**Report on the Assessment of Undergraduate Critical Thinking**

**California State University, Fresno**

**May 11th, 2021**

**Background**

California State University, Fresno initiated campus wide assessment of the WSCUC five core competencies in 2015. The former Director of Assessment, Melissa Jordine, established a timeline whereby an ad hoc committee is formed and carries out an assessment of one core competency every year. These efforts are now overseen by the current Director of Assessment, Dr. Douglas Fraleigh. The core competencies include Oral Communication (evaluated AY-2016-2017), Quantitative Reasoning (evaluated AY 2017-2018), Information Literacy (evaluated AY 2018-2019), Written Communication (evaluated AY 2019-2020), and Critical Thinking. During the 2020-2021 academic year, the Critical Thinking Core Competency was assessed.

**The Critical Thinking Assessment**

The critical thinking assessment was overseen by the Director of Assessment, Dr. Douglas Fraleigh, and led by two College Assessment Coordinators, Dr. Paul Price and Dr. Monica Summers. In Spring 2020, an ad hoc committee consisting of Dr. Price, Dr. Summers, Dr. Venita Blackburn, Dr. Raymond Hall, Dr. Ines Lima, Dr. Sunantha Teyarachakul, and Dr. Yuanyuan Xie met to revise a draft instrument that had been created by an earlier version of the committee. The final Fresno State Critical Thinking Assessment (CTA) took the form of a 22-item multiple-choice test that consisted of three distinct sections. The argument analysis section consisted of 11 items that required reading, interpreting, and evaluating arguments that had been presented in actual letters to the editor of the *Fresno Bee*. The formal logic section consisted of 7 items that required drawing conclusions from simple deductive arguments and recognizing the formal similarity between arguments with different content. The facts vs. opinions section consisted of 4 items that required determining whether a statement was about a fact or an opinion. The decision was also made to include a 20-item Inventory of Epistemically Unwarranted Beliefs (Dyer & Hall, 2019) for exploratory purposes only. This instrument requires rating one’s degree of belief in claims about things like the existence of Bigfoot and alien abductions.

The final version of the CTA was approved by the Department of Psychology IRB and, due to the Covid-19 pandemic, disseminated to students online. The instructors of 24 different courses, across six of the university’s eight schools and colleges, agreed to offer it in their classes for extra credit. There were 521 students who attempted the CTA. However, the target population was students within three semesters of graduation, which reduced the final usable sample to 283. (Students who did not respond to all 22 items on the CTA were also dropped from the analyses because it is generally impossible to tell whether they chose to skip the items, skipped them accidentally, or had technical issues.)

The benchmarks for this assessment were scores of 70% overall and scores of 70% on each of the three sections of argument analysis, formal logic, and facts vs. opinions. The Inventory of Epistemically Unwarranted Beliefs was included for exploratory purposes only, so no benchmark was established for it.

**Results of the Critical Thinking Assessment**

Sample Characteristics

As shown in Table 1, the students represented six of the eight colleges at Fresno State. In terms of transfer status, first generation status, gender, and race/ethnicity, this sample is reasonably representative of the student body as a whole.

**Table 1.** *Characteristics of the Main Sample (*N *= 294)*

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Number** | **Percentage** |
| **College / School** |  |  |
| Arts and Humanities | 15 | 5.1 |
| Business | 61 | 20.7 |
| Engineering | 21 | 7.1 |
| Health and Human Services | 38 | 12.9 |
| Science and Mathematics | 85 | 28.9 |
| Social Sciences | 73 | 24.8 |
| **Transfer** |  |  |
| Yes | 140 | 49.5 |
| No | 141 | 49.8 |
| Not Reported | 2 | 0.7 |
| **First Generation** |  |  |
| Yes | 213 | 75.3 |
| No | 66 | 23.3 |
| Not Reported | 4 | 1.4 |
| **Gender** |  |  |
| Female | 185 | 65.8 |
| Male | 95 | 33.6 |
| Nonbinary | 1 | 0.04 |
| Not Reported | 2 | 0.07 |
| **Race / Ethnicity** |  |  |
| African American | 7 | 2.5 |
| Asian America | 24 | 8.5 |
| East Indian | 1 | 0.4 |
| European American | 44 | 15.5 |
| Latino | 167 | 59.0 |
| Middle Eastern | 6 | 2.1 |
| Native American | 1 | 0.4 |
| Native Hawaiian / Pacific Islander | 1 | 0.4 |
| Other / Multiple | 32 | 11.3 |

