

Assessing Critical Thinking in GE

In Spring 2016 semester, the GE Curriculum Advisory Board (CAB) engaged in assessment of Critical Thinking (CT) across the General Education program. The assessment was focused on Upper Division GE courses to assess CT “at or near graduation.” The assessment followed a two-pronged approach:

1. Faculty were invited to identify a “signature assignment” that demonstrated students’ critical thinking abilities. Students uploaded papers to the STEPS web-based assessment system. Two members of CAB read each paper, assessing the work with an agreed upon rubric (see Attachment A). In the case of widely divergent scores, a third reader assessed the paper.
2. Faculty who did not participate in the “signature assignment” assessment were asked to allow CAB members to administer a “CT Quiz” in their classes. Participant faculty agreed to provide extra credit to students proportional to their performance on the quiz. That is, the better students scored on the quiz, the more extra credit points they received. The CT Quiz was a multiple choice test derived from a CT assessment designed by faculty and carried out in Area A3 courses in Fall 2012 and 2013 using a pre- and post-test format. Nine questions were drawn from the Fall 2012 assessment and 11 questions were drawn from the Fall 2013 assessment. See Attachment B.

Signature Assignment Assessment

Participants

101 student papers were selected from fall 2015 **General Education courses** and were assessed for critical thinking. These papers came from the following courses:

- | | | |
|-------------|-----------|-----------|
| • ECON 365 | fall 2015 | 13 papers |
| • GEOS 330Z | fall 2015 | 27 papers |
| • HUMN 300Z | fall 2015 | 23 papers |
| • PHIL 370 | fall 2015 | 38 papers |

Measure

Critical thinking was assessed according to five criteria: *avoids weak arguments, conclusions, identifies broader conditions, identifies issues, and use of evidence.*

Two different readers separately assessed papers using a scale of 1 to 3 to assess critical thinking in writing: Beginner (1), Competent (2), Accomplished (3). The scores from the two readers were then averaged to form one composite score. For example if a student received a 2 from the first reader and a 3 from the second reader the student received an overall score of 2.5 for that criteria. Papers went to a third reader if scores varied by more than 1 point. Average scores from each of the five criteria were then added to create a total critical thinking score for each student. For example if a student received a 2 on *avoids weak arguments*, a 2 on *conclusions*, a 2 on *identifies broader conditions*, a 2 on *identifies issues*, and a 2 on *use of evidence* the student received an overall score of 10. The lowest possible score is 5 and the highest possible score is 15.

Average Scores

Average total scores for the 101 student papers was 10.31, $SD = 2.64$, the lowest score was a 5 and the highest score a 15. Average scores on the five criteria of critical thinking: *avoids weak arguments* $M = 1.98$, $SD = .58$, *conclusions* $M = 1.85$, $SD = .61$, *identifies broader conditions* $M = 2.16$, $SD = 6.1$, *identifies issues* $M = 2.23$, $SD = .59$, and *use of evidence* $M = 2.08$, $SD = .59$.

Gender

To determine if there is a difference between genders on mean total critical thinking scores (5-15) an independent samples t -test was conducted between female ($N = 34$, $M = 10.65$, $SD = .44$) and male students ($N = 67$, $M = 10.14$, $SD = .33$). There was not a statistically significant difference between the two groups $t_{(99)} = .906$, $p \leq .367$. These results do not indicate a significant difference between females and males on their mean total critical thinking scores.

Enrollment Status

To determine if there is a difference in enrollment status on mean total critical thinking scores (5-15) an independent samples t -test was conducted between transfer students ($N = 54$, $M = 10.03$, $SD = .36$) and native students ($N = 44$, $M = 10.61$, $SD = .41$). There was not a statistically significant difference between the two groups $t_{(96)} = 1.068$, $p \leq .288$. These results do not indicate a significant difference between transfer and native students on their mean total critical thinking scores.

Remediation Status

To determine if there is a difference in remediation status on mean critical thinking scores (5-15) an independent samples t -test was conducted between students who took one or more college remediation courses ($N = 16$, $M = 11.55$, $SD = .260$) and students who did not take college remediation courses ($N = 84$, $M = 10.11$, $SD = 2.60$). The results show a statistically significant difference between the two groups $t_{(98)} = -2.021$, $p = .046$. The t value of -2.021 indicates a statistically significant difference on total critical thinking score means ($p = .046$) for students who took one or more college remediation courses and students that did not take college remediation courses. *The results show that students who took one or more college remediation courses had significantly higher total critical thinking scores than students who did not take college remediation courses.*

Class Status

A one-way analysis of variance (ANOVA) was conducted to compare mean total critical thinking score (5-15) across class status (freshmen, sophomore, junior, senior). Since only two students reported being “freshmen”, this group was combined with the 14 students in the “sophomore” group, leaving three comparison groups (lower classmen, junior, senior). There was homogeneity of variances, as assessed by Levene’s test of homogeneity of variance ($p = .225$). Mean total critical thinking scores were not significantly different across class status groups, $F_{(2,98)} = .425$, $p = .655$, partial $\eta^2 = .008$. These results do not indicate a significant difference in mean total critical thinking scores between lower classmen, juniors and seniors.

