CAE

Does College Matter?

Measuring Critical-Thinking Outcomes Using the CLA

January 2013

This research was conducted by the Council for Aid to Education (CAE), supported by internal funds

While there is disagreement regarding the desired outcomes of post-secondary education, most agree that critical thinking is among the important outcomes of the college experience. A value-added analysis of the assessment results obtained from the colleges and universities using the CLA conducted by CAE (Council for Aid to Education) in 2012 indicates that, by and large, there is significant growth in critical thinking skills. This stands in contrast to other studies suggesting limited growth between the freshman and senior years of college. Regardless of the research results found, the ongoing measurement of critical thinking is essential to the continued understanding and improvement of institutional learning and development.

Introduction

The title of this paper — Does College Matter? — asks whether there is evidence that students' set of real-world skills grow during their attendance in college (Pascarella and Terenzini,(2005). While this question has been of interest since the foundation of the Academy, it has recently come into increased focus, with concerns over rapidly rising tuition rates and student loan defaults. At the same time, the business community is wary of the skills students acquire in college, suggesting that in many cases students do not possess the necessary strengths and talents to be successful in the workplace. In short, educators, politicians, and the public at large are asking whether college is "worth it."

"...the business community is questioning the skills students acquire in college, suggesting that in many cases students do not possess the necessary skills to be successful in the workplace."

To fully address this issue, we need to first develop an understanding of what it means to matter—that is, what are the potential outcomes of a college education. At the risk of oversimplification, there are fundamentally two "camps" regarding outcomes. On the one hand there are those that emphasize the unique disciplinary knowledge and skills arising out of the major area of study. On the other, there are those who call for achieving outcomes that are common to all disciplines and seen as important regardless of area of study.

But even among those who assert that there is a common set of cross-disciplinary outcomes that should result from the college experience, there is disagreement as to what those critical-thinking skills should be. There are those who suggest that the most important outcomes of the college experience are moral; that is to produce ethical members of society with a clear philosophical understanding of right and wrong is often cited as the target here. Still others cite a range of effective outcomes as the primary goals of post-secondary education.

There is widespread agreement that critical-thinking skills, such as analysis of information, reasoning from text, and quantitative reasoning will enable students to be knowledgeable consumers of information, strong problem solvers, and effective decision makers (Shavelson, 2009).

It is this set of critical-thinking skills, along with communication, that are measured by the CLA, which is the assessment used to draw conclusions presented in the paper. A full treatment of the range of college outcomes is well beyond this paper; we raise it here at the outset solely to clarify the outcomes which we address in this piece and to place those outcomes in a broader context.

What Do We Mean by Critical Thinking?

Critical-thinking skills are an important "slice" of the generic, cross-disciplinary skills that can be achieved in college. These are the set of skills required for success in the workplace, higher academics, and life in general. While transcending specific disciplines or domains of knowledge, these skills are both learnable and teachable. (Indeed, the evidence presented in this paper suggests that this is the case.)

This set of skills sits firmly between those which are fixed (or at least enduring) that we generally refer to as intelligence (often seen as related to Spearman's concept of general intelligence, or "g") and the vast array of domain or discipline-specific skills.

For a more complete discussion of critical-thinking skills, consult Pascarella and Terenzini (2005) and Shavelson,(2009).

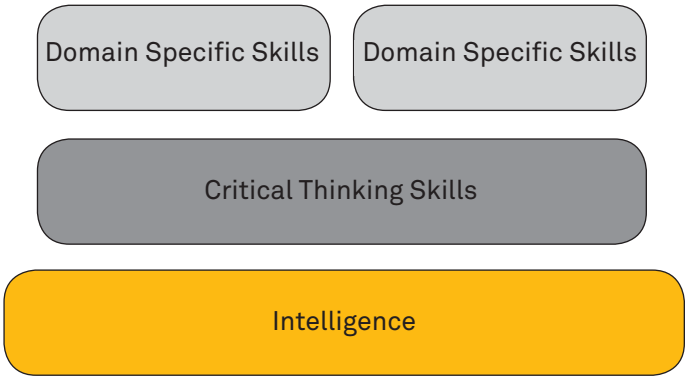

Figure 1: Critical Thinking Model

About the CLA

The CLA measures an array of critical-thinking skills, including the analysis and evaluation of information, analytic- and quantitative-reasoning skills, problem-solving, and written communication.