Overall CTA Scores

We began by computing the percentage of items answered correctly by each student. Table 2 shows the overall mean and standard deviation, along with the percentage of students meeting the benchmark of 70%. Table 2 also shows these statistics further broken down by transfer status, first-generation status, gender, and race/ethnicity.

**Table 2.** *Mean and Standard Deviation of Percent Correct and Percentage Meeting Benchmark*

|  |  |  |  |
| --- | --- | --- | --- |
| **Subset of the Sample** | **Mean** | **Standard Deviation** | **Percentage Meeting Benchmark** |
| **All Students** |  |  |  |
|  | 65.7 | 16.6 | 41.2 |
| **Transfer** |  |  |  |
| Yes | 65.4 | 16.6 | 39.5 |
| No | 65.9 | 16.6 | 42.8 |
| **First Generation** |  |  |  |
| Yes | 64.1 | 16.1 | 36.2 |
| No | 70.8 | 17.2 | 56.5 |
| **Gender** |  |  |  |
| Female | 65.1 | 15.6 | 38.0 |
| Male | 66.5 | 18.4 | 46.9 |
| **URM** |  |  |  |
| Yes | 64.2 | 15.9 | 35.5 |
| No | 73.4 | 18.0 | 71.7 |

The first thing to note is that the overall mean percentage correct was only 65.7% and that only 41.2% of the students met the benchmark. The second thing to note is that first-generation and URM status seem to matter, with first-generation students and URM students scoring somewhat lower on average and meeting the benchmark at a somewhat lower rate. Because first-generation status and URM status are correlated with each other, we also conducted a simple least-squares regression analysis to disentangle their roles. In this analysis, the effects of both first-generation status and URM status were statistically significant, indicating that they both play a role. First-generation status resulted in scores that were about 5 percentage points lower on average, while controlling for URM status. And URM status resulted in scores that were about 7.5 percentage points lower on average, while controlling for first-generation status.

CTA Section Scores

We also looked separately at scores for the argument analysis, formal logic, and facts vs. opinions sections of the CTA. Table 3 shows the means, standard deviations, and percentages of students meeting the benchmark for each of these three sub-scores. Here we see that the students did about as well at argument evaluation as they did overall. However, they did especially well at distinguishing facts from opinions, with 92% of them meeting the benchmark. And they did especially poorly at formal logic, with only 37% of them meeting the benchmark.

**Table 3.** *Mean and Standard Deviation of Percent Correct and Percentage Meeting Benchmark* *for Each of Three Sub-Scores*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mean** | **Standard Deviation** | **Percentage Meeting Benchmark** |
| **Sub-Score** |  |  |  |
| Argument Analysis | 65.9 | 16.5 | 43.0 |
| Formal Logic | 57.2 | 23.1 | 37.0 |
| Facts vs. Opinions | 91.6 | 17.7 | 92.0 |

CTA Scores and Time to Graduation

If critical thinking is improving over students’ time at the university, we might expect that those who were closer to graduation would perform better on the CTA. This did not turn out to be the case. The mean score for students who were *not* within three semesters of graduation was actually slightly higher (68.5%), as was the percentage of students meeting the benchmark (52.0%). These differences are not quite statistically significant by conventional standards and could very well represent random variation. However, it is also possible that they reflect the fact that the more advanced students are further removed from having taken their required lower-division critical thinking course (GE Area A3).

