THE TEAGLE ASSESSMENT PROJECT
A STUDY OF THE LEARNING OUTCOMES FOR MAJORS IN CLASSICS

Michael Arnush, Rachelle L. Brooks, and Kenneth Scott Morrell

ABSTRACT
This presentation will discuss a project to assess the cognitive outcomes of undergraduate students who major in classics, which has been underway since 2005. During the first phase of the project (2005-2007), "Measuring Undergraduate Cognitive Outcomes from a Disciplinary Perspective," a group of faculty members in classics representing both undergraduate and research institutions, in consultation with a social scientist, identified a set of intellectual skills associated with the study of classics. Broadly speaking, these skills fall into categories of critical thinking and postformal reasoning. The group then collaborated on the design of an instrument to measure those outcomes. This instrument comprised essay questions, portions of the Sternberg Triarchic Abilities Test, and a survey of demographic and educational information. In the fall of 2006 the project administered the instrument to a group of forty-six undergraduate students and developed an evaluative rubric.

This work formed the basis for a second phase (2008-2013), "A Longitudinal Study of Critical Thinking and Postformal Reasoning: Assessing Undergraduate Outcomes Within Disciplinary Contexts," which differed from the first in four ways: (1) it included majors in political science to determine whether instruments that feature disciplinary content can better measure the ability to think critically than interdisciplinary methods such as the Collegiate Learning Assessment; (2) it called for administering the instrument at two times, once during the students' first or second years and again during their senior years to measure change in the development of their cognitive abilities; (3) the initial cohort of participants expanded to include students at twelve colleges and universities; and (4) the study explored the connection between critical thinking and postformal reasoning through the use of the Reasoning about Current Issues (RCI) questionnaire.

1 On behalf of all those involved in the project, the authors gratefully acknowledge the support of the Teagle Foundation (http://www.teaglefoundation.org), which funded both phases of the study, the contributions of the participating institutions, and the Center for Hellenic Studies (http://chs.harvard.edu).
In the fall of 2009 the study administered the assessment to over 900 students. To maintain contact with the initial group of participants and maximize the response rate for the second administration, the project collected information annually in the spring and spring about the types of courses the students had completed and other campus-based educational experiences, which could account for changes in the cognitive outcomes. In the fall of 2011, the project re-assessed those who were sophomores in 2009, and in the fall of 2012, those who were first-year students at the time of the first administration. To compensate for attrition among the initial group of participants and allow for cross-sectional analysis, the project recruited additional senior students in 2011 and 2012. This group of respondents included majors in classics and political science who had not participated in the first administration.

This presentation will address three topics. First, we will discuss the process of identifying the outcomes and formulating the essay questions, which were to provide novel stimuli and require students to make interpretations, derive meaning, draw conclusions, and make persuasive arguments, but did not assume any prior knowledge of the discipline. We will also describe the steps in developing rubrics for evaluating the essays. This will be of particular interest for those who wish to collect baseline information about their students. The second topic will concern issues associated with administering the assessments, such as complications related to institutional review boards, calibrating incentives, the participants' levels of interest, engagement, and "assessment fatigue." Finally, we will present some of the findings from the study, offer a few interpretive observations, and comment on potential directions for further research.

I. Outcomes

In the spring and fall of 2002, Stephen Klein, George Kuh, Marc Chun, Laura Hamilton, and Richard Shavelson administered a battery of tests to 1365 students at fourteen colleges, with the goal of exploring "the feasibility and utility of using open-ended direct measures of student learning." This project grew out of a study of assessment by Shavelson and Leta Huang entitled "Responding Responsibly to the Frenzy to Assess Learning in Higher Education" and eventually evolved into the College Learning Assessment.²


The CLA concerns "broad abilities," which the authors of the study describe as follows:

Broad abilities are complexes of cognitive processes ("thinking") that underlie verbal, quantitative and spatial reasoning, comprehending, problem solving and decision making in a domain, and more generally across domains. These abilities are developed well into adulthood through learning in and transfer from non-school as well as school experiences, repeated exercise of domain-specific knowledge in conjunction with prior learning and previously established general reasoning abilities. As the tasks become increasingly broad—moving from a knowledge domain to a field such as social science, to broad everyday problems—general abilities exercise greater influence over performance than do knowledge structures and domain-specific abilities. Many of the valued outcomes of higher education are associated with the development of these broad abilities.