What distinguishes the CLA from other purported assessments of critical thinking is the structure of the test and assessment approach used. The first 60 minutes consist of an integrated performance task that mirrors a real world scenario that could be encountered in a work (or academic) environment. The student is provided with three or more authentic documents, such as a data table or graph, a newspaper article, a research report, or other critical information sources that students would typically encounter in a real world setting. Students are asked to analyze the information presented and then are given a series of questions that ultimately requires them to make a decision or solve a problem.

> "The CLA measures an array of critical thinking skills including the analysis and evaluation of information, analytic and quantitative reasoning skills, problem-solving, and written communication."

The second 30 minutes of the test provide a series of problem sets assessing quantitative reasoning and reasoning from text. Each problem set provides a scenario and authentic source documents, with 5-10 selected response type questions to be answered by the student for each scenario to gather further evidence of the student's critical-thinking skills.

The reliability and validity of the CLA

There is considerable evidence to support the reliability and validity of the CLA collected over the past decade.

Cronbach's alpha, which ranges from 0 to 1, measures the internal consistency of a set of items. (Values closer to 1 indicate higher reliability.) The reliability for the CLA+ obtained from field testing ranged from .85-.87.

With respect to validity, several studies (Klein, Benjamin, Shavelson, and Bolus, 2007) have shown that the CLA is an effective measure of critical thinking. The CLA correlates very highly (.80+) with other measures of critical-thinking skills (e.g., Academic Profile, CAPP). Moreover, the instrument has been shown to be sensitive to growth in college environments.

Critical Thinking Skill Growth During College

For the past decade, colleges and universities have been measuring critical-thinking skills gained in college using the CLA. About 700 schools have measured the level of critical-thinking skills when entering and again when exiting school. Most of these schools have conducted these measurements to obtain an estimate of the "value added" for colleges—the contribution to student growth attributable to the college experience.

Measuring Growth

There are two primary research designs that have been used to measure value added (growth) during the college experience. One way to measure growth is to assess the critical-thinking skills of a cohort of students at the beginning of college and then to measure that same cohort of students at the end of college. This "measure them at the beginning and measure them at the end" has an intuitive appeal but presents several

challenges. This requires waiting at least four years for answers and can be affected by student attrition.

An alternative to this cohort assessment approach is to measure a random sample of freshman and senior students at the same time. This produces similar results, statistically, to the measurement of a single cohort over time. This requires acceptance of certain assumptions such as the similarity between the freshman and senior classes and the reliance on random sampling, but is often preferred because it is a faster and more cost-effective way to get answers to the value-added question.

How much growth do we see?

In the 2011-12 academic year, approximately 158 post-secondary institutions administered the CLA to estimate the growth (value added) between the freshman and senior years of college. The change in the average CLA score between the freshman and senior years was calculated for each participating institution. We then calculated the effect size, a commonly used measure that allows comparisons over time, across studies and across measures.[1]

> "On average...we found that the effect size for student growth between the freshman and senior year was .78... on average, a school whose freshmen were at the 50th percentile coming into college grew to the 78th percentile"

On average (mean effect size across all institutions) we found that the effect size for student growth between freshman and senior years was .78. This means that, on average, students grew about three quarters of a standard deviation. As shown in Figure 2, the average school had a freshman score of 1046 and a senior score of 1154, with an average gain of 108 points. CLA scores are reported on a scale that ranges from approximately 400 to 1600.

