SUPPLEMENTAL MATERIALS – APPENDICES
SPECIAL VISIT REPORT
NOVEMBER 2014

The following materials are appended. The first three are in response to requests by the Visiting Team. The others are updates and/or supplemental materials.

1. Demographic Data on Transfer Students – Entering Class of 2014 2
2. Graduation Rates for Transfer Students Entering with Junior Status (2007-2011) 4
3. Schedule of Future Academic Program Reviews:
   - University Committee on Academic Review (UCAR)
   - Undergraduate Program Review (UPR) 5
4. Important Announcements Regarding USC’s Leadership 7
5. USC Organizational Chart 10
6. Quantifying Learning in Critical Thinking – supplementary article
   Writing and Critical Thinking Study 11
### Fall 2014 Transfer Students by Gender

<table>
<thead>
<tr>
<th>SEX</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>729</td>
<td>50.8%</td>
</tr>
<tr>
<td>Male</td>
<td>706</td>
<td>49.2%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>1435</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

### Fall 2014 Transfer Students by Ethnicity

<table>
<thead>
<tr>
<th>ETHNIC</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native American</td>
<td>3</td>
<td>0.2%</td>
</tr>
<tr>
<td>Asian/Asian American</td>
<td>291</td>
<td>20.3%</td>
</tr>
<tr>
<td>African American/Black</td>
<td>49</td>
<td>3.4%</td>
</tr>
<tr>
<td>Latino/Hispanic</td>
<td>228</td>
<td>15.9%</td>
</tr>
<tr>
<td>Non-Resident Alien</td>
<td>224</td>
<td>15.6%</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>3</td>
<td>0.2%</td>
</tr>
<tr>
<td>Two or More Ethnicities</td>
<td>57</td>
<td>4.0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>165</td>
<td>11.5%</td>
</tr>
<tr>
<td>White</td>
<td>415</td>
<td>28.9%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>1435</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

### Fall 2014 Transfer Students and First Generation

<table>
<thead>
<tr>
<th>First Generation Student?</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>921</td>
<td>64.2%</td>
</tr>
<tr>
<td>Yes</td>
<td>514</td>
<td>35.8%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>1435</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>
### Fall 2014 Transfer Students by Type of Sending Institution

<table>
<thead>
<tr>
<th>Sending Institution Type</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community College</td>
<td>829</td>
<td>57.8%</td>
</tr>
<tr>
<td>Four-Year University</td>
<td>604</td>
<td>42.1%</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>0.1%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td>1435</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

### Fall 2014 Transfer Students by School

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annenberg School for Communication</td>
<td>100</td>
<td>7.0%</td>
</tr>
<tr>
<td>Dornsife College of Letters, Arts &amp; Sciences</td>
<td>629</td>
<td>43.8%</td>
</tr>
<tr>
<td>Keck School of Medicine</td>
<td>18</td>
<td>1.3%</td>
</tr>
<tr>
<td>Leonard Davis School of Gerontology</td>
<td>2</td>
<td>0.1%</td>
</tr>
<tr>
<td>Leventhal School of Accounting</td>
<td>6</td>
<td>0.4%</td>
</tr>
<tr>
<td>Marshall School of Business</td>
<td>314</td>
<td>21.9%</td>
</tr>
<tr>
<td>Ostrow School of Dentistry</td>
<td>38</td>
<td>2.6%</td>
</tr>
<tr>
<td>Roski School of Fine Arts</td>
<td>14</td>
<td>1.0%</td>
</tr>
<tr>
<td>School of Architecture</td>
<td>25</td>
<td>1.7%</td>
</tr>
<tr>
<td>School of Cinematic Arts</td>
<td>57</td>
<td>4.0%</td>
</tr>
<tr>
<td>School of Theatre</td>
<td>33</td>
<td>2.3%</td>
</tr>
<tr>
<td>Sol Price School of Public Policy</td>
<td>29</td>
<td>2.0%</td>
</tr>
<tr>
<td>Thornton School of Music</td>
<td>28</td>
<td>2.0%</td>
</tr>
<tr>
<td>Viterbi School of Engineering</td>
<td>142</td>
<td>9.9%</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td>1435</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

### Fall 2014 Transfer Students by Permanent Residency

<table>
<thead>
<tr>
<th>Location of Permanent Residence</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within California</td>
<td>1032</td>
<td>71.9%</td>
</tr>
<tr>
<td>Within United States, Outside of California</td>
<td>193</td>
<td>13.4%</td>
</tr>
<tr>
<td>Outside of the United States</td>
<td>210</td>
<td>14.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1435</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
## Graduation Rates for Transfers Students with Junior Status by Cohort

<table>
<thead>
<tr>
<th>Entering Term</th>
<th>Graduated in Three Years</th>
<th>Graduated in Four Years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Didn’t Graduate</td>
<td>Graduated</td>
</tr>
<tr>
<td>20073 %</td>
<td>16.23%</td>
<td>83.77%</td>
</tr>
<tr>
<td>N</td>
<td>49</td>
<td>253</td>
</tr>
<tr>
<td>20083 %</td>
<td>14.97%</td>
<td>85.03%</td>
</tr>
<tr>
<td>N</td>
<td>72</td>
<td>409</td>
</tr>
<tr>
<td>20093 %</td>
<td>15.38%</td>
<td>84.62%</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>363</td>
</tr>
<tr>
<td>20103 %</td>
<td>13.50%</td>
<td>86.50%</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>423</td>
</tr>
<tr>
<td>20113 %</td>
<td>12.35%</td>
<td>87.65%</td>
</tr>
<tr>
<td>N</td>
<td>60</td>
<td>426</td>
</tr>
</tbody>
</table>
Academic Program Review Schedule
University Committee on Academic Review (UCAR)

**Fall 2014**

Department of History

Thornton School of Music

Department of East Asian Languages and Cultures

**Spring 2015**

Davis School of Gerontology

Department of Management and Organization

Department of Industrial and Systems Engineering

**Fall 2015**

Department of Sociology

Department of Physics and Astronomy

School of Dramatic Arts

**Spring 2016**

Comparative Studies in Literature and Culture Doctoral Program

Section in Human Evolutionary Biology in the Department of Biological Sciences

Master of Arts in Cinematic Arts
Academic Program Review Schedule
Undergraduate Program Review (UPR)

**Fall 2015**

Department of English
Political Science

**Spring 2016**

American Studies
Neuroscience

**Fall 2016**

Marine Environmental Biology
Art History

**Spring 2017**

Cognitive Science
Gender studies
Linguistics

**Fall 2017**

Molecular and computational Biology
Slavic Languages

**Spring 2018**

East Asian Languages and Cultures
East Asian area studies

**Fall 2018**

Middle East Studies
Overseas Studies Office
MEMORANDUM

To: Academic Deans, Faculty, Staff, and Students
From: C. L. Max Nikias, President
Date: October 6, 2014
Subject: National Search: USC Senior Vice President for Academic Affairs and Provost

I am pleased to formally announce a national search for USC’s next senior vice president for academic affairs and provost. As the second-ranking officer under the president, and the chief academic officer of the university, this person plays a central role in USC’s operations. All of the university’s deans and vice provosts report directly to this person, who also oversees the divisions of Student Affairs, Libraries, Information Technology Services, Student Religious Life, and Enrollment Services. I believe this appointment is particularly critical at this juncture in USC’s history, as we are in the midst of taking the university into the ranks of undisputed academically elite universities.

I have already formed a search committee that consists of a wide range of prominent faculty from an array of schools, and the full roster accompanies this memorandum. The committee members and I will actively consult the executive board of the Academic Senate, University and Distinguished Professors, Student Government (both undergraduate and graduate), the deans, and the academic department chairs of the Keck School of Medicine. We will use the feedback we receive from these consultations to determine the ideal qualifications of our next provost and to guide our evaluations throughout the process.

Isaacson, Miller, a highly-regarded search firm, is assisting the university in its search, and the firm’s president, John Isaacson, and vice president, David Bellshaw, will oversee the firm’s efforts. The members of the search committee and I are seeking input, and I invite you to submit nominations or applications for this position. You can direct these materials to uscprovost@imsearch.com. The search committee will make every effort to maintain absolute confidentiality throughout the process.

I want to take this opportunity to warmly thank the members of the search committee for their service. As I told each one, I consider this the most significant appointment of my cabinet as president. This person will play a pivotal role in USC’s continued ascent, and we all look forward to appointing an exceptionally accomplished individual to USC’s leadership team.
Provost Search Faculty Advisory Committee

Chair
C. L. Max Nikias, President

Members
Hortensia Amaro, Dean’s Professor in Social Work and Preventative Medicine, School of Social Work
Daniela Bleichmar, Associate Professor in Art History and History, Dornsife College of Letters, Arts and Sciences; Smithsonian Magazine’s one of “37 under 36” America’s Young Innovators
Leo Braudy, University Professor and Leo S. Bing Chair in English and American Literature, Dornsife College of Letters, Arts and Sciences
Ainsley Carry, Vice Provost for Student Affairs; Professor of Clinical Education, Rossier School of Education
Stephen Gruber, Professor of Medicine, H. Leslie Hoffman and Elaine S. Hoffman Chair in Cancer Research, Director of the USC Norris Comprehensive Cancer Center, Keck School of Medicine
Ange-Marie Hancock, Associate Professor of Political Science and Gender Studies, Dornsife College of Letters, Arts and Sciences; member of the Executive Board of the Academic Senate
Velina Hasu Houston, Professor of Theatre in Dramatic Writing, Director of Dramatic Writing Program, School of Dramatic Arts; Associate Dean of Faculty Recognition and Development
Deborah MacInnis, Charles L. and Ramona I. Hilliard Professor of Business Administration, Professor of Marketing, Marshall School of Business
Ellis Meng, Professor of Biomedical and Electrical Engineering; chair of the Women in Science and Engineering program; National Science Foundation and Wallace H. Coulter Foundation Career Awards; TR35 Young Innovator Under 35
Manuel Pastor, Professor of Sociology and American Studies and Ethnicity, Director, Center for the Study of Immigrant Integration, Dornsife College of Letters, Arts and Sciences
Jacob Soll, Professor of History and Accounting, Dornsife College of Letters, Arts and Sciences and Leventhal School of Accounting; MacArthur Fellow
David St. John, Professor of English and Comparative Literature, chair of the English Department, Dornsife College of Letters, Arts and Sciences; renowned poet
Rohit Varma, Grace and Emery Beardsley Chair in Ophthalmology, Professor of Ophthalmology and Preventative Medicine, chair of the Department of Ophthalmology, Director of the USC Eye Institute

Search Administrator
Sarah Bullock, Office of the President
sbullock@president.usc.edu, (213) 740-2111

Search Firm
Isaacson, Miller
John Isaacson, President
David Bellshaw, Vice President and Director
MEMORANDUM

To: Academic Deans, Faculty, Staff, and Students

From: C. L. Max Nikias, President

Date: October 17, 2014

Subject: Appointment of Interim Provost and Senior Vice President for Academic Affairs

I am pleased to announce that beginning December 1, 2014, Professor Michael Quick will assume the role of interim provost and senior vice president for academic affairs. Professor Quick has agreed to serve in this role until our national search for a new provost is completed. In this capacity, he will oversee all the duties and responsibilities of the Office of the Provost, and all of the provost’s direct reports will remain the same. I would like to thank him for assisting the university during this transition phase. We should all be grateful for his outstanding service to USC and our community.

