluate the meaning of a text, not just writing mechanics. Pearson has designed IEA for CLA+ using a broad range of real CLA+ responses and scores to ensure its consistency with scores generated by human scorers. Thus, human scorers remain the basis for scoring the CLA+ tasks. However, automated scoring helps to increase scoring accuracy, reduce the amount of time between a test administration and report delivery, and lower costs. The automated essay scoring technique that CLA+ uses is known as Latent Semantic Analysis (LSA), which extracts the underlying meaning in written text. LSA uses mathematical analysis of about 800 student responses per prompt and the collective expertise

of human scorers, and applies what it has learned from the expert scorers to previously unscored student responses.

Once tasks are fully operational, CLA+ uses a combination of automated and human scoring for its Performance Tasks. Each Performance Task response is double-scored: once by IEA, and once by a trained human scorer. Although nearly all responses are scored by a both a human scorer and IEA, occasionally IEA identifies unusual responses, which are then automatically sent to the human scoring queue.

All scorer candidates undergo rigorous training in order to become certified CLA+ scorers. Training includes an orientation to the prompts and scoring rubrics/guides, repeated practice grading a wide range of student responses, and extensive feedback and discussion after scoring each response.

To ensure continuous human scorer calibration, CAE developed the E-Verification system for the online scoring interface. The E-Verification system was developed to improve and streamline scoring. Calibration of scorers through the E-Verification system requires scorers to score previously-scored results, or "Verification Papers," when they first start scoring, as well as throughout the scoring window. The system will periodically present Verification Papers to scorers in lieu of student responses, though they are not flagged to the scorers as Verification Papers. The system does not indicate when a scorer has successfully scored a Verification Paper, but if the scorer fails to accurately score a series of Verification Papers, he or she will be removed from scoring and must participate in a remediation process. At this point, scorers are either further coached or removed from scoring.

Each response receives subscores in the categories of Analysis & Problem Solving, Writing Effectiveness, and Writing Mechanics. Subscores are assigned on a scale of 1 (lowest) to 6 (highest). Blank responses or responses that are entirely unrelated to the task (e.g., writing about what they had for breakfast) are flagged for removal from results.

Using data from the CLA, CAE used an array of Performance and Analytic Writing Tasks to compare the accuracy of human versus automated scoring. For 12 of the 13 tasks examined, IEA scores agreed more often with the average of multiple experts (r=.84-.93) than two experts agreed with each other (r=.80-.88). These results suggest that computer-assisted scoring is as accurate as—and in some cases, more accurate than—expert human scorers (Elliot, 2011).

How are scorers trained and evaluated?

All scorer candidates undergo rigorous training in order to become certified CLA+ scorers. Scorer training consists of two to three separate sessions and takes place over several days. A lead scorer is identified for each PT and is trained in person by CAE measurement scientists and editors. Following this training, the lead scorer conducts an in-person or virtual (but synchronous) training session for the scorers assigned to his or her particular PT. A CAE measurement scientist or editor attends this training as an observer and mentor. After this training session, homework assignments are given to the scorers in order to calibrate the entire scoring team. All training includes an orientation to the prompt and scoring rubrics/guides, repeated practice grading a wide range of student responses, and extensive feedback and discussion after scoring each response.

After participating in training, scorers complete a reliability check where they score the same set of student responses. 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.

SCALING PROCESS

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

For each PT, raw subscores are summed to produce a total PT score, which is then scaled to a total score, where PT and SRQ tasks are equally weighted at .50 each. Because not all tasks have equal levels of difficulty, raw total scores from the different tasks are converted to a common scale of measurement to reflect comparable levels of proficiency across tasks. Once converted, a given CLA+ scale score indicates approximately 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 average scale score across types.

To convert the raw scores to scale scores, CAE uses a linear scale transformation. This process results in a scale-score distribution with the same mean and standard deviation as the entering academic ability (EAA) as measured by SAT or cross-walked ACT scores of the entering freshmen from the fall 2012 administration of the CLA. This type of scaling preserves the shape of the raw score distribution and maintains students' relative standing on a given 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 ensures that a very high score earned on the task (not necessarily the highest possible score) corresponds approximately to the highest SAT (or converted ACT) score of any freshman who took that task. Similarly, a very low score earned on a task would be assigned a scale-score value that is close to the lowest SAT (or converted ACT) score of any freshman who took that task. However, unlike the SAT scale, the CLA+ scores do not have a ceiling. This means that if a student achieves an exceptionally high or low raw score, this scaling procedure may produce scale scores outside the normal SAT (Math + Verbal) score range of 400 to 1600.