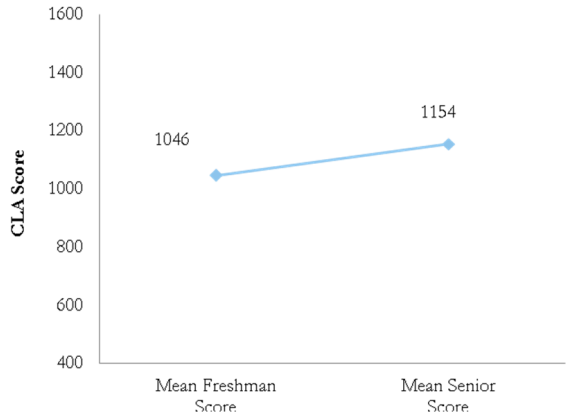

Figure 2. Mean CLA Scores of Schools Testing Freshmen and Seniors, 2011-12

An advantage of using effect size is that it can be readily and accurately interpreted in terms of average percentile gain. In this study, on average, a school whose freshmen were at the 50th percentile coming into college grew to the 78th percentile (for freshman) by the senior year. This is a significant amount of growth.

Similar effect-size patterns are seen across years, providing further evidence that college students exhibit significant growth on the critical-thinking skills measured by the CLA. Schools participating in the CLA in academic years prior 2011-12 showed levels of growth consistent with the average effect size seen in the 2011-12 academic year. The average effect size across all years combined (a sample of 1,307 cross-sectional cohorts) was also .73. An effect size of .73 indicates a 27-point percentile gain from the 50th percentile. Figure 3 shows the distribution of effect sizes among the cross-sectional cohorts that tested between 2005-06 and 2011-12.

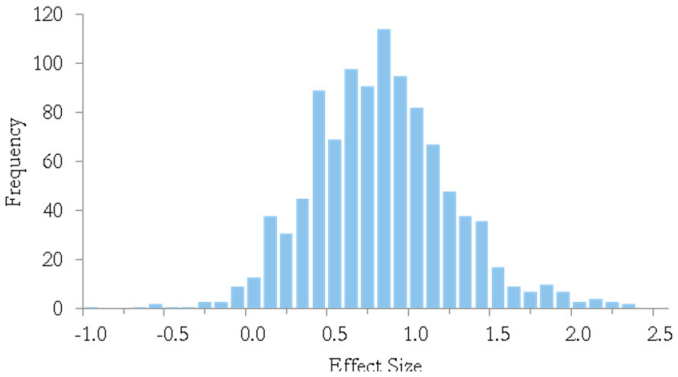

Figure 3. Distribution of Effect Sizes Across Schools, 2005-12 (N=1,037)

Colleges show similar levels of growth regardless of sector (public v. private), institution size, Carnegie Classification, selectivity, or minority-serving status. Though some types of schools show slightly higher growth on average, none of the categories of schools mentioned above had average effect sizes below .62.

Figure 4 below shows the distribution of effect sizes for private versus public schools among institutions that tested during the 2005-06 through 2011-12 academic years, and exemplifies the similarities in growth seen across different school types. Public universities tend to show slightly higher average growth (.77) and a little more variation within effect sizes than private universities (.69), though the differences in the two sectors' mean effect sizes are not statistically significant within individual academic years.

Figure 5 shows the distribution of effect sizes for institutions that are minority serving alongside the distribution of schools that are not minority serving[2]. Minority-serving schools have higher effect sizes, on average (.78), than other colleges and universities (.73), though these differences in average effect size are not statistically significant. Minority-serving institutions also tend to show greater variation in growth levels than their peers. The greater variation in performance among minority-serving institutions may be attributable to this institution type

1 This is a conversion of the test score to standard deviation units, with a score of "1" meaning one standard deviation in growth.

2 Minority-serving institutions include Historically Black Colleges and Universities and schools with student populations that are 25% or more Hispanic.