Our current provost, Elizabeth Garrett, will take her sabbatical from December 1 to June 30, 2015, as she prepares to transition to her new role as president of Cornell University, beginning July 1, 2015. In the spring, we will come together to honor her stellar contributions to USC, and wish her continued success at Cornell. Although we will miss her, I am certain she will be an extraordinary president, and that she will build an enduring and consequential legacy in American higher education.
QUANTIFYING LEARNING IN CRITICAL THINKING

Richard Fliegel and John Holland

ABSTRACT
This article describes a three-year study assessing change in critical thinking demonstrated in essays written for regular class assignments. A rubric was designed and scorers trained to assess critical thinking holistically without knowledge of the writing prompt or author’s status. The longitudinal improvement in scores earned by freshmen over three years was significantly related \( (p < .01) \) to the difference in scores awarded to freshmen and juniors at a single moment in time. This method can be used to provide an “academic snapshot” of a program, using authentic course work to quantify learning in critical thinking over the undergraduate experience.

In fall 2008 we requested and received financial support from the Spencer and Teagle foundations for a project that sought to document and ultimately quantify the change in critical thinking we observed in the writing of our undergraduates at the University of Southern California (USC). The principal investigators on our project served as the directors of the General Education and Writing programs in the College of Letters, Arts and Sciences (now USC Dornsife College). For each of our programs, growth in critical thinking was an articulated learning objective. We were motivated to demonstrate that our students had in fact achieved that outcome, as anecdotal evidence suggested. However, we did not believe that we could effectively measure the result of our classes as distinct from the totality of the curriculum, internships, service opportunities, and other
learning experiences in which our students engaged as undergraduates. Instead, we set out to quantify the change in critical thinking they could demonstrate as a result of their time at USC, from freshmen to juniors and seniors.

We focused on student essays as the work in which change in critical thinking would most likely be measurable. In so doing, we took advantage of two local factors. Our writing requirement had been restructured in 1997 so that students take their first course as freshmen and their second course as juniors or seniors; together the two writing courses presented occasions for before-and-after comparisons of relatively similar material. We have also had long experience with holistic scoring of essays, since our Writing Program was established as an independent academic unit in the 1970s.

To accomplish what we set out to do, we would have to succeed in the following:

1. Create a rubric that focused specifically on critical thinking rather than “good writing” more generally. That rubric would have to capture the implicit sense of critical thinking readers bring to a piece of writing when they decide it is exemplary in this respect.
2. Train a cohort of faculty and graduate students in the use of our rubric, according to the best practices of holistic scoring.
3. Collect essays written by freshmen, juniors, and seniors enrolled at an institution of higher education at a single moment in time; score the critical thinking evinced in those essays; and compare the scores awarded (without knowledge about their authors) to essays from each group.
4. Collect essays written by the same freshmen two years later; score them in another reading that included both their first-year essays and their third-year essays, again without indication as to the status of the authors.
5. Compare the scores of essays written by freshmen with the scores of essays written by juniors and seniors in both readings.

If the essays written by juniors and seniors scored significantly higher on a rubric designed to measure critical-thinking skills, we could conclude that our juniors and seniors demonstrated a higher degree of ability in critical thinking than our entering freshmen. If the essays written by those same freshmen at two different points in their academic careers showed a comparable rise in critical-thinking scores, we could conclude that a “snapshot” of this kind, assessing the critical-thinking abilities of freshmen, juniors, and seniors at one point in time, could be used to assess the improvement in critical thinking demonstrable by students longitudinally over their years at a college or university.

This project was undertaken with a double purpose. We hoped to respond to the national discourse on the assessment of college learning, which suggested...
that students are not learning much as a result of their higher education and that better assessment of learning outcomes is needed to improve student learning. The methods of assessment available to remedy this situation did not seem to us to reflect the values of academic culture or the learning objectives we hold in highest esteem. At the same time, we believed that the process of assessing our effectiveness in teaching students to think critically would help us improve our programs by making the participants in our study more aware of what is being done by their colleagues and the writing produced by their students. This report chronicles our progress toward both of these objectives.

The National Context

The national discourse today concerning critical thinking at the college level centers on failure. In their 2011 book, Academically Adrift: Limited Learning on College Campuses, sociologists Richard Arum of New York University and Josipa Roksa of the University of Virginia summarize their study of 2,322 students at twenty-four four-year colleges and universities by concluding that higher education is failing half of our students in this critical area: “At least 45 percent of students in our sample did not demonstrate any statistically significant improvement in Collegiate Learning Assessment [CLA] performance during the first two years of college. While these students may have developed subject-specific skills that were not tested for by the CLA, in terms of general analytical competencies assessed, large numbers of U.S. college students can be accurately described as academically adrift” (p. 121). In the February 14, 2011, issue of the Chronicle of Higher Education, Alexander Astin of the University of California, Los Angeles, took Arum and Roksa to task for their use of data in this provocative claim, arguing that the criteria for establishing that something is happening beyond random chance cannot be turned on their heads to establish that something is not happening. But their study has captured public attention. Writing as a critic-at-large, Louis Menand reviewed the book in the June 6, 2011, New Yorker, as an opportunity to discuss higher education and what students seem to be expecting of college. The Los Angeles Times published an editorial by Arum and Roksa the following week, explaining why so few students are learning to think critically.

As sociologists, Arum and Roksa are concerned with inequalities in the system for students from more and less affluent families, attending more and less privileged schools. Their report acknowledges that students at some schools show more change in critical thinking than others and that some students in all schools do better than others in this respect. Arum and Roksa focus on the factors that contribute to the acquisition of critical thinking both before
students attend college and while undergraduates are enrolled. On the basis of their data, they argue that the students who learn to think critically are assigned more reading and writing and spend more time studying—conclusions with which most college educators would agree. They also find that extracurricular activities, jobs, and peer-group studying do not improve performance in critical thinking—a series of more contentious assertions.

Arum and Roksa base their claim on a longitudinal comparison of students from one year to the next—that is, the scores of freshmen on the CLA were compared with the scores of the same students as sophomores. This does not reflect the developmental sea change one observes among students sometime between the sophomore and junior years, when they integrate the learning of the first two years, enroll in upper-division courses, and start to plan for life after college. We find much to endorse in the Arum and Roksa book and in their L.A. Times editorial, but there are points of contention, most serious of which is the sense of critical thinking they adopted with their method of assessment.

The Arum and Roksa study uses as its measure of critical-thinking skills the Collegiate Learning Assessment, which has become a widely adopted assessment tool. The CLA presents students with documents describing a situation and asks them to construct an argument; or it presents them with an argument advanced by a party in a situation and asks for their analysis of that argument. The students in the Arum and Roksa study were asked to write a memo about the wisdom of purchasing a private plane for the DynaTech sales force, after a similar plane had crashed. Students were given e-mails, newspaper articles, charts, a federal accident report, pictures of the plane, and an amateur pilot’s comparison with other models. Student responses were evaluated using the criteria published by Richard Shavelson in his 2010 book, Measuring College Learning Responsibly: Accountability in a New Era—How clear and concise is the argument? How effective is the structure? How well does the student defend the argument?—followed by a three- or four-point rubric for each of those questions. These “Criteria for Scoring Responses to Critical Thinking Prompts” incorporate an implicit definition of critical thinking that we believe needs to be questioned.

The CLA was developed by Richard Shavelson together with Roger Benjamin and Steven Klein, based on work begun by the latter two when both were at RAND in Santa Monica. We had opportunities to meet with each of them at that early stage of development. Professor Klein shared his doubts about holistic scoring in general; Professor Benjamin was more interested in what we were trying to do but asked, “How do you get the kids to take it seriously?” That is, how do you elicit from students their best work, or representative work, when they know that the assessment tool is not part of their curriculum, their instructor’s
assignments, or their grade—when it might have been imposed on all of them for the sake of institutional accountability?

Our answer to that question has been to design a process for assessing change in critical thinking that is accepted by both students and their teachers as a genuine part of the class material, consonant with academic values, and focused on the skills classroom instructors intend to teach their students. “Embedded” assessment measures suggest by the metaphor that they are extrinsic to the class material, slipped into the curriculum like a seed in soil, an invasive medical procedure, or a journalist in a military unit. Instead, we would try to measure change in critical thinking demonstrated through the materials developed by college instructors as part of their regular teaching materials. Before doing so, we ran a more controlled study.

Preliminary Study

Before undertaking our project for the Spencer and Teagle foundations, we conducted a study of critical-thinking skills demonstrated by students at different points in their careers at USC. We thought to take advantage of the structure of our writing requirement, which includes one class at the freshman level (WRIT 140, Writing and Critical Reasoning) and a second class at the junior or senior level (WRIT 340, Advanced Writing). This arrangement allowed us to compare essays of entering students with those of their upper-class colleagues. We solicited volunteers among the writing faculty who were willing to assign the same prompt to their classes for the same time period of in-class writing. The essays were coded as to their source and then stripped of any class identifiers and scored by a cohort of writing instructors who had been trained in the use of a rubric designed expressly for this purpose by Peter and Noreen Facione.

Peter Facione, the former provost at Loyola in Chicago, was involved in the Critical Thinking Movement in philosophy in the 1990s. He developed a fill-in-the-blanks measure of critical-thinking skills and then, with Noreen Facione, a holistic scoring rubric, designed to focus on critical thinking rather than on good writing more generally. Although we had reservations about the implied conception of critical thinking, which is similar to the model used by the CLA, we used the Facione rubric, which articulates complex ideas in a clear and concentrated form, with a different focus from the rubrics used to grade good writing. The Facione rubric has only four points, which makes training scorers simpler and its wider adoption more likely.

Our participants’ essays were collected on paper and scored without knowledge of their source, and the results were compared. The distinction between upper-half and lower-half scores on the rubric was sharp, although
finer distinctions within quartiles of the scale were not. When we tracked the sources of the essays, we found the distribution in Figure 1.