Do scaling equations change with each administration?

Since 2006, CLA scores have only changed with the introduction of major innovations to the assessment. In fall 2010, CLA equations were revised to reflect the establishment of subscores. The new format of the assessment (CLA+), introduced in fall 2013, has also led to new scaling equations to adapt to the new components of the assessment.

Beginning with the fall 2013 administration of CLA+, scaling equations will remain fixed across administrations, and will only be updated if the population of participating institutions has changed enough to significantly alter score distributions.

Schools wishing to compare CLA+ results to prior year results can use the following equation, which was derived by comparing CLA and CLA+ total scores from the 134 institutions that tested students on both the new and old forms of the assessment (r=0.822):

CLA scores from fall 2010 – spring 2013: $score_{CLA+} = 184.188 + (0.812 \cdot score_{CLA})$
CLA scores from before fall 2010: $score_{CLA+} = 268.066 + (0.6897 \cdot score_{CLA})$

VALUE-ADDED SCORING

What do my school's value-added scores on CLA+ mean?

CLA+ includes two forms of growth estimates—value-added scores and effect sizes. Value-added scores compare growth on CLA+ within a school to the growth seen across schools testing similar populations of students, as determined by their EAA and freshman CLA+ performance. Effect sizes reflect the standardized difference in performance between freshmen and other class levels tested, and are described in more detail in the "What do my school's effect sizes for the CLA+ mean?" section.

When the average performance of seniors at a school is substantially better than expected, this school is said to have high "value added." For instance, consider several schools admitting students with similar average performance on general academic ability tests (e.g., the SAT or ACT) and on tests of higher-order skills (e.g., CLA+). If, after four years of college education, the seniors at one school perform better on CLA+ than is typical for schools admitting similar students, one can infer that greater gains in critical-thinking and writing skills occurred at the highest performing school.

Note that a low (negative) value-added score does not necessarily indicate that no gain occurred between freshman and senior year; however, it does suggest that the gain was lower than would typically be observed at schools testing students of similar EAA.

Value-added scores are placed on a normalized (z-score) scale and assigned performance levels. Schools that fall between -1.00 and +1.00 are classified as "near expected," between +1.00 and +2.00 are "above expected," between -1.00 and -2.00 are "below expected," above +2.00 are "well above expected," and below -2.00 are "well below expected." Value-added estimates are also accompanied by confidence intervals, which provide information on the precision of the estimates; narrow confidence intervals indicate that the estimate is more precise, while wider intervals indicate less precision.

Please note that schools must test freshmen in the fall CLA+ window and seniors in the spring window of the same academic year to be eligible for value-added scores. However, if schools have CLA+ data for both groups of students, but have not tested those students in the standard windows (e.g., schools interested in tracking students longitudinally), schools can also use the CLA+ value-added model parameters to create their own growth estimates. Instructions are provided on the following page.

What value-added model does CLA+ use?

CLA+ estimates the value added as the difference between freshman and senior deviation scores through an enhanced regression model known as hierarchical linear modeling (HLM), which accounts for CLA+ score variation within and between schools.

Through spring 2009, CLA estimated value added as the difference between freshman and senior deviation scores through an ordinary least squares (OLS) regression model. Beginning in fall 2009, the CLA moved to the HLM approach. Under the HLM model, a school's value-added score indicates the degree to which the observed senior average CLA+ score meets, exceeds, or falls below expectations established by the senior average EAA score and the average CLA+ performance of freshmen at that school, which serves as a control for selection effects not covered by EAA. Only students with EAA scores—SAT Math + Verbal, ACT Composite, or Scholastic Level Exam (SLE) scores converted to the SAT scale—are included in institutional analyses.

The decision to move from an OLS to HLM model was made after analyses for the CLA showed that the two methods produce similar results. Correlations between the value-added scores resulting from the two approaches were .79 in the 2006-07 administration and .72 in the 2007-08 administration. Reliability estimates, however, were higher for the newer model than the original. Average split-sample reliabilities were .81 (HLM) and .73 (OLS) for 2006-07, and .75 (HLM) and .64 (OLS) in 2007-08. Year-to-year value-added score correlations also increased with the new approach (.58) from the original (.32). The HLM model, therefore, is more efficient because, when the number of tested students is held constant, scores from the new approach are more precise within a given year and are also more realistically stable across years. The HLM model also provides school-specific indicators of value-added score precision, which improve the interpretability of scores.