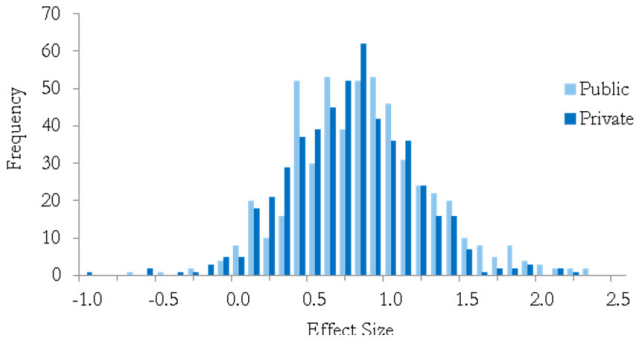

Figure 4. Distribution of Effect Sizes by School Sector, 2005-12

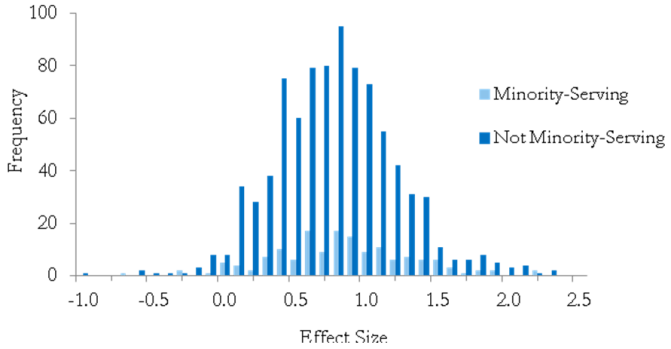

Figure 5. Distribution of Effect Sizes by Minority-Serving Status, 2005-12

making up a small proportion of the overall sample; less than 15% (151) of the 1,037 academic cohorts that tested between 2005-12 came from minority-serving schools.

Figure 6 shows the distribution of effect sizes for small, midsized, and large schools (with school size determined by the number of full-time equivalent students.) Midsized schools have a larger average effect size (.77) than small or large schools (.72 and .71, respectively). They also vary somewhat more in performance, though these differences are also not statistically significant.

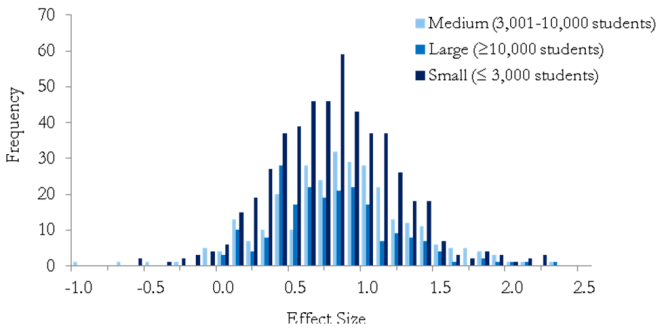

Figure 6. Distribution of Effect Sizes by School Size, 2005-12

There are, however, highly statistically significant differences in average effect size by Carnegie Classification (p<.001). Baccalaureate institutions show more average growth (.81) than master's colleges and universities (.72), which in turn show more average growth than doctoral/research institutions (.63). Doctorate-granting universities also show less variation in effect sizes than master's or bachelor's institutions. These differences were consistent across academic years, as well. Figure 7 shows the distribution of effect scores for the three Carnegie Classification levels.

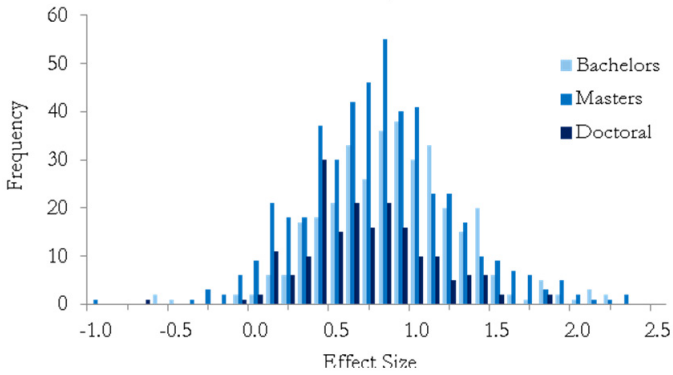

Figure 7. Distribution of Effect Sizes by Carnegie Classification, 2005-12

There are also highly statistically significant differences in average growth seen according to institutional selectivity (p<.001). Non- to Less Competitive institutions (those with Barron's 28th edition values of 1-2) show the highest average effect size (.83), compared to Competitive to Competitive Plus institutions (Barron's values of 3 and 3.5; average effect size of .76), which in turn have a higher average effect size than Very to Most Competitive institutions (Barron's ratings of 4 or higher; average effect size of .62). These distributions are shown in Figure 8. The least competitive institutions also show slightly more variation in effect sizes, with the most competitive institutions performing more similarly to each other.