Correlation with Epistemically Unwarranted Beliefs

Turning to the Inventory of Epistemically Unwarranted Beliefs, we computed three scores for each student. One was the mean agreement rating across the 5 epistemically warranted beliefs, one was the mean belief rating across the 15 epistemically unwarranted beliefs, and one was the difference between these two scores. Here we found that the percentage correct on the CTA was moderately positively correlated with the strength of epistemically warranted beliefs (*r* = +.349), moderately negatively correlated with epistemically unwarranted beliefs (*r* = -.363), and strongly positively correlated with the difference (*r* = .469). It is notable that measures of two very different conceptions of critical thinking (the ability to evaluate arguments vs. the strength of belief in epistemically warranted and unwarranted claims) are correlated to a degree that is suggestive of an important connection between the two.

Other Notes on the CTA

It is also worth noting that the entire 22-item CTA had reasonably good (but not great) internal consistency, with a Cronbach’s α of .71. What this indicates is that answering one item correctly is positively correlated with answering other items correctly, so that it is at least somewhat defensible to interpret the total score as a measure of a single underlying construct like “critical thinking.” This is also consistent with the fact that the scores on the three sub-scales were reasonably strongly correlated with each other. The argument analysis and formal logic sub-scores were correlated at *r* = +.478. The argument analysis and facts-opinions sub-scores were correlated at *r* = +.345. And the formal logic and facts-vs.-opinions sub-scores were correlated at *r* = +.256.

**Conclusions and Recommendations**

The results of the Critical Thinking Assessment were discussed in two committees. The first was the group of College Assessment Coordinators (including Drs. Fraleigh, Price, and Summers). The second was the above mentioned ad hoc Critical Thinking Committee. The conclusions and recommendations below come directly from those discussions.

General Conclusions

We recognize that there are many different conceptions of critical thinking (e.g., general vs. discipline-specific, skills vs. attitudes, and so on) and that no single assessment is likely to get at all of them. However, we are also confident that the present approach addresses one important aspect of critical thinking--the ability to understand and analyze informal and formal arguments--and that it does so in a valid way. Thus we believe it is reasonable to conclude that Fresno State undergraduates’ critical thinking skills are less developed, and appear to improve less over time, than would be ideal.

Recommendations

The committees came up with a number of recommendations that fall into two categories: strengthening the teaching of critical thinking and strengthening the assessment of critical thinking.

* Strengthening the Teaching of Critical Thinking
  + The University’s GE Committee should revisit the learning outcomes for GE Area A3 (Critical Thinking). Perhaps these learning outcomes need to be clarified or elaborated. And the results of this effort should be communicated clearly to instructors who teach A3 courses.
  + The extent to which critical thinking is reinforced *after* students have completed their required lower-division GE course in critical thinking should be assessed. It could be that it receives very little emphasis in the junior and senior years. Upper division courses that explicitly include critical thinking as a learning outcome should be available across the curriculum. Assignments in these courses that aligned with critical thinking learning outcomes would be a resource for core competency assessment.
  + There should be more professional development opportunities for faculty who want to learn to teach critical thinking more effectively, both at the lower-division and upper-division levels. Professional development opportunities should include best practices for teaching critical thinking to our diverse student body and achieving equity in assessment results.
* Strengthening and Extending the Assessment
  + The current instrument should continue to be used in the future but perhaps with an additional qualitative element to help understand how students are engaging with it and why they are falling short of expectations.
  + Future assessments should continue to break out results by characteristics such as first-generation students, gender, and ethnicity to provide evidence of the results of changes in critical thinking instruction and indicate whether further changes to achieve equity are warranted.
  + Future assessment teams should explore alternative conceptions and measures of critical thinking, including measures of attitudes such as skepticism and tolerance of uncertainty.
  + The current data set should be made available to other faculty who are interested in exploring it in more detail for research and assessment purposes. Many of the students consented to the use of data from their university records, which would allow for a more fine-grained analysis of the results (with respect to things like SAT scores, specific courses taken, and so on).