For more information about the difference between the OLS and HLM models, as well as the rationale for moving to the newer model, please see Improving the Reliability and Interpretability of Value-Added Scores for Post-Secondary Institutional Assessment Programs (Steedle, 2010a).

How can I calculate my value-added score?

Institutions may want to conduct their own analyses in which, for example, they calculate the value-added scores within certain subpopulations of students for whom they have conducted in-depth sampling. To calculate these scores, you need:

- samples of entering and exiting students with both CLA+ and EAA scores. This information is available in the Student Data File, which is distributed to institutions with each administration's results,
- and the estimated parameters for the value-added model, which will be provided in the appendices to institutional reports starting in spring 2014.

Following the completion of the spring administration of CLA+, the table below will be updated to show the estimated parameters for the CLA+ value-added model. Using these estimated parameters and the instructions below, one can compute the expected senior CLA+ score for a given school. In combination with the observed mean score for seniors at that school, this can be used to compute the school's value-added score. These values can also be used to perform subgroup analyses or estimate value-added for groups of students that have been tested longitudinally.

Estimated Parameters for the Value-Added Model					
	γ_{00}	γ_{10}	γ_{01}	γ_{02}	Standard Deviation
Total CLA+ Score	--	--	--	--	--
Performance Task	--	--	--	--	--
Selected-Response Questions	--	--	--	--	--

1. Refer to your CLA+ Student Data File to identify your subgroup sample of interest. The subgroup must contain freshmen and seniors with CLA+ scores and EAA scores.

2. Using your CLA+ Student Data File, compute:

- The mean EAA score of seniors (exiting students) in the sample
- The mean CLA+ score of freshmen (entering students) in the sample
- The mean CLA+ score of seniors (exiting students) in the sample

3. Calculate the senior sample's expected mean CLA+ score, using the parameters from the table above. Please note that the same equation can be used for the individual sections of the CLA+, as well as for the total CLA+ score, by placing the appropriate parameter values from the appropriate row in the table into the equation:

$$\textit{Expected Score} = \gamma 00 + \gamma 01\ (\textit{senior mean EAA}) + \gamma 02\ (\textit{freshman mean CLA+ score})$$

4. Use your expected score to calculate your subgroup sample's value-added score:

$$\textit{Value-added score, unstandardized} = (\textit{Observed senior mean score}) - (\textit{Expected senior mean score})$$

5. Convert that value-added score to standard deviation units:

$$\textit{Value-added score, standardized} = \frac{(\textit{Value-added score, unstandardized})}{(\textit{Standard Deviation})}$$

What do my school's CLA+ effect sizes mean?

Effect sizes represent the amount of growth seen from freshman year, in standard deviation units. They are calculated by subtracting the mean freshman performance at your school from the mean of your sophomore, junior, or senior performance, and dividing by the standard deviation of your freshman scores. Effect sizes do not take into account the performance of students at other CLA+ institutions.

Like value-added scores, effect sizes are only provided for institutions that have tested specific class levels in specific windows—freshmen in the fall, and sophomores, juniors, or seniors in the spring. Schools that have tested students outside of these windows (e.g., those that are assessing students longitudinally) or that want to estimate effect sizes from a different baseline group can easily calculate effect sizes by subtracting the mean performance of one group of students from that of another, and then dividing by the standard deviation of the first group's scores.

ANALYSIS

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

The requirements for including students' results in institutional reporting are dependent upon the type of report an institution is looking to receive.

For cross-sectional (value-added) institutional reports, students must:

- be in the correct class year, as verified by the registrar (freshmen must test in the fall window and sophomores, juniors, or seniors must test in the spring);
- have either an ACT Composite, SAT Math and SAT Verbal/Critical Reading score, or an SLE score;
- and have completed CLA+, which includes submitting a scorable response to the Performance Task, attempting at least half of the Selected-Response Questions, and responding to the CLA+ survey questions.