> "Colleges show similar levels of growth regardless of sector... Though some types of schools had slightly higher growth on average, none... had an average effect size below .62"

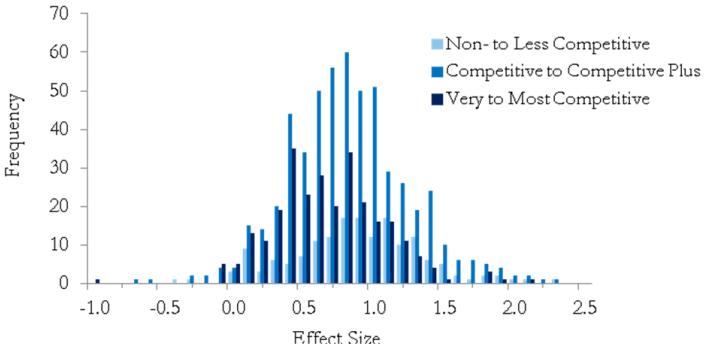

Figure 8. Distribution of Effect Sizes by Selectivity, 2005-12

This stands in contrast to the findings of Academically Adrift (Arum and Roska, 2011) who also examined student growth using the CLA. They suggest that there is little growth in critical thinking as measured by the CLA. They report an effect size of .18, or less than 20% of a standard deviation. However, Arum and Roska used different methods of estimating this growth, which may explain the differences in growth shown here with that reported in Academically Adrift. For instance, the level of

growth they report is across the first two years of college, not the entire four-year college experience. The Arum and Roska study is also limited to a small sample of schools that are a subset of the broader group of institutions that conduct value-added research using the CLA, and so may not be representative of CLA growth in general.

Measuring Senior Outcomes: The Competency Based Model[3]

In addition to within-school growth from freshman to senior year, the CLA also measures the absolute scores achieved by students in their final semester of college. Across the 2005-06 through 2011-12 academic years, the average institutional senior sample scored 1173 on the CLA, which suggests that, on average, colleges are graduating seniors with the critical-thinking and written-communication skills needed to succeed beyond college.

To give context to this score, students receiving total scores in the high 1100s on the CLA typically provide responses that receive subscores of or close to 4. A subscore of 4 might indicate that the student: provides some careful analysis with a clear position and some sound reasons in support of that position; demonstrates an accurate understanding of the task content, though disregarding some information; organizes the response in a way that makes arguments and logic apparent but not obvious; demonstrates good control of grammatical conventions with few errors; and writes well-constructed sentences with some varied structure and length. Subscores of 1 (the lowest possible score) indicate minimal evidence of the subscore skills (analytic reasoning and evaluation, writing effectiveness, writing mechanics, and problem-solving), and subscores of 6 (the highest possible score) indicate outstanding evidence of these skills.

> "... the average institutional senior sample [score] ...suggests that, on average, colleges are graduating seniors with the critical thinking and written communication skills needed to succeed beyond college."

As with growth estimates, the distribution of institutional senior CLA scores is bell-shaped, with most schools' seniors performing close to the average—about half of all schools scored between approximately 1100 and 1250—though with a trail of schools lagging behind the group and a handful of schools exhibiting performance well above average (see Figure 9).

Public- and private-sector schools' seniors perform similarly on the CLA, with private-sector seniors scoring slightly higher on average (1183) than public-sector seniors (1165), though private colleges and universities also have a somewhat larger variation in senior scores than public schools. While the difference in average senior scores between the two sectors is small, it is also statistically significant (p=.001). Figure 10 shows the distribution of senior CLA scores by sector, for those institutions that tested between 2005-06 and 2011-12