In brief, 45.3 percent of the essays written by freshmen scored in the upper half of the Facione rubric, while 58.2 percent of all essays written by seniors scored in the upper half of the scale. Among juniors and seniors who began as freshmen at USC, 65.5 percent scored in the upper half of the Facione scale, while 73.8 percent of those who majored in the USC Dornsife College of Letters, Arts and Sciences scored in the upper half of the scale.

The Spencer/Teagle Project

Supported by the grant provided for systematic improvement in undergraduate education, we set out to improve on what we had previously done. We hoped to develop a technological process for collecting and presenting student essays to scorers that would enable us to gather essays from more students. At the same time, we wanted to create a new rubric that captured an idea of critical thinking closer to our own conception and to the intuitive sense of the term used by generally well-educated readers.

We began by gathering faculty on our campus from a variety of disciplines in the humanities and social sciences for a discussion about critical thinking. Before the discussion began, however, we distributed blank index cards and asked the participants to answer one question: “When you decide that a piece of writing demonstrates good critical thinking, to what in the writing are...
you responding?” The cards that came back emphasized originality—an idea or insight “that makes me want to read further,” which was not a feature of particular emphasis in the Facione rubric and is not in the CLA scoring criteria.

We repeated our index card exercise at several other academic gatherings, including a meeting of other recipients of Spencer/Teagle grants. On the basis of those responses and discussions among our faculty, we designed two new rubrics focused on critical thinking. A six-point rubric was created by the faculty of the Writing Program; and we created a four-point rubric ourselves, taking advantage of what we had learned from the six-point rubric discussions (Appendixes A and B).

Each of these rubrics incorporates a functional definition of critical thinking that emphasizes originality of thought more than has been the case in other researchers’ work. We ultimately favored our four-point rubric, for the ease with which it can be understood and applied by scorers who may have little prior experience with holistic scoring. The crucial determinant of an upper-half score on that rubric requires evidence of a writer’s ability to analyze a question or problem and to synthesize an individual response to his or her analysis. These skills may be evident at the level of the paragraph or the essay as a whole, but the ability to take ideas apart and put them back together in an innovative way seems crucial to a reader’s judgment that a piece of writing demonstrates skill at critical thinking.

The functional definition of critical thinking employed by Benjamin, Klein, and Shavelson in the CLA includes skill at what is frequently called quantitative reasoning. It makes sense to do so. Faculty across the “two cultures” articulated in 1959 by C. P. Snow at Cambridge have identified critical thinking as the goal of their pedagogy, although “literary intellectuals” and scientists seem to have different sets of skills in mind (see 1959/1998). Our rubric does not require an essay to display its author’s facility with quantitative reasoning as a necessary constituent of critical thinking. To our minds, a prerequisite facility with numbers may prevent a writer from demonstrating what he or she can do with other modes of thinking. While we value quantitative reasoning among our undergraduates, we would recommend assessing it independently from the critical-thinking skills a student can demonstrate in writing.

Methodology

Working with Instructional Technology Services at USC, we collected hundreds of student essays (using Blackboard) and retrieved them in a usable form (which required more extensive programming than we had anticipated). Software was utilized that brought up individual essays randomly, without identifying
information, so that scorers could rate them easily and we could track their scores. We used this system to read a new set of essays written by freshmen in WRT 140 classes and juniors and seniors in WRT 340 classes, in response to different prompts assigned by their classroom instructors. In the 2009 reading, raters were first trained to use the six-point rubric, without access to the writing prompts. Then we repeated the process with the four-point rubric, using the same set of essays.

Results: 2009 Reading

The results of the rating using the six-point rubric showed a nice contrast between the scores earned by freshmen in WRT 140 and the scores earned by juniors and seniors in WRT 340 (Figure 2). The blue line indicates the scores of the freshmen, and the pink line, of juniors and seniors. When the essays were scored using the four-point rubric, a similar distinction emerged (Figure 3). In both figures, the points indicate the percentages of essays that scored in the upper half of

![Critical Thinking Study: 6-Point Rubric](image1)

**FIGURE 2**

![Critical Thinking Study: 4-Point Rubric](image2)

**FIGURE 3**

Quantifying Learning in Critical Thinking
TABLE I  Summary of Critical Thinking Study Findings, 2009

<table>
<thead>
<tr>
<th></th>
<th>6-Point Rubric</th>
<th>4-point Rubric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count 140 Essays</td>
<td>259</td>
<td>Count 140 Score</td>
</tr>
<tr>
<td>Sum 140 Scores (3 Readings)</td>
<td>2450</td>
<td>Sum 140 Scores</td>
</tr>
<tr>
<td>Mean 140 Score</td>
<td>9.459459459</td>
<td>Mean 140 Score</td>
</tr>
<tr>
<td>Average Score</td>
<td>3.153153153</td>
<td>Average Score</td>
</tr>
<tr>
<td>Count Upper Half</td>
<td>91</td>
<td>Count Upper Half</td>
</tr>
<tr>
<td>Count Lower Half</td>
<td>168</td>
<td>Count Lower Half</td>
</tr>
<tr>
<td>Count 340 Score</td>
<td>261</td>
<td>Count 340 Score</td>
</tr>
<tr>
<td>Sum 340 Scores</td>
<td>3117</td>
<td>Sum 340 Scores</td>
</tr>
<tr>
<td>Mean 340 Score</td>
<td>11.94252874</td>
<td>Mean 340 Score</td>
</tr>
<tr>
<td>Average Score</td>
<td>3.980842912</td>
<td>Average Score</td>
</tr>
<tr>
<td>Count Upper Half</td>
<td>185</td>
<td>Count Upper Half</td>
</tr>
<tr>
<td>Count Lower Half</td>
<td>76</td>
<td>Count Lower Half</td>
</tr>
</tbody>
</table>

the scale, judged by raters using our rubrics. Table 1 shows a summary of scores in numerical form. (Note, for other representations of results, see Appendix C.)

As usual, our result posed as many questions as it answered. But it did seem to support the contention that our students are improving their critical-thinking skills over their time at USC, and we can use the process created for our study to quantify the change.

Our purpose in gathering these data was twofold: to accrue evidence for program accountability and to learn how to improve our instruction to increase student learning. The former goal was important for the university’s
reaccreditation, while the latter was more interesting to our writing faculty. The experience of reading hundreds of essays written for other instructors in response to a variety of prompts provoked productive conversations among our scorers as well as follow-up studies, which will be discussed below.

Results: 2011 Reading

During the 2011–12 academic year, the students who entered as freshmen in fall 2008 had advanced to junior standing, when many of them enrolled in sections of WRIT 340, Advanced Writing. This enabled us to collect samples of their essays, to record information about the authors of those essays, to strip them of identifiers, and to have them scored by readers trained in our four-point rubric.

Our purpose was to determine how closely longitudinal data approximated the data we had elicited from different groups of students on campus at the same time. We compared the difference in scores between the (2008) freshman essays and (2011) junior essays by the same students with the difference in scores received by the essays written by different freshmen, juniors, and seniors in our 2009 reading. If those differences proved to be comparable, a school should be able to assess the improvement in critical thinking among its students over time by measuring the critical abilities of its entering freshmen and those of its juniors and seniors.

We collected essays from all students registered in WRIT 340, Advanced Writing, in spring 2011 and searched for essays written by students whose freshman essays had also been collected for our earlier reading. We found over three hundred essays by 2011 juniors that fit that criterion. These essays were cleaned of any information that identified the class standing of the author and were aggregated with essays that had been written by the same students as freshmen, similarly stripped of identifying information. The latter collection comprised over five hundred essays, including more than one example from several students. The result was a set of 829 essays, each of which had to be scored at least three times using the four-point rubric. Fourth readings were required by the chief reader for interrater reliability, when essays received “splits”—for example, scores of 2, 2, and 4 from the first three readers. The mean scores awarded to all these essays in 2,500 readings, and the scores of the essays in our 2009 reading session, are summarized in Table 2.

All of the essays in our 2011 reading received slightly lower scores than the essays scored in our 2009 reading, but the similarity of results is striking. The students who wrote essays both as freshmen and as juniors showed a degree of improvement in critical-thinking skills closely resembling the difference quantified by our scorers in 2009, comparing freshmen who entered USC in 2008 with students who were juniors or seniors at the time. If anything, the increase in critical thinking demonstrated by the cohort of freshmen who entered
The longitudinal improvement in critical-thinking skills for freshmen who entered in fall 2008 appears slightly stronger than the difference between students at the freshman and junior or senior level in our 2009 reading. One contributor may be the rise in SAT scores among entering freshmen at USC during the period between 2005 and 2008. Another contributor may be the improved training of our scorers, who made better use of the full scale of scores in 2011 than in our 2009 reading. In any case, the difference in mean scores (2.093 to 2.038, a total of 0.055) suggests that a simultaneous evaluation of freshman and junior essays should be comparable, if not identical, to the results of a longitudinal study.

The results of our 2011 reading are represented in Figure 4. The summary of scores reveals that almost 24 percent of the essays written while students were freshmen in WRIT 140 scored in the upper half of the rating scale; more than 76 percent of those essays scored in the lower half of the scale. Two years later, almost 66 percent of the essays written by the same students as juniors scored in the upper half of the rating scale, with just over 34 percent scoring in the lower half of the scale.

A breakdown by point score is provided in Table 3. Essays that received reading scores up to 7 (scored a maximum of 2-2-3 by three readers) were considered to have fallen in the “lower half” of the scale, while those that received scores of 8 or higher (scored at least 3-3-2 by three readers) were deemed to have fallen in the “upper half” of the scale. This binary distinction was intended to approximate a reader’s informal judgment that a piece of writing demonstrates “good critical thinking.”

These figures include all samples of work submitted by students in Advanced Writing classes. Some of those essays were written in genres much

<table>
<thead>
<tr>
<th>TABLE 2 Comparing Freshman Essays to Junior and Senior Essays</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Scores</strong></td>
</tr>
<tr>
<td><strong>Reading</strong></td>
</tr>
<tr>
<td><strong>2009</strong></td>
</tr>
<tr>
<td>Freshman Essays (from WRIT 140 classes)</td>
</tr>
<tr>
<td>Junior and Senior Essays (from WRIT 340 classes)</td>
</tr>
<tr>
<td>Difference</td>
</tr>
</tbody>
</table>
more conducive to critical thinking than others. If the comparison were limited to \textit{writ 340} papers in critical genres for each student, the contrast between freshman- and junior-level work would likely prove even stronger (Table 4). If we look at the individual point scores we see that the preponderance of freshman papers from \textit{writ 140} that did score in the upper half (below the black line in Table 4) fell at the 8 or 9 point level (9.24 and 9.57 percent, respectively, or 18.81 percent of the total freshman essays). In contrast, the papers written at the junior level climbed considerably higher in the upper half of the scale (17.11, 15.21, and 14.07 percent at each point score, or 46.39 percent of the total \textit{writ 340} essays).