Table 1. Class Status Means and Standard Deviations on Total Critical Thinking Scores

Class Status	<i>N</i>	<i>M</i>	<i>SD</i>
Lower Classmen	16	10.07	3.11
Juniors	40	10.12	2.34
Seniors	45	10.59	2.76

College

A one-way analysis of variance (ANOVA) was conducted to compare mean total critical thinking score (5-15) across eight colleges (College of Agriculture, College of Behavioral and Social Sciences, College of Business, College of Communication and Education, College of Engineering Computer Science and Construction Management, College of Humanities and Fine Arts, College of Natural Sciences, and Undergraduate Education). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p=.816$). Mean total critical thinking scores was not significantly different across college groups, $F_{(7,93)} = .691, p = .679$, partial $\eta^2 = .05$. These results do not indicate a significant difference in mean total critical thinking scores across eight colleges.

Table 2. College Means and Standard Deviations on Total Critical Thinking Scores

Class Status	<i>N</i>	<i>M</i>	<i>SD</i>
Agriculture	2	9.08	2.95
Behavioral & Social Science	16	10.72	2.38
Business	18	9.89	2.85
Communication & Education	16	9.63	3.02
Engineering Computer Science & Construction Management	15	10.51	2.73
Humanities & Fine Arts	15	11.19	2.27
Natural Sciences	17	10.44	2.57
Undergraduate Education	2	8.58	3.65

Ethnicity

A one-way analysis of variance (ANOVA) was conducted to compare mean total critical thinking scores (5-15) across ethnic groups (not specified, Asian, Black/African American, Hispanic/Latino, two or more Ethnicities/Races, White). Since only one student identified as Asian, the Asian Ethnic group was dropped from the analysis, leaving five comparison groups. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p=.231$). Total critical thinking scores was not significantly different across Ethnic groups, $F_{(4,95)} = .464, p = .762$, partial $\eta^2 = .02$. These results do not indicate a significant difference in mean total critical thinking scores across ethnic groups.

Table 3. Ethnic Group Means and Standard Deviations on Total Critical Thinking Scores

Ethnic Group	<i>N</i>	<i>M</i>	<i>SD</i>
Not specified	9	10.87	2.58
Black/African American	3	11.50	.87
Hispanic/Latino	15	9.69	2.00
Two or more Ethnicities/Races	7	10.40	2.35
White	66	10.39	2.84

Critical Thinking Quiz

The CT Quiz is designed to cover the standard dimensions of critical thinking: identifying issues, assessing the relevance of evidence, detecting flawed logic and similar “habits of mind.” It does not focus on formal logic. And it endeavors to use examples from everyday life that reflect the kinds of decisions and judgments students (and others) make in assessing a variety of arguments and assertions. As a home-grown test, it is not nationally normed. For comparison we have the pre- and post-test scores from the administration of the test in Fall 12 (scores from 1,194 students) and Fall 13 (scores from 1,237 students). The Fall 12 and 13 test-takers were mostly freshman and sophomores. In the discussion that follows, the GE CT Quiz 2016 results are compared with *post-test* scores from the Fall 12 and 13 administration.

CT Quiz administration

The CT Quiz was administered in 13 classes between April 21 and May 16, 2016. The classes were associated with all GE Pathways except Diversity. See Table 1. CAB faculty administered the CT Quiz with the cooperation of class instructors. Students received an explanation of the purpose of the assessment using a standard script. Students were given 20 minutes to complete the quiz. A total of 439 valid responses were obtained. The students used a scantron sheet to record their answers and provided their name and student number so that responses could be linked to social and educational information such as class level, major, native vs. transfer status, etc. Of the valid responses, 31 identified as freshman, 72 as sophomores, 161 as juniors and 174 as seniors.

CT Quiz results

The overall average score on the CT Quiz was 11.14 answers correct or 55.6% (Table 2, Column F). The lowest score was 2 and the highest score was 20. This compares to an average score on the same questions of 56.4% in Fall 12 and 74% on the Fall 13 post-test quizzes. A question by-question comparison is provided in Column G, which records the difference score between the Fall 12 or 13 post-test scores and the GE Quiz administered in Spring 16. For all questions but two, the percentage of correct answers was *lower* on the Spring 16 CT Quiz compared to student performance on the same questions in the earlier administration of the quiz.