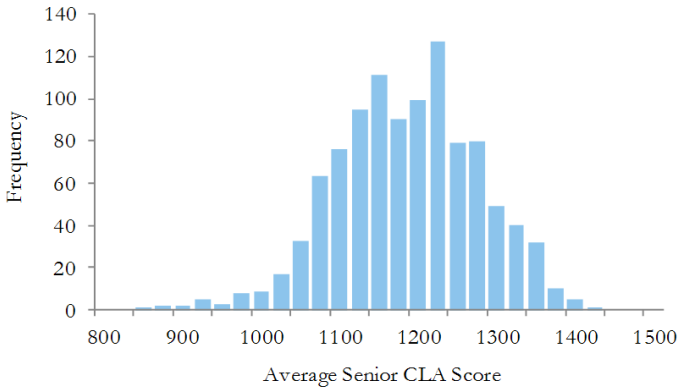

Figure 9. Distribution of Mean CLA Scores for All Schools Testing Seniors, 2005-12

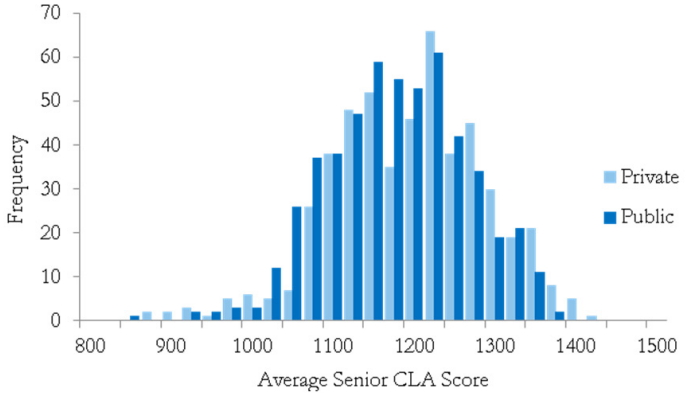

Figure 10. Distribution of Mean Senior CLA Scores by Sector, 2005-12

Figure 11 shows the distribution of average senior scores for minority-serving institutions alongside schools that are not minority serving. Minority-serving schools whose seniors have taken the CLA have a considerably lower average score (1100) than the average cohort of seniors at non-minority CLA schools (1186), and those differences are statistically significant (p<.001). Both types of schools have similar levels of variance in their scores.

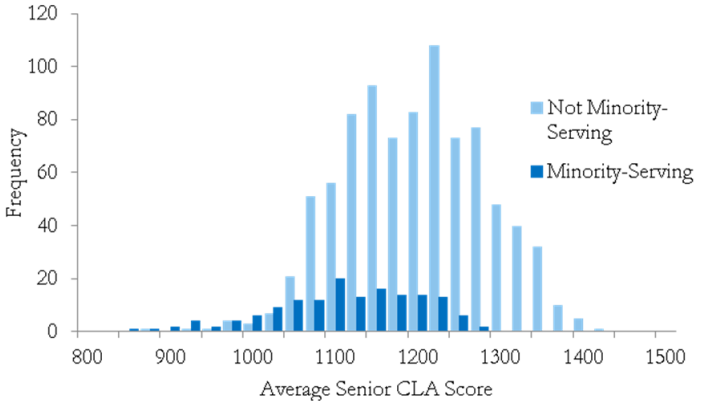

Figure 11. Distribution of Mean Senior CLA Scores by Minority-Serving Status, 2005-12

3 This model may be used by undergraduate programs that use the competency-based approach to certification for graduation rather than accumulation of a certain number of credits. The institution offers the student the opportunity to demonstrate that they meet the minimum level of proficiency required for the academic program the student is enrolled in. Absolute scores are focused on rather than value added or growth from freshman to senior levels. This model may be of most interest to undergraduate programs that are largely online in nature.

The differences in average senior CLA scores between small, medium, and large institutions is also statistically significant (p<.001), with larger institutions' seniors performing approximately 40 and 50 points higher (1206) than their peers at small- (1169) and medium-sized institutions (1157), respectively. The variation in scores was similar across the small- and medium-sized institutions, while large schools testing seniors on the CLA performed similarly to one another (see Figure 12).

When comparing schools by Carnegie Classification, seniors at doctorate-granting universities on average perform about 40 points better on the CLA (1213) than baccalaureate colleges (1175), which in turn perform about 20 points better on average than seniors at master's colleges and universities (1156, p<.001). Figure 13 shows the distribution in senior CLA performance for these institution types, with bachelor's institutions having a relatively wide variation in average senior scores and master's institutions performing similarly to each other, showing lesser variation in senior CLA scores. The differences in average senior scores for these three school types are significantly different from each other.