For single-administration institutional reports with cross-CLA+ comparison, students must:

- have a class standing provided by the school's registrar;
- and have completed the CLA+, which includes submitting a scorable response to the Performance Task, attempting at least half of the Selected-Response Questions, and responding to the CLA+ survey questions.

For institutional reports with one cohort of students and no cross-CLA+ comparisons, students must:

- have completed CLA+, which includes submitting a scorable response to the Performance Task, attempting at least half of the Selected-Response Questions, and responding to the CLA+ survey questions.

On the student level, total scale scores are computed as the weighted sum of the Performance Task (weighted at .50) and the Selected-Response Question scores (weighted at .50). Only students that have provided scorable responses to both sections will receive a total score. However, section scores and subscores will be provided for all students in the institution's Student Data File, where available.

On the school level, each score is the average of scores from those students that have met the criteria outlined above. Note that, during the registrar data collection process, schools can identify students (e.g., those that are part of an in-depth sample) for exclusion from institutional analyses by assigning them a program code.

Does CLA+ analysis account for ceiling effects?

No school-level scores approach the theoretical maximum of scaled CLA+ scores. There are, however, individual students who have achieved a maximum scale score on the CLA or CLA+, as a function of exceptional performance. Originally, CAE capped the distribution of the CLA 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, and CLA+ also does not have a capped distribution.

Does CLA+ correct for range restriction?

Correcting for range restriction is not necessary here because the institution is the unit of analysis, and we don't have a range-restricted population of institutions. Summary statistics of SAT scores for students sampled in the CLA are similar to national figures. Specifically, looking at the 2008 estimated median SAT (or ACT equivalent) of the freshman class across 1,398 four-year institutions in the U.S, we find a minimum of 726, mean of 1057, maximum of 1525, and standard deviation of 127. Across CLA+ schools (fall 2013, n=124) for the same variable we find a minimum of 752, mean of 1031, maximum of 1354, and standard deviation of 110 (College Results Online, 2010).

CONTROLLING FOR EAA

How does CLA+ "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 1560, and so forth (ACT, 2008).

A full crosswalk is printed in the sample institutional report available on our website, and is also provided to participating institutions in both the fall and spring reports. The correlation between ACT Composite and SAT Math + SAT Verbal/Critical Reading has been shown to be as high as .92 (Dorans & Schneider, 1999).

CORRELATIONS WITH OTHER MEASURES

CLA+ has not been correlated with other measures yet. However, given that CLA+ is essentially the CLA assessment with the addition of selected-response questions, the following information might be of interest to the reader.

To what degree is the National Survey of Student Engagement (NSSE) correlated with the CLA?

Correlations between the National Survey of Student Engagement (NSSE) and CLA were explored using data from the CLA feasibility study. Findings were presented at the 2004 annual meeting of the American Educational Research Association, and published in *Research in Higher Education* (Carini, Kuh, & Klein, 2006). The researchers found statistically significant—but small—correlations between CLA outcomes and student engagement scores. Partial correlations between CLA outcomes and student engagement scales were .10 or higher for level of academic challenge, supportive campus climate, reading and writing, quality of relationships, institutional emphases on good practices, and self-reported general education gains. None of the CLA-engagement partial correlations were negative, and they were also slightly higher than GRE-engagement correlations. 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.

Are there linkages or relationships between your test 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 EAA of students?

To date, we have not conducted research to determine whether any linkages or agreements exist between the CLA and various standardized placement tests that would determine an initial freshman course. 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.

RELIABILITY

What is the reliability of CLA+?

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

Performance Tasks are scored through a combination of automated and human scoring. More specifically, each PT is double-scored—once by a machine using Latent Semantics Analysis and once by a trained human scorer. The degree of agreement between scorers is known as the inter-rater reliability or inter-rater correlation. Scores close to 1 indicate high agreement, whereas scores close to 0 indicate little or no agreement. The inter-rater correlation was used as the reliability coefficient for the PT, whereas Cronbach's alpha was utilized for the SRQs. Cronbach's alpha measures the internal consistency of a set of items and can range from 0 to 1. Values closer to 1 indicate higher reliability; values closer to 0 indicate lower reliability. Table 1 shows the reliability statistics for the different components of CLA+.