Shavelson and his team recognized that academic work within a major field of study contributed to the overall learning outcomes but concluded that other means were available for measuring student learning in specific domains, such as tests "in a capstone course or by an integrated examination." Consequently, their interests lay in assessing broad abilities that would allow, most importantly, for comparisons on an institutional basis. The Teagle Assessment Project began as a response to this approach with the goal of devising a way to measure generalizable cognitive outcomes related to a particular discipline, in this case, classics.

To identify the cognitive outcomes specific to the study of classics Rachelle Brooks recruited five faculty members to serve as a Faculty Advisory Board.\(^4\) This group met in the fall of 2005, to discuss the project and formulate a set of outcomes. Their findings clustered around three broad areas of engagement, which they considered relevant to most undergraduate programs of study in classics. The first concerned the acquisition of one or more ancient languages, typically ancient Greek and Latin, which results in a heightened understanding of natural human language and the role of language in the construction of reality. The second related to the development of a historical perspective, which yields a sense of how culture shapes a people’s understanding of the world and how cultural conventions and institutions evolve or persist over time, and the third reflected the multi-disciplinary nature of the field, which requires students to work with different types of

\(^{4}\) Members of the Faculty Advisory Board were Michael Arnush, Skidmore College, Barbara Gold, Hamilton College, Kenneth Morrell, Rhodes College, David Porter, Skidmore College, and Peter Struck, University of Pennsylvania.
literary, documentary, and material evidence, recognize the limitations of each type of evidence, and to critically evaluate assumptions, methodologies, and conclusions based on them. From a cognitive perspective, these outcomes were examples of formal and postformal reasoning. The former represents the ability to assess evidence and draw logical conclusions in the process of applying knowledge and skills to domain-specific problems and obtaining verifiable results. The latter is the ability to recognize and engage with questions for which the parameters may be unspecified and fluid, the data may be incomplete, conflicting, and compromised, and the solutions may be variable, non-verifiable, and subject to multiple interpretations.

II. DEVELOPING AND TESTING THE INSTRUMENT

With these outcomes in mind the team developed an instrument that consisted of four components:

1. Two analytic problems that presented students with novel stimuli and called for them to assess a narrative description or set of artifacts (presented in photographs), develop an interpretation, organize a series of arguments in support of their position, and identify what additional information they would find helpful. The consulting classicists designed the questions with two primary criteria in mind. First, they would focus primarily on the second and third outcomes. Each question referred to a specific historical context and pertained to a different cultural setting (one concerned Greece in the fifth century BCE, and the other, Roman influence on the coast of North Africa in the first century CE). Each question also drew on different types of evidence (one problem related to literature, and the other to archaeological artifacts). Second, neither question would require specific prior knowledge, but each would offer students, who had acquired some experience in the field, opportunities to apply their learning. At the same time, students who had no experience with the subject could respond cogently to the prompts by drawing on their ability to think critically. In other words, the goal of the questions was not to test the students’ command of classics-related material.

---

5 These outcomes were consistent with those identified in a study conducted by the Center for Hellenic Studies on "Classics and Undergraduate Liberal Education," the results of which appeared in “The Classics Major and Liberal Education,” *Liberal Education* 95 (2009): 14-21.
2. Portions of the Sternberg Triarchic Abilities Test, which distinguishes among academic-analytical, synthetic-creative, and practical-contextual intelligence. This section provided a benchmark for assessing performance on the analytic questions.

3. A series of questions to collect demographic information and data related to the respondents’ educational experiences in high school and college, such as SAT scores, AP courses, GPAs from high school and college, majors or intended majors, participation in “high impact practices,” which positively contribute to learning outcomes, and a series of five questions designed to determine how much effort the respondents put into completing the assessment.  

4. The Reasoning about Current Issues (RCI) assessment, which asks students to evaluate opinions about contemporary issues based on how similar the opinions are to their own. The RCI measured the respondents’ level of postformal reasoning.  

The project recruited two other faculty members to develop rubrics for evaluating the responses to the analytic essays. This process involved three phases. First, the faculty members independently drafted model answers based on how they thought a good student would respond to the questions. They then met with a member of the Faculty Advisory Board to discuss their answers, and from these discussions emerged the first draft of the rubric, which consisted of (a) a set of baseline expectations, (b) a list of specific points the students might mention in their responses, and (c) some suggestions about what the respondents might identify as additional helpful information. To gain a more accurate sense of how undergraduates might perform, the faculty members asked five students to respond to the questions in a simulated testing environment. The five students included a sophomore, who had recently declared a major in classics but was already taking advanced courses in ancient Greek, Latin, and Spanish, two juniors who were both preparing to spend the spring semester at the Intercollegiate Center for Classical Studies in Rome, and two seniors, both of whom had had experience abroad, one at the ICCS and the other on the excavations of the Porta Stabia.