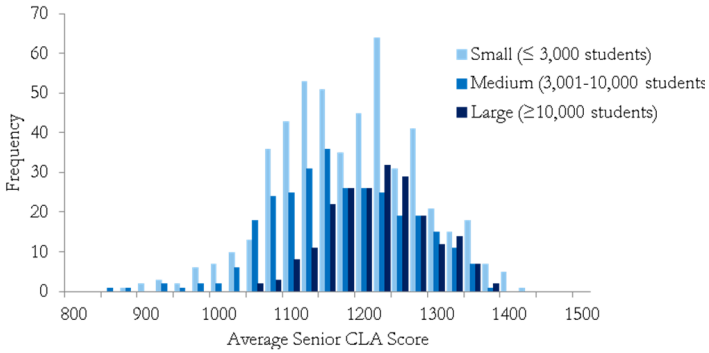

Figure 12. Distribution of Mean Senior CLA Scores by Institution Size, 2005-12

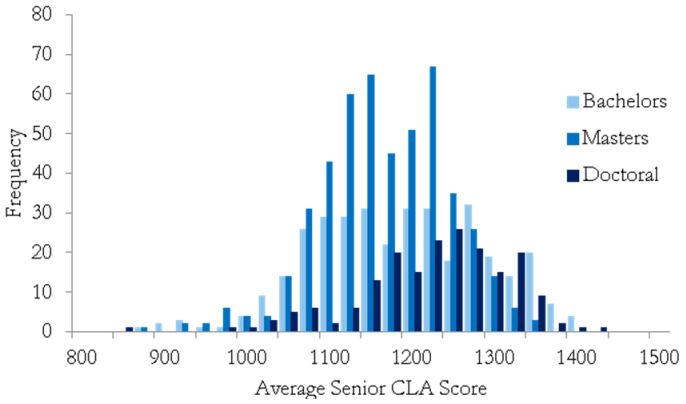

Figure 13. Distribution of Mean Senior CLA Scores by Carnegie Classification, 2005-12

Figure 14 shows the distribution of average senior CLA scores by school selectivity, where there are also statistically significant differences (p<.001). As one would expect, Very to Most Selective institutions' seniors score much higher—about 150 points—on average (1254), than Non- to Less Competitive institutions (1102). Competitive and Competitive Plus institutions, on average, perform in between, though considerably below their more selective peers (1158). All three school types show low variation in scores, showing that schools within a given selectivity category tend to perform similarly to other schools within the same category in terms of average senior scores.

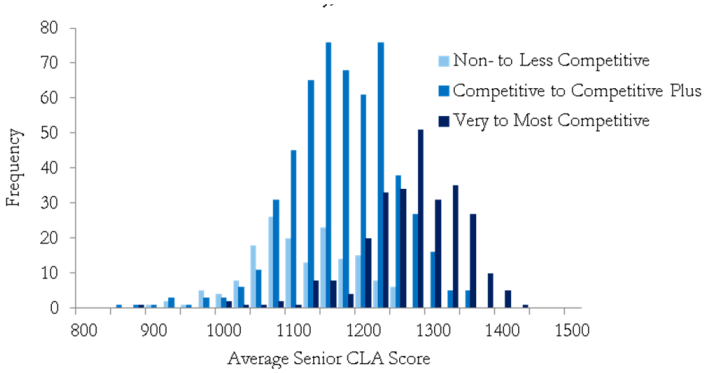

Figure 14. Distribution of Mean Senior CLA Scores by Selectivity, 2005-12

The average total scores seen across CLA institutions show that these schools are, on average, graduating seniors with critical-thinking and written-communication skills—both across school types, as well as within the group as a whole. We also see that certain categories of schools tend to graduate students whose have a better grasp on these skills than their peers; just as we might expect (as suggested by their performance on the CLA), more-selective and doctoral institutions graduate seniors with greater attainment of these skills relative to seniors at other post-secondary institutions. There is, however, considerable variation in performance on the CLA both across all schools and within certain categories of schools, suggesting that not all schools are graduating students of the same caliber when it comes to the skills that the CLA measures.