Table 1: Reliability indices for CLA+ Sections

CLA+ Section	Reliability
Total CLA+	.81
Performance Task	.77
Selected-Response Questions	.76
Scientific & Quantitative Reasoning	.51
Critical Reading & Evaluation	.58
Critique an Argument	.52

Reliability for the PT (*r*= .77) is comparable to the reliability for the SRQs (α=.76). Stratified alpha (Cronbach, Schonemann, & McKie, 1965) was used to combine the PT with the SRQs, resulting in a reliability coefficient of .81. Previous research has indicated that CLA scores have been very reliable at the institution level (α =.80) (Klein, et al., 2007), but not at the individual student level (α = .45). However, with the addition of SRQs to the PT, the reliability of individual student scores is high enough to ensure the appropriateness of making interpretations at the individual student level and for making inferences in regard to grading, scholarships, admission, or placement.

VALIDITY

Do you have any evidence of construct validity?

In the fall semester of 2008, CAE (CLA) collaborated in a construct validity study with ACT (CAAP) and ETS (MAPP) to investigate the construct validity of these three assessments (Klein et al., 2009). Construct validity refers to whether an assessment measures the particular skill (i.e., construct) that it purports to measure and is often evaluated by examining the pattern of correlations between a test and other tests of similar and different skills (Campbell, 1959). For example, if the CLA measures critical thinking skills, then it should be highly (positively) correlated with other tasks that measure critical thinking.

Results from the study show that for critical thinking, the CLA is indeed strongly positively correlated with other tasks that measure critical thinking. The correlation between CLA Performance Tasks and other tests of critical thinking range from .73 to .83. The correlation between CLA Critique-an-Argument tasks and other constructs that measure critical thinking range from .73 to .93. A full report of the Test Validity Study (Klein et al., 2009) can be found on CAE's website, http://www.cae.org/content/pdf/TVS_Report.pdf.

Information on the construct validity of CLA+ will be reported in the near future.

What about the face validity of your measure?

A test is said to have face validity when, on the surface, it appears to measure what it claims to measure. For CLA+ to have face validity, CLA+ tasks must emulate the critical thinking and writing challenges that students will face outside the classroom. These characteristics of CLA+ were vetted by a sample of judges who participated in the CLA+ standard setting study.

After reviewing CLA+ tasks in depth and reading a range of student responses, these judges completed a questionnaire to express their perceptions of the tasks.

As shown in Figure 1, results indicate that the judges perceived the CLA+ tasks to be good assessments of critical-thinking, writing, analytic-reasoning, and problem-solving skills. Responding on a 1-5 scale, judges felt, for example, that CLA+ measures important skills that college graduates should possess (Mean 4.92, SD 0.29); students need good analytical reasoning and problem solving skills to do well on the task (Mean 4.75, SD 0.45); students need good writing skills to do well on the task (Mean 4.08, .90) and students who do well on the task would also perform well in a job requiring good written communication (Mean 4.20, SD 0.78) or analytic reasoning and problem solving skills (Mean 4.75, SD 0.45). Respondents also agreed, after viewing the tasks, that successful performance on CLA+ may help students compete in a global market (Mean 4.40, SD 0.70).