### Table 3

<table>
<thead>
<tr>
<th>6-Point Rubric</th>
<th>4-point Rubric</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Count 140 Essays}</td>
<td>\textit{Count 140 Score}</td>
</tr>
<tr>
<td>\textit{Sum 140 Scores (3 Readings)}</td>
<td>\textit{Sum 140 Scores}</td>
</tr>
<tr>
<td>\textit{Mean 140 Score}</td>
<td>\textit{Mean 140 Score}</td>
</tr>
<tr>
<td>\textit{Average Score}</td>
<td>\textit{Average Score}</td>
</tr>
<tr>
<td>\textit{Count Upper Half}</td>
<td>\textit{Count Upper Half}</td>
</tr>
<tr>
<td>\textit{Count Lower Half}</td>
<td>\textit{Count Lower Half}</td>
</tr>
<tr>
<td>\textit{Count 340 Score}</td>
<td>\textit{Count 340 Score}</td>
</tr>
<tr>
<td>\textit{Sum 340 Scores}</td>
<td>\textit{Sum 340 Scores}</td>
</tr>
<tr>
<td>\textit{Mean 340 Score}</td>
<td>\textit{Mean 340 Score}</td>
</tr>
<tr>
<td>\textit{Average Score}</td>
<td>\textit{Average Score}</td>
</tr>
<tr>
<td>\textit{Count Upper Half}</td>
<td>\textit{Count Upper Half}</td>
</tr>
<tr>
<td>\textit{Count Lower Half}</td>
<td>\textit{Count Lower Half}</td>
</tr>
</tbody>
</table>

| Count 140 Essays | 303 |
| Sum 140 Scores (3 Readings) | 1947 |
| Mean 140 Score | 6.425742574 |
| Average Score | 2.141914191 |
| Count Upper Half | 72 |
| Count Lower Half | 231 |
| Count 340 Score | 526 |
| Sum 340 Scores | 4475 |
| Mean 340 Score | 8.507604563 |
| Average Score | 2.835868188 |
| Count Upper Half | 347 |
| Count Lower Half | 179 |
Analyses of our data confirm that the difference between scores awarded to essays written by the same students while enrolled in WRIT 140 (in 2008) and WRIT 340 (in 2011) is significant beyond the .01 level—that is, there is less than one chance in a hundred that these results were obtained by random chance. If our rubric measures what we designed it to measure, these students in aggregate demonstrated a quantifiable improvement in their critical-thinking skills over their time at USC. (See Appendix D.) When the results of our simultaneous study (scoring essays written by freshmen, juniors, and seniors in fall 2008) are compared with the results of a longitudinal study (scoring essays written in 2008 and those written in spring 2011 by the same students), we find an equally significant difference.

On the basis of these results we would argue that our method has demonstrated a quantifiable improvement in critical-thinking skills over three years of our students’ educational experience and that this longitudinal improvement can be approximated by comparing essays written by freshmen with those written by juniors or seniors at a single moment in time. An institution of higher education can meet the demand for accountability in this learning outcome by taking a “holistic snapshot”—assessing the critical thinking demonstrated by students in the essays they write for their classroom instructors and using the difference between the scores awarded to freshmen, juniors, and seniors as an index of the improvement in critical thinking accomplished by students over their years at an institution.

<table>
<thead>
<tr>
<th>4-Point Rubric</th>
<th>140</th>
<th>%</th>
<th>Cum%</th>
<th>340</th>
<th>%</th>
<th>Cum%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12</td>
<td>3.96%</td>
<td>3.96%</td>
<td>6</td>
<td>1.14%</td>
<td>1.14%</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>11.88%</td>
<td>15.84%</td>
<td>27</td>
<td>5.13%</td>
<td>6.27%</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>11.55%</td>
<td>27.39%</td>
<td>34</td>
<td>6.46%</td>
<td>12.74%</td>
</tr>
<tr>
<td>6</td>
<td>84</td>
<td>27.72%</td>
<td>55.12%</td>
<td>57</td>
<td>10.84%</td>
<td>23.57%</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>21.12%</td>
<td>76.24%</td>
<td>55</td>
<td>10.46%</td>
<td>34.03%</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>9.24%</td>
<td>85.48%</td>
<td>52</td>
<td>9.89%</td>
<td>43.92%</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>9.57%</td>
<td>95.05%</td>
<td>90</td>
<td>17.11%</td>
<td>61.03%</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>3.30%</td>
<td>98.35%</td>
<td>80</td>
<td>15.21%</td>
<td>76.24%</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>1.65%</td>
<td>100.00%</td>
<td>74</td>
<td>14.07%</td>
<td>90.30%</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0.00%</td>
<td></td>
<td>51</td>
<td>9.70%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>303</td>
<td></td>
<td>526</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conceptual Outcomes

In addition to the assessment process we set out to create, our project has resulted in a number of conceptual outcomes, unplanned but not unintended, among the faculty who participated. These conceptual outcomes have taken three forms: (1) follow-up studies in response to what we were learning; (2) discussions about critical thinking, what it comprises, and how it should be assessed; and (3) resulting changes to the curriculum.

Faculty of the Writing Program who participated in our study decided to learn more about the students coming into the university, their history with critical thinking, and the attitudes they bring with them. They found that critical thinking is presented in high school as puzzle solving and crisis management and that students are taught “the three p’s”—pastiche, personal narratives, and paragraphs (five). Students bring little skepticism to what they read, except about the advertising they encounter; “blogspeak” and tweeting are two frequent forms of their discourse.

In our sessions, each essay was scored without reference to the writing prompt that elicited it, but many could be deduced from the essays themselves. It became apparent to our scorers that some prompts allowed more latitude than others for writers to demonstrate their abilities with critical thinking. In the discussions that followed, it appeared that some kinds of prompts were more likely to invite students to demonstrate superior critical-thinking skills. Those that required students to confront open-ended questions without predictable answers, engaging areas of uncertainty and ethical dilemmas, enabled them to write essays that articulated more personal, specific, and original thought—elements scorers identified as determinant in their assessment of the critical thinking articulated in an essay.

There are clearly contexts in which questions that call for correct answers are academically useful, especially when testing for particular knowledge, but prompts designed to engage uncertainty seem more likely to elicit what general readers consider “good critical thinking.” If that is true, critical thinking should be taught as a rhetorical act, an effort to construct a personal perspective with sufficient appeal to be shared. To that end, a good deal of rhetorical theory developed over the last forty years is applicable. Process theory, reading theory, and social constructionist theory all have something to offer the writer who seeks to improve his or her ability to think critically in writing, intended as an exchange between writer and reader. Some of the faculty of our Writing Program who were involved with this project devoted a good deal of thought and discussion to this line of reasoning.
What should count as critical thinking in the college context? In *The Philosophy of Literary Form*, the rhetorical literary theorist Kenneth Burke reminds us that “all questions are leading questions. . . . [They] need no ‘weighting’ other than the weighting implicit in the choice of topic itself” (1941/1973, p. 67). What do we want our writing prompts to lead our students to do? To write memos arguing for or against the purchase of a corporate jet? Or to think through questions that defy clear resolutions in provocative, original ways?

Our study suggests that informed readers respond to the latter rather than the former as evidence of critical thinking—an inducement to read further. The *cla* prompt has evidently been designed to incorporate numeracy: students are given charts and figures about the SwiftAir jet to include in their arguments concerning its purchase. How much weight should be given to the fact that the same model recently crashed, compared with other data and reports about its safety record? We value numeracy and respect the need for its assessment but chose to assess critical thinking on one side of the “two cultures” divide. The *cla* prompt makes critical thinking dependent on numeracy: a student who lacks facility with numbers is unlikely to score well, even if he or she can think creatively, with depth, for example, in linguistic analysis. We believe that the two sets of skills should be assessed separately, so that less attention is paid to utilitarian analysis and more is paid to the synthesis of original ideas.

We are not suggesting that student writers should incorporate ideas for the sake of their novelty. Responsiveness and consideration of alternative perspectives are always required for effective critical thought. Rather, we would side with the Irish poet James Stephens, who wrote, “Originality does not consist in saying what no one has ever said before, but in saying exactly what you think yourself” (see 2011).

**Technical Difficulties**

At the start of our study we decided to use Blackboard to collect student essays because it was the course management system supported by Instructional Technology Services on our campus. We hoped to create a system that students and faculty would consider a reasonable extension of their regular course work, with regular access to technical assistance, when necessary. It also needed to be supported by university resources to continue beyond the duration of the grant.

Unfortunately, we soon discovered that pulling essays out of Blackboard in a form that would be usable required more programming time than we had anticipated. Even worse, every time the Blackboard program was updated, we lost access to our collected essays, and the software required to retrieve them from the system had to be programmed all over again. Ultimately we were
forced to compensate Blackboard for retrieving the essays we had collected using its course management system. This resulted in a delay in completing the last part of our project.

This is not a challenge we have entirely overcome. On the basis of our experience we would not recommend the Blackboard course management system for collecting and storing student essays. We are working with our campus Instructional Technology Services to find an alternative.

Improving Instruction in Critical Thinking

In reviewing our results, we identified certain questions we thought could be answered by targeted readings of student essays. Collectively, the juniors and seniors enrolled in WRIT 340 demonstrated stronger skills in critical thinking than the freshmen writers in WRIT 140, but there was still a substantial number of juniors and seniors whose essays had not been scored in the upper half of the rating scale. What particular underdeveloped skills had prevented those writers from earning higher scores?

We used funds from the Spencer/Teagle grants to hire a writing instructor who served as a chief reader in our scoring sessions, to do a focused study, pulling up individual essays—written by juniors and seniors—that had fallen into the lower half of the rating scale and rereading them to learn what had prevented those essays from earning higher scores. After reviewing a large number of those essays, the writing instructor concluded that this clustering of essays resulted from reluctance among our scorers to recognize synthetic thinking at the paragraph level rather than in the whole essay, in the form of an integrative thesis. We adjusted the training of scorers in our next reading to address this reluctance, but the insight could also improve the prompts assigned in advanced writing classes with a professional focus.

After each reading we debriefed the scorers who had rated student essays. Their comments frequently turned to a discussion of writing prompts—both the assignments they used in their own classes and the prompts from their colleagues that had presumably elicited the essays. This was in part a consequence of our decision to assess critical thinking demonstrated in actual class assignments, rather than assigning a common prompt of our own design, as we had done in our preliminary study. The debriefing sessions made clear that reading hundreds of essays without a common assigned prompt allowed us to learn about the kinds of writing assignments that created occasions for (and elicited) the analysis and synthesis of original ideas.