> "'What you measure is what you get' It is only through a sustained program of critical thinking assessment and research that we will produce the thoughtful college graduates the current global knowledge economy so desperately needs."

Conclusion

In short, we conclude that college does matter. Students show considerable growth in critical-thinking skills between the freshman and senior years of college, as measured by the CLA. Likewise, on average, schools are graduating seniors whose level of performance on the CLA matches what one would expect, given institutional selectivity and Carnegie Classification, for example. As a whole, CLA schools provide evidence that seniors are graduating with the critical-thinking and written-communication skills that are vital for success beyond college. But the answer may well be less important than the question.

We must continue to investigate whether we should expect a greater impact on important skills such as critical thinking. And, research should move beyond this fundamental question, to explore the factors that impact student growth in critical thinking. There are an array of collegiate environmental variables, teaching variables, and student variables that need to be examined to gain a full understanding of what matters.

The need for ongoing measurement of students' critical-thinking skills is perhaps the most important conclusion we can draw. We would do well to heed the often heard refrain in the educational community: "What you measure is what you get." It is only through a sustained program of critical-thinking assessment, research, and instruction that we will produce the thoughtful college graduates the current global Knowledge Economy so desperately needs.

Find Out More

There is much written about college outcomes and critical thinking. These three resources are an excellent starting place if you want to learn more.

Klein, S., Benjamin, R., Shavelson, S. & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review* 31(5):415-439.

Pascarella and Terenzini, P.T. (2005) How College Affects Students: Volume II A third decade of research. San Francisco: Jossey Bass

Shavelson, R. (2009). *Measuring College Learning Responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.

This research was conducted by the Council for Aid to Education, a nonprofit assessment and research firm located in Manhattan, New York. The research was supported by internal research funds.

References

Arum, Richard and Roska, Josipa (2011) Academically Adrift. Chicago: University of Chicago Press.

Barrie, S. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education*, 51, 215–241.

Benjamin, R., Chun, M. & Shavelson, R. (2007). *Holistic tests in a sub-score world: The diagnostic logic of the Collegiate Learning Assessment*. New York: CAE.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. New York: Houghton Mifflin

Clanchy, J., & Ballard, B. (1995). Generic Skills in the Context of Higher Education. *Higher Education Research and Development*, 14(2), 155-166.

Green, W., Hammer, S., & Star, C. (2009). Facing up to the challenge: why is it so hard to develop graduate attributes? *Higher Education Research & Development*, 28(1), 17-29.

Klein, S., Benjamin, R., Shavelson, S. & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review 31*(5):415-439.

Klein, S., Chun, M. Hamilton, L. Kuh, G. & Shavelson, R. (2005). An approach to measuring cognitive outcomes across higher education institutions. *Research in Higher Education, 46*(3), 251-276.

Klein, S., Freedman, D., Shavelson, R. Bolus, R. (2008). Assessing school effectiveness, *Evaluation Review 32*(6): 511-525.

Martinez, M. E. (2000). *Education as the cultivation of intelligence*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pascarella and Terenzini, P.T. (2005) How College Affects Students: Volume II A third decade of research. San Francisco: Jossey Bass

Rychen, D.S.E., & Salganik, L.H.E. (2003). *Key competencies for a successful life and a well-functioning society*. Cambridge, MA, USA: Hogrefe & Huber Publishers.

Shavelson, R. & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change 35*(1), 10-19.

Shavelson, R. (2008). The Collegiate Learning Assessment. *Forum for the Future of Higher Education*. Cambridge, MA: Ford Policy Forum.

Shavelson, R. (2009). *Measuring College Learning Responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.

Shavelson, R. (2011). An approach to testing and modeling competence. *Paper presented at the Conference on Modeling and Measurement of Competencies in Higher Education*. Berlin: Germany.

Spearman, C. (1904). "General intelligence" objectively determined and measured. American Journal of Psychology, 15, 201–293.