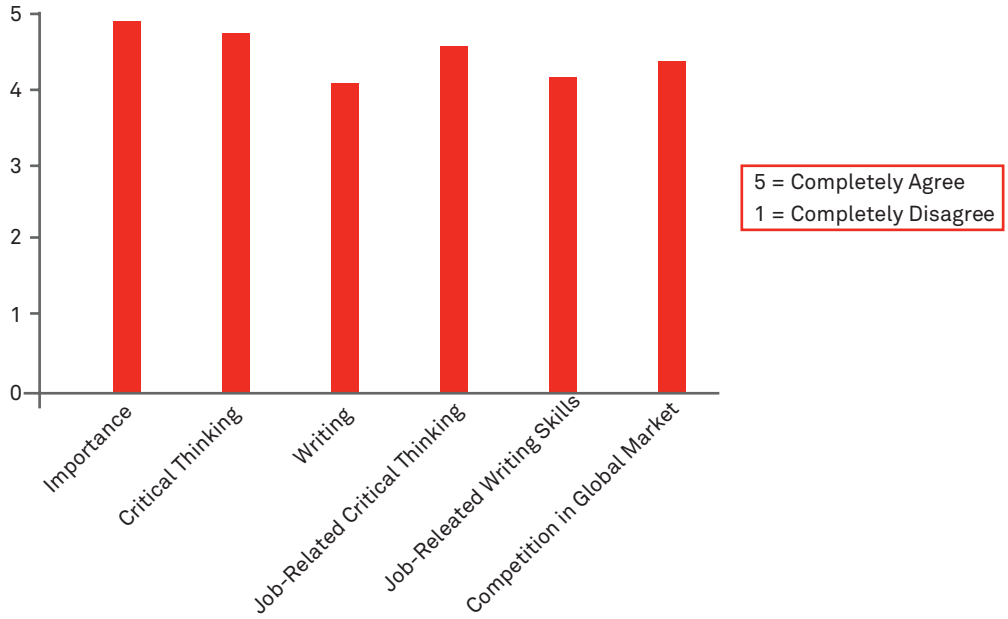


Figure 1: Average face validity assessments of the CLA+

How are cut scores determined?

On December 12, 2013, a standard setting study was conducted to formally establish fair and defensible levels of mastery for CLA+. 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, 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.

CAE recruited a panel of 12 subject matter experts based upon industry standards (Jaeger, 2005). The participants in this study, representing various sectors of both higher and secondary education and content experts, either supervise or work with students/new graduates.

CAE employed the bookmark(Lewis, Mitzel, Green, & Patz, 1999) methodology to establish the three different cut scores for four different levels of mastery: Below Basic, Basic, Proficient, and Advanced. Under the bookmark method, the CLA+ SRQ items and PT responses are arranged in order of difficulty and the expert judges on the panel are individually asked to pick the point at which, using the SRQs as an example, a basic/proficient/advanced student would answer this item correctly. So if a judge thought that out of 25 items, a basic student would answer the first seven items correctly, a proficient student would answer the first 14 items correctly, and an advanced student would answer the first 21 items correctly, their scores would be 7, 14, and 21. The overall cut scores for each section and each level of mastery is computed using the average across all 12 panel participants.

STUDENT EFFORT

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

CLA+ does not control for self-reported student effort, but has conducted some research on the role that motivation plays in CLA+ achievement. Analyses of the relationship between Performance Task scores and self-reported effort suggest that, controlling for EAA, student effort only explains about three to seven percent of the variance in school-level scores (Klein, Benjamin, Shavelson, & Bolus, 2007).

Additional research, presented at the 2010 Annual Meeting of the American Educational Research Association, focused on the relationship between incentives, motivation, and CLA performance. Using the Student Opinion Survey (SOS)—a motivation scale that measures a student's effort and belief that performing well is important—CAE found that (after controlling for average EAA) motivation was a significant predictor of CLA scores on the student level, but not on the school level (Steedle, 2010b).

Tying stakes to an assessment has also been shown to increase motivation and—in turn—test scores, based on analyses of college students' performance on the ETS Proficiency Profile (Liu, Bridgeman, & Adler, 2012). Students who were informed that their scores might be shared with faculty at their college or with potential employers performed better than students who were told that their results would be averaged with those of their peers before possibly being shared with external parties. Both of these groups of students performed better than those who were informed that their results would not be seen by anyone outside of the research study.

Because CLA+—unlike its predecessor, the CLA—is reliable at the student level, stakes can be tied to student performance to increase motivation and improve outcomes. To increase motivational opportunities, CAE has embedded results-sharing components into the assessment, delivering electronic badges to students who have performed at or above the proficient level on CLA+, and entering into partnerships will online transcript services and virtual job fairs to allow high-performing students to share their results.

Student Effort and Engagement Survey Responses

Tables 2 and 3 are the summarized results for the questions from the student survey that was administered to participants following the completion of the CLA+ assessment.

Tables 2 and 3 show that 53.6% put more than a moderate amount of effort into their CLA+ responses and 66.2% of students found the tasks to be moderately to extremely engaging. These results are encouraging because low student motivation and effort are construct-irrelevant threats to the validity of test score interpretations. If students are not motivated, their scores will not be accurate reflections of their maximum level of competency. Although these responses are self-reported, the validity of CLA+ should be enhanced given that stakes are attached to the assessment. Previous research suggests that student motivation and performance is improved as a direct function of attaching stakes to an assessment (Liu, Bridgeman, & Adler, 2012).