The essays were "scrubbed" of identifying information, but were the raters able to guess the source of the essays anyway? Sometimes the answer was yes,
from the sophistication of the language, for example, but those guesses were often mistaken. The preponderance of our raters were drawn from the ranks of WRIT 140 instructors, who had little motivation to score essays by students in WRIT 340 higher than the essays written by their own students. One set of essays written by students in WRIT 340 most often received scores in the lower half of the scale. This assignment asked students in the arts to write about their own creative processes. The prompt did not call for consideration of a second perspective, although some student writers did in fact incorporate more than one. This result showed us that our raters were not assigning higher scores to essays they guessed had come from WRIT 340 students.

Analyzing hundreds of essays also helped us identify the kinds of prompts that invite students to demonstrate facility with critical thinking, as discussed under “Conceptual Outcomes,” above. We have tried to build this focus on questions of genuine uncertainty, the consideration of multiple perspectives, and the articulation of an original synthesis more systematically into our assignments and classroom pedagogy.

Certain changes have already been made in the curriculum we offer to all undergraduates at USC and in the training we provide graduate students who serve as writing instructors. We have redesigned a common assignment for all sections of WRIT 140, a course enrolling about eighteen hundred students each fall and twelve hundred students each spring. This redesign was intended to allow students more latitude for critical thinking as discussed above, and it incorporated a focus on “metacognition,” reflecting discussions at the annual meeting of the Spencer/Teagle grant recipients. Students are now asked to write about their own critical-thinking processes early in their careers at USC. The final assignment for WRIT 140 in spring 2011 and ancillary materials designed to help students respond to that prompt appear in Appendix E.

In addition, we have changed the information provided to writing instructors about the goals for each of their five assignments. We have shifted the materials we provide new instructors from assignment packets specific to different disciplinary areas to a Writing Program Tool Kit, with resources that should be applicable over a spectrum of academic emphases. And we have integrated instruction in writing assignment development into the graduate course that must be taken by all graduate students teaching writing for the first time.

Dissemination

Our project proved particularly useful and timely because it was conducted while USC underwent the process of reaccreditation by the Western Association
of Schools and Colleges. As one might expect, the accreditation association expressed sincere interest in the university’s commitment to academic program improvement and accountability through systematic assessment, and our project, which undertook to assess an outcome at the center of the educational enterprise, was given considerable attention in the site visits and written reports. This sort of external pressure is probably necessary for most institutions to think seriously about developing a culture of evidence on campus. However, we believe strongly that the forms of assessment should not be dictated to schools by outside agencies but, rather, must be created by the academic community itself, so that we measure what we genuinely believe our students need to learn. We have tried to design an assessment process that reflects the values of our faculty while it demonstrates that our students are learning what we have promised to teach them.

We ran a small study with the cooperation of faculty members at USC who were not involved with our scoring of essays. We gave ten faculty members in the humanities and social sciences copies of the essays used as “anchors” in our training of holistic scorers and asked them to rank order the eight papers according to whatever definition of critical thinking they used in their classes. There were no discussions about critical thinking, socialization, or prompts, and no guidelines were provided. Nine of the faculty returned their results in the time period requested. No two of them had ranked the eight papers in the same order, but the aggregate closely resembled the evaluations of our scorers. Our intent was to show that we are in fact measuring what the faculty consider important in their classes. In this study we discovered differences by discipline and across time: while

![Figure 5](image-url)
there was strong agreement as to whether the essays fell in the upper or lower half of the scale, within the upper half one paper (about the war in Iran) gained ground over another, apparently as a result of the changing public discourse.

To encourage faculty to participate in our study by collecting their students’ essays electronically, we asked an associate professor of writing (teaching) to create a blog interface, providing an opportunity for students to make their writing available to a wide online audience. The professor created the interface in Figure 5 and has made wonderful use of it in his own writing classes. Other members of our faculty have not yet taken advantage of it, but it is available to them.

Limitations of the Study

From the outset, in this project we have tried to balance the clarity of our research design with the possibility that our assessment method might be utilized by other college programs seeking to quantify the change that takes place in the critical thinking of their students over time. In an earlier study we used a more orthodox design, ensuring that all student essays would be written in response to the same prompt, chosen for the project, in the same amount of time. To score those essays, we used a rubric created not by us but by a researcher with a long history of work in the assessment of critical thinking. Our results in that study were encouraging, but we found that the specific conditions of the research design made the process less likely to be adopted by the college faculty and administrators we hoped to reach.

Timed writing to a supplied prompt has little to do with the current practice of teaching writing by instructors at USC and other progressive programs around the country, and it has less to do with the critical thinking students are likely to use in their professional careers or private lives. Neither the students who participated in our earlier study nor the instructors who allowed us to use their class time treated the conditions of that study as a regular part of their curriculum or as a task they would consider “embedding” in their course work. As a result, the effort made by students for that study was different from the effort they exerted on actual class assignments, and their essays did not reflect their best work. Nor were their teachers eager to incorporate a comparable assignment into the regular requirements of their classes.

If we hoped to develop an assessment method that would capture change in critical thinking convincingly for the faculty as well as the students, we would need to collect and evaluate essays written by students in response to genuine assignments by their classroom instructors. Those essays would be written in response to different prompts, under different conditions; and we would need to design a rubric that did not depend upon the particular writing prompt that had elicited the essays to be scored.
This meant that the project had to accomplish several goals at the same time. We had to (1) design and test a new scoring rubric, (2) use it to demonstrate change over time for a certain cohort of students, and (3) differentiate the critical-thinking ability evidenced by that cohort from that of a more advanced cohort in a single moment in time.

There was an obvious risk with this approach: if our results did not indicate clear differences and correspondences, we would have no way to determine whether the fault lay in the design of the rubric, a lack of difference in demonstrated skills, or the research design of our project. Fortunately, our results were statistically significant enough for us to argue that they could not have resulted from random chance: the rubric must be accurate enough, the difference over time sharp enough, and the skills of the two cohorts similar enough to produce the pattern of data collected score by score after hundreds of individual readings of student essays. But there were also limitations that arose from this balancing act.

For example, attrition is a factor that might have affected one dimension of our outcomes. In the last part of our study we compared the essays of students who persevered at USC with the essays of those same students as freshmen; attrition should not affect that comparison at all. In the first part of our study we compared the essays of those students as freshmen with their junior and senior colleagues at the time; attrition might have affected that comparison, if it had “weeded out” former freshmen from the junior pool, so that only the best thinkers remained. However, if that were the case, we might have expected to find a larger difference between those scores than we found between the two scores of the longitudinal cohort. The difference was in fact remarkably similar. Furthermore, the undergraduate attrition rate at USC is quite low: over 90 percent of fall freshmen graduate within five years. For these reasons we believe that our results were not affected by attrition in the junior class in the first year of our study.

A bigger limitation of the study from the standpoint of research design concerns the fact that student writers at all levels were responding to different essay prompts—actual assignments of their classroom instructors, assigned as part of their regular course work. Some of the prompts more easily invited critical thinking than others. In looking over the scored essays we did find prompts, for example, that did not require students to consider multiple points of view, such as a prompt that asked students to reflect on their own creative processes. However, this prompt was assigned at the upper-division (340) level, not at the freshman level, and the readers who scored these essays gave them suitably low scores.

All identifying information was removed from each essay, but in some instances readers were inclined to guess whether the writer was in fact a freshman or a junior. We tried to make some guesses ourselves and often discovered that we were wrong in our assumptions. Nevertheless, those guesses may have

Quantifying Learning in Critical Thinking
influenced some readers in the scores they gave the essays they read. However, our holistic scoring process was designed to catch any scores affected by extraneous information. Since each essay had to be read at least three times, by different readers, we could identify an aberrant score when one was assigned. Essays that showed gaps in scoring (those that received individual reader scores, e.g., of 1, 1, and 3) were read again by a chief reader, who had been trained for this purpose and given the authority to override an inappropriate score. Of course, the principal reason a reader was tempted to guess the status of a student writer was the quality of the essay at hand, so those guesses—right or wrong—rarely influenced the final scores.

Our decision to use essays regularly assigned by writing instructors posed a challenge for our data sampling as well. In the first year of our study, we asked all instructors who were willing to participate in the project to collect their students’ essays using Blackboard. We took a random sample of those essays, at both the freshman (WRIT 140) level and the junior (WRIT 340) level, and used the scores they received as our first set of data. In the third year of our study, we asked all junior-level (WRIT 340) writing instructors to collect student essays using Blackboard and searched through our data set for essays written by the same students two years before. We used all of those essays in our comparison during the third year of our study. The first-year essays were drawn from classes whose instructors volunteered to participate, and there may have been some outside factors influencing those instructors in their willingness to participate. However, we could hardly have gained access to essays written by students whose instructors were not willing to participate.

On the theoretical level a more intriguing limitation concerns the interaction of critical thinking with the ability to express one’s thoughts in writing. Separating those two closely connected abilities was the focus of much discussion in designing a rubric specifically for this study. We use a six-point rubric in assigning grades in our Writing Program, one that is designed to score general writing ability—which of course includes critical thinking as a constituent element. Teasing out critical thinking from other elements of good writing was a challenge. We began by asking a wide range of faculty to answer the following question: To what did they think they were responding when they judged the author of a piece of writing to have demonstrated good critical thinking? Originality of thought, a new idea, and the ability to spark interest in the reader were all qualities emphasized more than is usually the case in rubrics for general writing ability. But our success in differentiating between skill at critical thinking and skill at general writing remains a possible limitation of the study.

Finally, we encountered technical difficulties in collecting data that need to be solved before the process can be readily emulated by any other institutions.
We used Blackboard to collect student essays, which ended up involving additional programming, extra costs, and project delays. The USC Dornsife College instructional technology office has created an alternative online system for collecting and storing student essays, designed to complement the custom software we have used to present individual student essays anonymously for scoring. We have used it this year with the students in our Writing Program and are now engaged in scoring the essays we have collected in this way.

Conclusion

In a convocation address in 1997, Tom Gerety, president of Amherst College, summed up the intentions of his own college teachers when he told his students, “The bigger, more enduring question, the one you will have to ask yourself as long as you live is this: Are my thoughts and stances my own? Do I deserve the respect of these teachers, deserve it because I am what they deeply want me to be—not a clone of theirs or a convert but a free, independent thinker, an adult with the courage and tenacity to come to my own convictions?” (2007, p. 160). Any review of the academic requirements of undergraduates reveals the centrality of critical thinking in American colleges and universities. The mission statement of Harvard University encourages students “to respect ideas and their free expression, and to rejoice in discovery and in critical thought.” The Yale handbook informs freshmen that their distributional requirements provide competence in “the languages of thought” because “such competence gives you the tools to think critically and analytically, and to enlarge your imagination.”