---


in Pompeii. The two faculty members independently evaluated the essays using the rubric. They scored the essays on a scale of 0 to 4, assigning 2 points to essays that met the baseline criteria and then subtracting deductions in half-point increments for elements from the baseline expectations that were missing and adding to the scores for additional points the essays addressed. The faculty members then met to compare the results of their assessment, reconcile differences, and revise the rubric. At this point the project had tested the instrument with a group of forty-six respondents from a mythology course at Hamilton College. The two faculty members independently graded all ninety-two essays, following the procedure they had developed during the second phase, and met a third time to compare the results and revise the rubric.

III. The Study

Having developed and tested the instrument, the project began a three-year study to measure the development of learning outcomes, specifically critical thinking and postformal reasoning, in two academic disciplines, classics and political science. The study called for respondents to participate in the assessment at two times during their undergraduate experience, initially as first-year students or sophomores and again during their junior or senior years. In the interim, the project would collect information about their plans of study. At both times, respondents would answer one question pertaining to classics and one question related to political science. This would hopefully yield data to address two questions: (1) does an approach that uses content from a specific discipline offer a more accurate means of measuring the development of critical thinking than those that claim to be "nondisciplinary" and applicable to the undergraduate population as a whole, such as the CLA, and (2) to what extent do students transfer knowledge and skills developed within the major to other academic contexts?

---

8 Before going into the field in the fall of 2009, the Faculty Advisory Board slightly modified the questions, removing references to historical periods, individuals, and specific geographical locations with the goal of minimizing the potential impression among respondents that the questions required or sought discipline-specific information. They also revised the rubrics accordingly.

9 The project selected political science as the counterpart of classics because, like classics, political science also ranges over multiple disciplines such as economics, history, and philosophy. For certain concentrations with the major, for example, international politics, it also requires students to develop some degree of proficiency in a second language.
In 2008, during the first year of the study, the project formed an advisory group of faculty members from political science to go through the process of developing analytical questions for undergraduate majors in the field. Meanwhile, the project developed the infrastructure of administering the instrument, which included recruiting institutions and obtaining approvals from their review boards. Twelve colleges and universities participated in the study during the first administration of the assessment in the fall semester of 2009. There were three main challenges for the study. The first was recruiting a sufficiently large pool of first-year and sophomore respondents, i.e., those who had yet to declare a major, to ensure that an adequate number of majors from both disciplines would participate in the assessment during their junior or senior years. Ultimately the study relied on the judgment of faculty members at the participating institutions to determine which courses would likely include students who would go on to major in the field, for example, gateway or required courses for the major.

The second challenge was in maintaining contact with the respondents and tracking their academic work between the two administrations of the assessment while complying with provisions outlined by institutional review boards for ensuring the anonymity of the participants. The third, which is common to all studies of this nature, was finding appropriate and effective ways of encouraging students to do their best, when there were no academic or professional incentives. (The brevity of this paper prevents us from discussing these points at length, but future disseminations of the project’s findings will include further details.) During the first administration of the instrument in the fall of 2009, 744 students completed the assessment. Of those 453 participated two years later, in the spring semester of 2012 or 2013, and from that group the study collected responses from eighteen classics and twenty-seven political science majors.

IV. Findings

These were the College of the Holy Cross, Davidson College, Grand Valley State University, Hamilton College, Monmouth College, Northwestern University, Oberlin College, Rhodes College, Skidmore College, Swarthmore College, the University of Pennsylvania, and Wesleyan University. The classics programs at all of the institutions were involved in the study. Political science programs at Grand Valley State University, Hamilton College, Monmouth College, Northwestern University, Rhodes College, Skidmore College, and Swarthmore College also participated. Swarthmore College did not participate in the second assessment. The project added the University of Maryland, Baltimore County, during the second administration in the spring of 2012 and 2013 to obtain additional cross-sectional data.
The process of analyzing this extensive collection of data is underway and still preliminary. For now we will look at some initial results based on the responses of thirteen classics majors. They participated in the first administration of the instrument as first-year students in introductory classics courses at seven different institutions. At the time none of them expressed an intention of majoring in the field. All had high school GPAs of between 3.6 and 5.2, and above average SAT or ACT scores. Seven were female and eight were male, and one was a student of color. All identified English as their first language except for one who did not report. The chart below summarizes what we found. The participants are ranked according to the change in their performance on the classics essay. Columns L through T indicate whether students participated in high impact practices.¹¹

¹¹ The only high-impact practice the study did not track was the capstone course or project.
Six of the thirteen students performed better on the classics essay in their senior year than they did as first-year students. All six also showed improvement on the political science essay. Given this improvement in critical thinking in both disciplinary domains, we looked for other commonalities among these students. Five of the six did not have a second major, so their educational program could have involved additional classics courses beyond the requirements for the major or a broader array of courses than those with double majors are likely to experience.