Table 1: Reliability indices for CLA+ Sections

How much effort did you put into these tasks?	
No effort at all	1.6%
A little effort	8.4%
A moderate amount of effort	36.4%
A lot of effort	32.8%
My best effort	20.8%

Table 3: Engaging

How engaging did you find the tasks?	
Not at all engaging	11.3%
Slightly engaging	22.5%
Moderately engaging	39.2%
Very engaging	22.2%
Extremely engaging	4.8%

Face Validity

Students were asked about their perceptions in terms of how well CLA+ measures writing and analytic reasoning and problem solving skills (Table 4).

In an attempt to establish face validity for CLA+, the tasks are designed to emulate critical-thinking and writing tasks that students will encounter in nonacademic endeavors.

As shown in Table 4, results indicate that the students perceived the tasks to be moderately to extremely good assessments of writing (77.5%) and analytic reasoning and problem solving skills (79.8%) for the Performance Tasks, and analytic reasoning and problem solving skills (70.9%) for the SRQs.

Table 4: Face Validity

How well do you think these tasks measure the following skills:	Writing - PT	Analytic Reasoning and Problem Solving - PT	Analytic Reasoning and Problem Solving - SRQ
Not well at all	5.4%	4.2%	6.8%
Slightly well	17.2%	15.9%	22.3%
Moderately well	45.2%	39.9%	44.0%
Very well	27.3%	31.5%	22.5%
Extremely well	5.0%	8.4%	4.4%

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+ PTs (Table 5) and between SAT scores and time spent on CLA+ tasks. This relationship is not surprising given that the average test time for tasks in minutes (Table 6) was moderate. Well-developed responses are expected to take longer to compose, although it is possible that students can achieve a high score with a brief response. Table 6 also indicates that students did not invest much time in the SRQs and consequentially, a low correlation is observed between the time spent on SRQs and the total score (Table 5).

Table 5: Relationship between Time Spent on CLA+ sections and CLA+ Total Scores

	Time SRQ	Time PT	Total Time	Total Score
Time SRQ	1			
Time PT	.251	1		
Total Time	.594	.928	1	
Total Score	.121	.307	.302	1

Table 6 shows the average testing time for each component of CLA+. Results indicate that on average students finished the different components of the assessment with some time remaining in each section.

Table 6: Test Time for Tasks in Minutes

	Mean	Standard Diviation
Time Spent PT	33.11	14.90
Time Spent SRQ	20.68	6.92
Total Test Time	53.80	17.93

REFERENCES

ACT. (2008). ACT-SAT Concordance. Retrieved July 21, 2011, from http://www.act.org/aap/concordance/

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards of Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Campbell, D. T. (1959). *Convergent and discriminant validation by the multitrait-multimethod matrix*. Psychological Bulletin, 56(2), 81-105.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). *Student Engagement and Student Learning: Testing the Linkages*. Research in Higher Education, 47(1), 1-32.

Cronbach, L. J., Schonemann, P., & McKie, D. (1965). *Alpha coefficients for stratified-parallel tests*. Educational and Psychological Measurement, 25, 291-312.

Dorans, N. J., & Schneider, D. (1999). *Concordance between SAT I and ACT scores for individual students* Research Notes, College Entrance Examination Board.

Elliot, S. (2011). *Computer-assisted scoring for Performance tasks for the CLA and CWRA.* New York: Council for Aid to Education.

Jaeger, R. R. (2005). *Selection of judges for standard setting*. Educational Measurement: Issues and Practice, 10(2), 3-14.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). *The Collegiate Learning Assessment: Facts and Fantasies*. Evaluation Review, 31(5), 415-439.

Klein, S., Liu, O. L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., . . . Steedle, J. (2009). Test Validity Study (TVS) Report. Supported by the Fund for the Improvement of Postsecondary Education. from http://www.cae.org/content/pdf/TVS_Report.pdf

Lewis, D. M., Mitzel, H. C., Green, D. R., & Patz, R. J. (1999). *The Bookmark standard setting procedure*. Monterey: McGraw-Hill.

Liu, L., Bridgeman, B., Adler, R. (2012). *Measuring learning outcomes in higher education: motivation matters*. Educational Researcher 2012 41(9), 352-362.

Steedle, J. T. (2010a). *Improving the Reliability and Interpretability of Value-Added Scores for Post-Secondary Institutional Assessment Programs*. Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Steedle, J. T. (2010b). *Incentives, Motivation, and Performance on a Low-Stakes Test of College Learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.