Despite its centrality, critical thinking has proved difficult to assess in a form that is recognized by faculty as consistent with their learning objectives and consonant with academic values. Karl Shilling has called critical thinking “the Vietnam of assessment” with good reason, but it is essential that colleges and universities be able to demonstrate that our students are learning what we promise to teach them. If we cannot, other bodies will impose assessment methods that are more disruptive and less well suited to the core mission of higher education.

In this study we have created a method designed to capture the values of college faculty while it quantifies the added value of a college degree with respect to the acquisition of skill in critical thinking. We have endeavored to show that improvement in this area can be quantified, using the essays students write for their college instructors, and that a “snapshot” analysis conducted at one moment in time, comparing freshmen with their junior and senior colleagues, can be used to approximate the improvement in critical-thinking skills those freshmen will demonstrate over time at their school.
This method builds upon our belief that the faculty are the best assessors of their students’ progress toward the learning outcomes articulated for their program, if we can provide a context in which they are not assessing their own pedagogical performance. Our rubric was designed to be easily trainable, using a scale of four categories with recognizable criteria.

We prefer this method to the Collegiate Learning Assessment because our system uses essays assigned by college faculty as part of their regular class curricula, increasing the likelihood that students will do their best work. It disentangles mathematical reasoning from a specifically literate form of critical thinking that emphasizes a different set of analytical skills: we do not ask whether a particular corporate jet should be recommended for purchase but, rather, rely upon our faculty to pose questions eliciting more personal and significant analysis and synthesis. And our method allows us to explore the kinds of writing prompts that enable those articulations.

We would not argue that students’ improvement is solely a result of their experience in the Writing Program or the General Education Program but, rather, of their entire experience at the university. Improving the critical-thinking skills of undergraduates is a shared learning outcome of general education and writing programs, but responsibility for this area is not limited to the faculty who teach in those programs.

Rather, it is an enterprise for the faculty as a whole, although differently emphasized in the disciplines. Our method can be adapted for different dimensions of critical thinking, so that a rubric could be designed to assess a student’s facility with the particular learning outcomes identified for each discipline. David Pace, Leah Shopkow, and their colleagues in the History Learning Project at Indiana have done thoughtful work in this area, identifying the intellectual abilities that constitute historical perspective (Middendorf, Pace, Shopkow, & Diaz, 2007). We would argue that the method for assessing critical thinking explored in this study merits further attention because it respects the work done by college faculty teaching the most elusive kind of critical thinking—which is central to our mission and the learning objectives of all institutions of higher education.
Appendix A
Six-Point Critical Reasoning Scoring Rubric Developed at the University of Southern California Dornsife College

6 Exceptional

Genuinely impressive in the depth, agility, and integrity of reasoning.
The argument or analysis

. . . is consistently insightful in terms of its cogency, creativity, and depth.
. . . scrupulously recognizes relevant interests and viewpoints and consistently reflects intellectual skepticism, intellectual honesty, and intellectual integrity.
. . . consistently recognizes the complexities of the presumed issue or task and fully supports its reasoning using fully resonant integration of evidence and explanation.
. . . employs modes of reasoning that are entirely appropriate to the presumed issue or task; does so in a manner that demonstrates exceptionally capable reasoning, with no indication of gaps, inconsistencies, contradictions, or fallacies.

5 Strong

Generally admirable in its reasoning; may at some points rise to the Exceptional or dip to the Acceptable, but without dwelling in either.
The argument or analysis

. . . is generally cogent, insightful, and/or creative.
. . . generally recognizes relevant interests and viewpoints and consistently reflects fair and ethical discursive behavior.
. . . generally recognizes the complexities of the presumed issue or task and reliably supports its reasoning using an appropriate integration of evidence and explanation.
. . . employs modes of reasoning that are appropriate to the presumed issue or task; does so in a manner that demonstrates capable reasoning, with little evidence of gaps, inconsistencies, contradictions, or fallacies.

4 Acceptable

Generally acceptable, or slightly better, in its reasoning.
The argument or analysis
. . . is generally clear and may show occasional instances of insight or efforts at creativity; may also be marked by occasional lapses of insight.
. . . evinces an awareness of relevant interests and viewpoints and displays few instances of narrow-minded or marginally ethical discursive behavior.
. . . acknowledges many of the complexities of the presumed issue or task and provides adequate if incompletely integrated support of its reasoning.
. . . employs modes of reasoning that are largely appropriate to the presumed issue or task; does so in a manner that demonstrates competent reasoning, but with occasional instances of gaps, inconsistencies, contradictions, or fallacies.

3 Limited

Sometimes acceptable in its reasoning.
The argument or analysis
. . . offers claims that lack clarity and have little depth of insight; lapses of insight equal or outweigh instances of insight.
. . . is minimally aware of relevant interests and viewpoints and/or displays more noticeable tendencies toward narrow-minded or ethically questionable discursive behavior.
. . . acknowledges some of the complexities of the presumed issue or task but provides support for its reasoning that is insufficient at points and poorly integrated.
. . . employs modes of reasoning that are occasionally inappropriate to the presumed issue or task; does so in a manner that evinces instances of questionable reasoning disrupted by gaps, inconsistencies, contradictions, or fallacies.

2 Flawed

Frequently unacceptable in its reasoning.
The argument or analysis
. . . offers limited insight into, and only a superficial understanding of, its own claims.
. . . shows limited awareness of relevant interests and viewpoints and/or displays narrow-minded or unfair discursive behavior.

Richard Fliegel and John Holland
. . . overlooks most of the complexities of the presumed issue or task and provides support for its reasoning that is consistently weak and disconnected.

. . . employs modes of reasoning that are frequently unsuited to the presumed issue or task, with many instances of gaps, inconsistencies, contradictions, or fallacies in the line of reasoning.

1 Incompetent

Generally unacceptable in its reasoning.
The argument or analysis
. . . offers little insight and only a muddled understanding of its own claims.
. . . shows little or no recognition of relevant interests and viewpoints or is conducted in a fundamentally unfair or deceitful manner.
. . . largely fails to recognize the complexities of the issue or to provide minimally adequate support.
. . . employs modes of reasoning that are almost entirely unsuited to the presumed issue or task, and in a manner that is dominated by gaps, inconsistencies, contradictions, or fallacies in the line of reasoning.

[Special thanks go to Jack Blum.]
Appendix B
Four-Point Critical-Thinking Project Scoring Rubric Developed at the University of Southern California Dornsife College

4 Compelling Critical Thinking

The scorer finds the essay’s argument compelling for a particular insight or sophistication of ideas; the analysis results in a justified judgment concerning actions or beliefs.
The writer demonstrates facility with both analysis (differentiation) and synthesis (integration) of ideas: distinctions and conclusions are specific and proportionate to their evidence.
The writer appreciates the complexity of the issue, respects competing perspectives, and responds with intellectual honesty and an original point of view.
The claims are contestable but advanced by discrete ideas, the supporting evidence is well chosen, and the reasoning linking evidence to ideas is consistently sound.
The prompt or genre may not enable a strong personal thesis, but the essay demonstrates a thoughtful synthesis in its treatment of ideas and evidence.

3 Clearly Competent Argumentation

The scorer encounters a cogent argument or analysis, capably constructed, perhaps with some idea or element deserving of praise.
The writer engages in both analysis and synthesis: ideas are presented in arguments that contribute to an integrated perspective. The thesis may be modest in ambition, but the essay develops a central idea, methodically if predictably.
The writer recognizes the complexity of the issue, acknowledges other perspectives, and advances an individual point of view.
There is an explicit structure of claims, supporting ideas, and evidence, with reasoned connections between them, though there may be questionable gaps in conclusions or evidence.
An essay may be comprehensive but plodding, or imaginative but incomplete, yet in each instance the writer demonstrates an awareness of the elements necessary to advance a logical argument.
2 Underdeveloped Analysis

The scorer encounters an essay that makes distinctions but does not inte-
grate or synthesize ideas.
The writer demonstrates ability with analysis or synthesis but does not
coordinate them: ideas may be sorted into paragraphs, or a claim may
be advanced with evidence, but the essay lacks an organized structure of
ideas.
The essay demonstrates facility with parts of an argument, but only parts:
the paper might identify topics to be compared, or suggest original ideas,
but fails to articulate their relationship to a particular point of view.
The writer may recognize the complexity of the issue or alternative
perspectives but does not address them sufficiently or appropriately.
The paper may lack a contestable claim, or fail to demonstrate the
relationship between ideas and evidence, which might not be entirely
relevant or sufficient.

1 Rudimentary Reasoning

The scorer may sense an argument, but it is never articulated; the essay
contains occasional ideas, but they never cohere to form distinct lines of
argumentation or analysis.
The writer requires instruction in analysis and synthesis, in drawing clear
distinctions and conclusions: similar ideas appear in different parts of
the essay; unrelated ideas are grouped together; ideas and observations
may be repeated.
The essay lacks skepticism toward received ideas, fails to define the context
of the discourse, makes broad claims with scant evidence, or presumes
unearned conclusions.
The writer does not recognize the complexity of the issue, consider other
perspectives, or address them with intellectual integrity.
The limitation may be in the prompt, which requires no more than
rudimentary reasoning, but the score should reflect the evidence of
thinking on the page.

Richard Fliegel and John Holland, University of Southern California Dornsife
College, December 2011

Quantifying Learning in Critical Thinking 187
A total score of 3 indicates that an essay received three separate scores of 1 each from the three readers who scored each essay; a score of 4 indicates two scores of 1 plus one score of 2. Similarly, a score of 12 indicates three separate reading scores of 4 each. An essay that received a total score of 11 (4-4-3 from three readers) was not thought to have received a “split” score but, rather, to have been scored as slightly less successful than an essay with a straight score of 12. An essay that received two scores of 4 and one score of 2, or two scores of 1 and one score of 3, was considered a “split” to be reviewed by the chief reader, who also monitored the general pattern of each reader’s scores during the session to ensure interrater reliability.
Appendix D
Statistical Results

The 2011 scoring session included one ("initial") sample essay written by each student in fall 2008 and usually two ("final") sample essays written by the same student in 2011. The 2009 scoring session included essays written by freshmen enrolled in WRIT 140 and juniors and seniors enrolled in WRIT 340 in fall 2008.