Four of the six students (rows two through five) improved by at least two points (out of ten) on the classics essay. All four reported participating in the same five high impact practices: a first-year seminar, a writing-intensive course, one or more courses that featured collaborative learning, independent study or research, and a course that focused on issues of diversity. Three of
the four reported being able to read and write in another language in their first year of college, and all four, by the time they were seniors. Three of the four could also speak another language by the time they were seniors.

By contrast, the two students, whose performance on the classics essay declined the most, also performed worse or only marginally better on the political science essay. Both participated in a writing-intensive course and courses that included collaborative learning and addressed issues of diversity, but neither reported enrolling in a first-year seminar. Neither reported having the ability to read or write in another language when they entered college, and neither developed the ability to read, write or speak in a second language during their college experience.

Some patterns are emerging from the data for students who show improvement in critical thinking and postformal reasoning during college. For example, participating in a first-year seminar, especially those that emphasize critical thinking about challenging topics, may actuate more advanced learning approaches at an earlier stage in a student’s collegiate experience, which may yield cross-disciplinary benefits later on. Also, the acquisition of a second language appears to play a significant role in the development of critical thinking. This latter finding is consistent with a study conducted at Kalamazoo College during the 2005-2006 academic year using the CLA. For that study, 186 first-year students and a stratified random sample of sixty-seven seniors took the CLA. The study then looked for factors that might account for the difference between the mean performance of the first-year students (eightieth percentile) and seniors (ninety-ninth). The authors grouped the seniors according to the college's five academic divisions, fine arts, foreign languages, humanities, natural sciences and mathematics, and social sciences, and concluded, "The academic division in which students majored seems to have had an effect on CLA performance. Adjusted CLA scores differed significantly among divisions, even though actual CLA scores did not, with students majoring in Foreign Languages having the highest AdjCLA and students in Natural Sciences having the lowest AdjCLA."\textsuperscript{12}

IV. CONCLUSIONS

In closing here are three observations. Although the analysis of the data from this project is only now getting underway, the study has already raised three issues that warrant further consideration and study. First, the number of respondents in the target group, that is, students who participated in the assessment during their first or second years, declared a major in classics or political science, and participated again in their junior or senior year was surprisingly small—only about 6.5% of those who completed the assessment at time one. Second, questions remain about how much effort the participants put into completing the assessments and how accurately they recalled their collegiate experiences. The responses from Students 377 and 419 provide two examples. The rather steep decline of Student 377’s performance on the classics essay, while showing a slight gain on the political science question and RCI, raises doubts about the amount of time or effort the student may have put into answering the question. Among the thirteen students in this analysis, Student 419 reported expending the lowest level of effort on the assessment as a first-year student. At that time this respondent also reported having participated in four high impact practices: a first-year seminar, an independent research project, a course on service learning, and an internship. As students gain more exposure to high impact practices over the course of their undergraduate years, some variation in responses to questions about those practices between the students’ first and senior years is expected. There is generally an increase the number of reported experiences. That students occasionally forget having participated in some is also likely. But a difference between participating in four at time one and zero at time two suggests a significant reporting error.

With regard to the impact of acquiring a second language on cognitive development or improvements in critical thinking and postformal reasoning, further study needs to determine whether the effects are comparable for (1) students who acquired a degree of proficiency in a second language during high school but did not go on to study the same language or acquire another language in college, (2) students who began acquiring a language in high school and continued studying the same language in college, and (3) students who did not acquire a degree of proficiency in a second language during high school but did in college. In other words, is it the ability itself, the development of the ability, or a combination of the two that accounts for the effects? These questions all point to the need for further study and refinements in the approach, which might include (1) conducting the study at one or more institutions with better coordination between the participating
departments over a longer period of time, so that the study could yield more statistically significant numbers and (2) embedding the assessment in the curriculum, for example, as graded components of gateway and capstone courses as a way to ensure a higher level of engagement.

In summary, we conclude that this approach to assessing generalizable cognitive outcomes among undergraduates has promise and can help inform faculty members as they seek better ways of measuring and understanding how their curricula and teaching strategies affect the intellectual development of their students. Given the relative lack of assessment tools available to even those institutions that are able and willing to pay for them, additional efforts, such as this one, can contribute to conversations about what students gain in general from a college education and in particular from a major field of study.