An overview of the data is provided below. The chart below compares the scores awarded to essays written in 2008 by freshmen with the mean scores of the two essays written by the same students in 2011 as the principal result; other comparisons make up the subsidiary results:

### Essay Set

2011 Scoring Session (essays written by 265 students as freshmen and then as juniors or seniors)
- SI: Initial Papers (WRIT 140; mean score = 6.45) 265
- SF1: Final Papers—Group 1 (primarily WRIT 340 final portfolio papers) 262
- SF2: Final Papers—Group 2 (primarily WRIT 340 final portfolio papers) 235
- SFμ: Mean Final Scores (mean of Group 1 and Group 2 papers for each student, when available; otherwise the Group 1 or Group 2 paper itself; mean score = 8.50) 265
- SFMin: Minimum Final Scores (lower scoring of Group 1 and Group 2 papers, when available) 232

2009 Scoring Session (essays written by different freshmen, juniors, and seniors in fall 2008)
- Initial Papers (WRIT 140) 166
- Final Papers (WRIT 340) 157

Both Scoring Sessions
- All Papers 1,085
- All Readings (not counting chief reader adjudications) 3,255

### Chi-Square Analysis

All chi-square tests were run in a 2 × 5 matrix in which the two writing samples—initial and final—were run against five score categories—low, mid-low, mid, mid-high, and high—derived from the full range (3–12) of possible scores resulting from three readers using a four-point rubric.

Richard Fliegel and John Holland
Range | Initial Score (SI) | Final Score (SFm) |
---|---|---|
Low (3–4) |  |  |
Mid-Low (5–6) |  |  |
Mid (7–8) |  |  |
Mid-High (9–10) |  |  |
High (11–12) |  |  |

The resulting chi-square data thus have four degrees of freedom ($df = 4$); their significance can be estimated using the values given below:

\[
p \quad .050 \quad .020 \quad .010 \]
\[
\chi^2 (df = 4) \quad 9.49 \quad 11.67 \quad 13.28
\]

The principal and subsidiary chi-square results are given below. The difference in scores between freshman essays (in writ 140) and those written by juniors and seniors (in writ 340) is significant beyond the .01 level for each reading and for both together, while there are no significant differences in scores awarded to 140 (initial) papers, or to 340 (final) papers, when those from each scoring session are compared:

<table>
<thead>
<tr>
<th>Set Comparison</th>
<th>(\chi^2)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principal Result</strong>&lt;br&gt;2011 Scoring Session&lt;br&gt;SI × SFm Initial Papers and Mean of Final Papers</td>
<td>98.83</td>
<td>&lt;.01</td>
</tr>
<tr>
<td><strong>Subsidiary Results</strong>&lt;br&gt;2011 Scoring Session&lt;br&gt;SI × SFMin Initial Papers and Minimum Final Papers</td>
<td>35.92</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>SF1 × SF2 Group 1 Final Papers and Group 2 Final Papers</td>
<td>4.79</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>2009 and 2011 Scoring Sessions&lt;br&gt;SI09 × SI11 Initial Papers 2009 and Initial Papers 2011</td>
<td>2.85</td>
<td>&gt;.10</td>
</tr>
</tbody>
</table>

*Quantifying Learning in Critical Thinking*
The results of the chi-square analysis were confirmed using the Wilcoxon Signed-Rank Test, which found a significant difference between the freshman essays written in WRIT 140 in 2008 and the two sets of essays written by those students as juniors in 2011, with z-scores of –8.931, \( p = .000 \), for one set of 340 essays (Late) and z-scores of –8.957, \( p = .000 \), for the other set of 340 essays (Final). Paired t-tests showed no statistically significant difference between the two sets of 340 essays. They are identified as “340 Late” and “340 Final” in the following charts, but they were simply two batches of essays submitted for WRIT 340, sorted in no particular way:

<table>
<thead>
<tr>
<th>Essays</th>
<th>Rank</th>
<th>n</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRIT 340</td>
<td>Negative</td>
<td>53a</td>
<td>89.98</td>
<td>4,769.00</td>
</tr>
<tr>
<td>WRIT 340 Late</td>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 140</td>
<td>Positive</td>
<td>185b</td>
<td>127.96</td>
<td>23,672.00</td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Positive</td>
<td>170c</td>
<td>117.22</td>
<td>19,927.00</td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Late</td>
<td>Negative</td>
<td>83d</td>
<td>102.76</td>
<td>8,529.00</td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>27a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Late</td>
<td>Ties</td>
<td>37b</td>
<td>101.48</td>
<td>12,177.00</td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24f</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Ties</td>
<td>24i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Ties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Total</td>
<td>265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Total</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Total</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Total</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>Total</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340 Final</td>
<td>Total</td>
<td>240</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( a \) WRIT 340 Late < WRIT 140.

\( b \) WRIT 340 Late > WRIT 140.

\( c \) WRIT 340 Late = WRIT 140.

\( d \) WRIT 340 Final < WRIT 140.
\[ \text{WRIT 340 Final} > \text{WRIT 140}. \]
\[ \text{WRIT 340 Final} = \text{WRIT 140}. \]
\[ \text{WRIT 340 Final} < \text{WRIT 340 Late}. \]
\[ \text{WRIT 340 Final} > \text{WRIT 340 Late}. \]
\[ \text{WRIT 340 Final} = \text{WRIT 340 Late}. \]

<table>
<thead>
<tr>
<th>Essays</th>
<th>n</th>
<th>Mean Score</th>
<th>S.D.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRIT 140</td>
<td>265</td>
<td>6.45</td>
<td>1.806</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>WRIT 340</td>
<td>265</td>
<td>8.22</td>
<td>2.513</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Late</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRIT 340</td>
<td>240</td>
<td>8.57</td>
<td>2.615</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Final</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E
Revising the Final Assignment in Writing 140

When Professor Robert J. Thompson visited the University of Southern California to meet with some of those engaged in the Spencer and Teagle grant, he commented that assessment should not only reveal values and objectives but change them, and not simply through recommendations after the fact but through the process of assessment itself. Professor Thompson stressed that he had little interest in grant recipients producing yet another binder to be shelved and forgotten, and he encouraged us to approach our focused assessment of student writing not only as a means of demonstrating the development (or lack thereof) of students’ critical-thinking skills—our primary charge—but also as an opportunity to reconsider and revise the practices through which we seek to foster critical thinking in our curriculum and pedagogy.

One way we tried to take advantage of the opportunity offered through the Spencer and Teagle grants involved the final assignment given in Writing 140 (Writing and Critical Reasoning), our principal lower-division composition course. Each of our Writing 140 courses is taught in co-enrolled affiliation with a general education “Social Issues” lecture course, and for several years we had mandated a “retrospective” fifth assignment, one that would ask students to reflect upon the issues and concepts that they studied in their particular Social Issues affiliation. Following the lead suggested by Professor Thompson, we decided to focus the final assignment on critical reasoning itself and to make the assignment common to all Writing 140 affiliations: while different affiliations might “tweak” the assignment in such a way as to fit a particular topical context, the dimension of critical reasoning specified in the writing prompt was to remain stable across affiliations.

To achieve this fundamental stability without compromising the topical flexibility, a committee of faculty members representing a range of affiliations was appointed to develop the assignment. The committee decided to engage students by basing the assignment on an issue that was current, controversial, and immediately relevant to those embarking upon a college career. In spring 2011, the campus was abuzz with debate about a longitudinal study conducted by sociologists Richard Arum and Josipa Roksa, who cited college students’ limited improvement in critical reasoning as an indication that higher education institutions were “academically adrift”—the premise of their aptly titled book, Academically Adrift: Limited Learning on College Campuses. Our prompt proceeded to ask students to analyze how reasoning had gone adrift in our culture as a whole, not necessarily in academia. In fall 2011, the passing
of Steve Jobs, co-founder, chairman, and chief executive officer of Apple Inc., presented an opportunity to reflect upon the ways in which digital media might affect reasoning in our culture, for better or for worse.

The given issues were used as thematic springboards to conceive of and phrase the eventual assignment sheets, which are below. While the committee insisted that the issue remain consistent regardless of the affiliation to ensure that all students were faced with a writing task of equal difficulty (rendering the final assignment a genuine standard of comparison within the portfolio-grading session), instructors were encouraged to make minor changes in wording and topical emphases to fit their particular affiliations.

The revisions to our final assignment that were motivated through our participation in the Spencer and Teagle grants have had several significant consequences for the Writing Program. First and most important, refo-cusing the final assignment on critical reasoning per se has caused us to give greater curricular and pedagogical attention to dimensions of critical reasoning throughout the semester and in particular has provided a meta-cognitive opportunity at the culmination of the semester for students to reflect upon their own analytical abilities and habits of mind. At the same time, the final assignment also involves students in a common intellectual experience, one that is pondered and discussed both within and beyond the classroom. In addition, the final assignment has greatly improved the assessment value of the final portfolio, as each portfolio (consisting of an impromptu essay, a revised paper, and the final “critical reasoning” assignment) now contains an element that is common across all affiliations and all writing sections.

Provided below are the assignment sheets and ancillary materials distributed with the final assignments for spring 2011. The ancillary materials include prewriting suggestions and two glossaries of potentially useful terms. Instructors were encouraged to use the glossaries as a basis for generating similar affiliation-specific lists in class with their students. The model essays were intended to provide instructors with examples of relatively successful responses to the prompt but were not distributed to students, as we wished to avoid any suggestion of an approved “template” that students must or should follow.

Assignment 5: Responding to Reasoning Gone Adrift

PURPOSE
This semester you have been encouraged to practice the principles of sound critical reasoning, which include exercising intellectual skepticism and honesty,
using validity claims, identifying logical fallacies, addressing counterarguments, and balancing rhetorical appeals (see pp. 1–14 in the Writing 140 Course Book). This final assignment will provide you with an opportunity to showcase what you have learned by demonstrating these sound critical-reasoning skills in your own writing while critiquing the critical reasoning of others on the sorts of issues you have studied this semester in your Social Issues (SI) affiliation.

**TOPIC**
When surveyed about the current state of teaching and learning in our universities, faculties rank critical reasoning as the primary goal of a college education. Accordingly, efforts are being made to determine whether or not students actually gain proficiency in critical reasoning over the course of their academic careers. One such study is being conducted here at the University of Southern California, where preliminary data indicate that students’ critical-reasoning abilities do significantly improve between their freshman and junior years of college. Though the population is limited to University of Southern California and the findings are tentative, this study suggests that students are indeed learning, thinking, and writing competently about the complex issues facing our world.

Other studies suggest, however, that improvements in collegiate critical reasoning may be minimal or even nonexistent. Derek Bok, former president of Harvard University, claims that “many seniors graduate without being able to write well enough to satisfy their employers. Many cannot reason clearly or perform competently in analyzing complex, non-technical problems” (2006, p. 8). A recent study corroborating Bok’s claims, conducted by sociologists Richard Arum and Josipa Roksa, concludes that, in fact, “higher-education institutions are currently academically adrift.” Based on the data, Arum and Roksa believe that there is “barely” improvement in students’ “complex reasoning and writing during their first two years of college,” with almost half “demonstrating no appreciable gain in these skills” (2011, p. 54).

In any case, the fact that critical reasoning is being called into question within academia—a purported bastion of reasoning—raises larger concerns about the quality of critical reasoning in our culture as a whole. If critical-reasoning practices are under question in an academic context, where they are explicitly taught and supremely valued, then they are even more suspect in lay contexts (e.g., the media, social networks, political organizations), where social issues are discussed or debated with a much less pronounced commitment to critical reasoning.

**TASK**
You have contemplated a number of social issues this semester, and in doing so you have likely recognized instances of both sound (see the Purpose section

196 Richard Fliegel and John Holland
above) and weak (see the handout “Some Ways . . .”) reasoning. With the latter in mind, select an issue or set of issues derived from your SI affiliation that, in your view, has not received the degree of close and careful reasoning that it deserves. Then locate venues in our culture where the reasoning on your chosen issue(s) is evidently adrift (see the handout “Some Venues . . .”). Finally, assume the role of social commentator and respond to the following prompt in a five- to seven-page thesis-driven essay:

How is the reasoning on the given issue(s) adrift, and how might that reasoning be improved?

In planning and writing your essay, make the fullest relevant use of the material you have worked with this semester in the two affiliated courses. Keep in mind, however, that you may also refer to examples that were not considered in your affiliation so long as these have direct relevance to the issue(s) you have elected to analyze. If you choose to employ outside sources, take care to subordinate them to your own main line of reasoning. As always, sources should support, not submerge, your own point of view.

PREWRITING SUGGESTIONS
As with any invention process, keep in mind two fundamental objectives: You will want to develop more ideas than you can possibly use in your paper so that you will thereby have the opportunity to select a small and coordinated set of insightful concepts for actual use; in doing so, you will need to find some main idea (i.e., thesis) that the other concepts will serve to support and advance. In this regard, it is always good practice to record potential paragraph-level claims on a points-to-make list and to remain alert to the possibility that your main claim may arise from one of these points. To that end, here is one way of approaching this assignment, though this is by no means the only way:

Step 1 Review the issues you have encountered this semester in your SI affiliation, whether in the lecture, the discussion section, or Writing 140. As you do so, make a list of all these issues.

Step 2 Referring to the handout “Some Venues Where Reasoning Occurs,” consider the contexts in which you have heard these issues discussed or debated, making a list of all these venues.

Step 3 Referring to the handout “Some Ways in Which Reasoning Can Go Adrift,” evaluate the quality of reasoning by identifying themes within and across venues that perpetuate or exacerbate your chosen issue(s). (i.e., Ask yourself, “How is the reasoning on the given issue[s] adrift?”)

Quantifying Learning in Critical Thinking 197
Step 4 Based on your analysis, address what could be done to make the reasoning on your issue(s) more reasonable. There are multiple ways to present this call to action. You might propose it alongside the analysis in the body of the essay, or you might reserve it for the conclusion. Regardless, make a conscious effort to include it. (i.e., Ask yourself, “How might that reasoning be improved?”)
Some Venues Where Reasoning Occurs

In looking for venues where the social issues you have studied this semester are discussed or debated, you might want to consider the list below. Be aware, however, that this list is not exhaustive, the terms should not be used indiscriminately (e.g., topically), and the best essays will transcend these terms by bringing something original to the discussion.

**OFFICIAL INSTITUTIONS**

**Academic Settings:** equipping students (K–12 schools districts, universities) with the skills to explore and address social issues

**Governmental Settings:** proposing, debating, and voting on formal legislation

**PUBLIC ARENAS**

**Written Discourse:** carrying signs or posters, displaying bumper stickers, T-shirts, or other paraphernalia

**Physical Discourse:** marching, congregating, protesting, or otherwise making one’s presence felt

**Verbal Discourse:** participating in conversations, either informal (grocery store, barbershop, neighbors) or formal (government or law enforcement officials etc.)

**Intrafamilial Gatherings:** discussing social issues at the dinner table, holiday gatherings, etc.

**Interfamilial Gatherings:** discussing social issues outside the realm of one’s immediate family (religious groups, sports teams)

**MEDIA, TECHNOLOGY, AND ART**

**Traditional/One-Way Media:** reporting on current events via newspaper articles and editorials, radio and television updates, etc.

**Social/Interactive Media:** inviting discussion through blogs/microblogs, file sharing, social networking (Twitter, Facebook, “Comments” sections, etc.)

**Entertainment News Media:** presenting current events in a way that is satirical (*The Daily Show, The Colbert Report*) or sensational (*The O’Reilly Factor, Hannity*)

**Vehicles of Artistic Expression:** providing indirect commentary on social issues through the artistic avenues of film, literature, and music
Some Ways in Which Reasoning Can Go Adrift

In looking for examples of critical reasoning related to the social issues you have studied this semester, you might want to consider the list below. Be aware, however, that this list is not exhaustive, the terms should not be used indiscriminately (e.g., topically), and the best essays will transcend these terms by bringing something original to the discussion:

**Logical Fallacies** (see pp. 7–9 in the *Writing 140 Course Book*)

**Conspiracy:** claims that require no evidence and that reject all evidence to the contrary as “part of the conspiracy” itself (e.g., 9/11 attacks were planned by George Bush, Barack Obama is not a U.S. citizen)

**Certainty:** approaching argument from an absolute, close-minded position that cannot take other positions into consideration

**Conformity:** while conformity may be somewhat reasonable (and necessary) in a social context, it can also hinder one’s ability to think beyond the boundaries of social norms and expectations

**Distraction:** good reasoning requires sustained attention. What distracts us? (e.g., pop culture, Facebook updates, personal life, consumerism)

**Apathy/Laziness:** unwillingness to invest in the work required for sound reasoning

**Dependence on “Experts”:** wherein participants give up their agency to authorities (e.g., the Milgram electric-shock experiments in psychology; consider phrases like “I was just following orders”)

**Diffusion of Responsibility:** the refusal or inability to take personal responsibility for social issues may stem from the belief that others will take care of problems

**Scapegoating:** transferring accountability (and usually guilt) from one group or person to another, thereby obscuring the record of events pertinent to an issue and thus concealing genuine responsibility for outcomes

**Avoiding Conflict:** sometimes a heated debate or discussion may appear like a tense conflict to be avoided rather than joined (e.g., consider phrases like “I don’t want to rock the boat”)

**Metaphysical/Mythical Thinking:** reasoning that assigns responsibility to factors beyond the “human” (e.g., declaring Hitler evil/satanic and thus absolving us of the task of explaining Nazism as a real social phenomenon)
Self-Selection and Confirmation Bias: the practice of only participating in the conversation through filters that you yourself select—and that tend to confirm rather than challenge preconceived ideas (e.g., understanding American social issues only through the lens of Glenn Beck or Keith Olbermann, FOX News or MSNBC)

Barriers to Entry: if entering the conversation requires resources/qualifications/access available to only a select few (e.g., the poll tax in the Jim Crow South or running for office requiring independent wealth—Meg Whitman for governor of California or Michael Bloomberg for mayor of New York)

Factual Consensus: it is hard for the reasoning to advance if the parties cannot agree on the same set of basic facts (i.e., hard to argue productively about what to do if there is disagreement on what is happening)

Agreeing to Disagree: engaging in critical reasoning requires one to find common ground between contrasting ideas

Self-Consciousness: worries about how others perceive us can limit our engagement with social issues (e.g., being labeled a radical, flip-flopper, idealist)

Static Reasoning: that which doesn’t adapt to new social attitudes/practices/realsies (e.g., fighting the War on Terror with conventional, Cold War-era strategic thinking)

Tradition: reasoning that ignores tradition will not resonate with a public, and reasoning that is overly beholden to tradition will be static and discourage new ideas

Naive Romanticism: reasoning that is naively ideal cannot genuinely engage the complex realities of most social issues

Data-Driven: reasoning only concerned with raw numbers or cold instrumental analysis will be devoid of pathos and ineffective at dealing with real human issues

Motive: if the conversation is entered for reasons other than advancing the best answer to the issues at hand, problems can arise, e.g.,

Popularity/Profit: commercially driven reasoning that is primarily concerned with ratings and selling advertisements

Sabotage/Corrupt: the goal is not to advance the discussion but, rather, destabilize or discredit it

Pleasure Principle: when the drive for pleasure is the chief motivating force in behavior it may affect reasoning.
We would like to express our appreciation to all of the people whose assistance and support allowed us to conduct our project. Robert Thompson of Duke University was the guiding intelligence behind the national set of projects in this program. The insights that he offered during his site visit and our annual meetings helped us develop our original idea in ways that are much more likely to be sustainable.

The contributions of other fund recipients, scholars and researchers from around the country, were also essential to the evolution of our thinking. We are grateful to the Spencer and Teagle foundations, and to Robert Connor and Michael McPherson in particular, for their continuing support and vision over these last three years.

In addition, we need to thank the many people at the University of Southern California who assisted us throughout the life of the project. Vice Provost for Undergraduate Programs Gene Bickers made the initial connection. Our provost, Elizabeth Garrett, has been unfailingly supportive, as have Dornsife College Dean Howard Gillman and Vice Dean for Academic Programs Steven Lamy. Sidney Harmon provided a valuable perspective as we developed our rubric. John Parker and Susan Chang of the Dornsife College Instructional Technology staff supplied essential pieces in the technical chain, and we could not have begun or completed the project without help from the university’s Instructional Technology Services, including Russ Kaurloto, Leon Nguyen, Otto Khera, Patrick Crispen, and Susan Metros. Geoffrey Middlebrook and Robin Romans helped coordinate that relationship. We are grateful to Jack Blum and Janice Schafrik for calculating the significance of our results, to Denise Wiegel and Kim Price for helping us manage our budgets, and for many other forms of staff support to Marsha Chavarria-Winbush and Maria Velasquez of the General Education Office and Barbara Leaks and Lettie Littlejohn of the Writing Program.

Finally, but most important, we thank the faculty members and graduate students who participated in our training and scoring sessions, especially Chief Readers Lacey Donohue and Kevin Egan, as well as readers LauraAnne Adler, Jennifer Ansley, Justin Bibler, Mary Byram, Samantha Carrick, Jeffrey Chisum, Mary Ann Davis, Elizabeth Durst, Caley O’Dwyer Feagin, Alicia Garnica, Devin Harlow, Trisha Hubbs, Troy Johnson, Kate Levin, Kevin Long, Andrew Loschert, Barbara Mello, Indra Mukhopadhyay, Lindsay Nelson, Eric Rawson, Saba Razvi, Suraj Shankar, and John (JJ) Strong.
References