Class	Enrollment	GE Area	Pathway
HCSV 370	39	Social Sciences	Health & Wellness
HIST/WMST 335	48	Humanities	Gender & Sexuality
HIST/MEST 363	25	Humanities	International Studies
NFSC 310	43	Natural Sciences	Food Studies, Global Development
ENGL 350	37	Humanities	Science, Tech & Values
NSCI 300	20	Natural Sciences	Great Books & Ideas
PSYC 321	44	Social Sciences	Science, Tech & Values
BIOL 318	89	Natural Sciences	Health & Wellness
POLS 365	33	Social Sciences	Ethics, Justice & Policy
POLS 365	30	Social Sciences	Ethics, Justice & Policy
HIST/MEST 363	47	Humanities	International Studies
HIST 341	50	Humanities	Food, Sustainability
HIST 341	50	Humanities	Food, Sustainability

The question with the largest *positive* point differential (+ 11 percentage points) was Q. 10 on the Spring 16 Quiz, which matches with Q. 15 from Fall13:

10. The Affordable Care Act (“Obama care”) requires most adults to purchase a health insurance policy. Advocates of the requirement say it will reduce the costs of health insurance for the majority of people. Critics argue that forcing people to buy health insurance is like forcing them to eat broccoli, which, they say, is clearly unconstitutional.

Which of the following assertions is logically the most relevant criticism of the critics’ argument?

- a. The critics are inconsistent in attacking the individual mandate because in fact they were the ones who originally proposed it.
- b. The critics do not genuinely oppose the individual mandate; they are simply mouthing opposition to embarrass political opponents.
- c. It has not been conclusively established that broccoli or exercise makes people healthier.
- d. There is a difference between forcing people to do something for their own good and compelling them to act for the greater good.

Correct answer is d.

The question with the largest *negative* point differential (-28 percentage points) was Q. 11 on the Spring 16 Quiz, which corresponds to Q. 15 from Fall 12:

11. “Hey, one English course won’t turn anyone into a perfect writer so there is no reason we should be required to take one.”

Which of the following statements is true about the speaker’s reasoning?

- a. The speaker offers the listener a false choice.
- b. The speaker offers a strong argument.
- c. The speaker assumes what she is trying to prove.
- d. The speaker tries to play on the listener’s emotions

Correct answer is a.

Table 2. Results of Spring 16 CT Quiz Compared to Fall 12, 13 Results

(A) F12 question#	(B) F12 % correct	(C) F13 question#	(D) F13 % correct	(E) Sp16 question#	(F) Sp16 % correct	(G) Sp12/13% - Sp16%
12	71			1	63	-8
8	60			2	37	-23
13	58			3	47	-11
		11	94	4	91	-3
		5	93	5	91	-2
		6	83	6	80	-3
		7	55	7	41	-14
		9	89	8	77	-12
		14	72	9	65	-7
		15	71	10	82	11
15	55			11	27	-28
17	47			12	32	-15
19	50			13	30	-20
		16	72	14	51	-21
		19	78	15	70	-8
4	45			16	27	-18
		18	70	17	70	0
3	44			18	43	-1
		8	37	19	20	-17
7	78			20	68	-10
Average	56.4		74.0		55.6	-10.5

There was a negative 21 point differential on Q. 2:

2. Dr. Mella has five brothers, all Democrats. Chances are he too is a Democrat.

Which of the following suppositions increases the strength of this inference, everything being equal?

- a. Dr. Mella's brothers all live in Dallas—Dr. Mella lives somewhere else.
- b. Dr. Mella's brothers all live in different parts of the country.
- c. Dr. Mella is the oldest brother.
- d. Dr. Mella and his brothers are all over six feet tall.

Correct answer is b.

Gender

To determine if there is a difference in gender on mean critical thinking scores (0-20) an independent samples t -test was conducted between female ($N = 230, M = 10.73, SD = 3.15$) and male students ($N = 208, M = 11.58, SD = 3.23$). The results show a statistically significant difference between the two groups $t_{(436)} = -2.791, p = .005$. The t value of -2.791 , indicates a statistically significant difference in mean critical thinking scores ($p < .000$) between male and female students. *Male students scored significantly higher than female students.*

First Generation

To determine if there is a difference on mean critical thinking scores (0-20) between first generation college students ($N = 214, M = 10.547, SD = 3.12$) and non-first generation college students ($N = 185, M = 11.703, SD = 3.21$) an independent samples t -test was conducted. The results show a statistically significant difference between the two groups $t_{(397)} = -3.639, p < .000$. The t value of -3.639 , indicates a statistically significant difference in the total critical thinking score means ($p < .000$) between first generation college students and non-first generation college students. *The scores of non-first generation college students are significantly higher than first generation college students.*

Remediation Status

To determine if there is a difference in remediation status on mean critical thinking scores (0-20) an independent samples t -test was conducted between students who took one or more college remediation courses ($N = 96, M = 9.44, SD = .24$) and students who did not take college remediation courses ($N = 343, M = 11.61, SD = .17$). The results show a statistically significant difference between the two groups $t_{(437)} = 6.107, p < .000$. The t value of 6.107 indicates a statistically significant difference on critical thinking score means ($p < .000$) for students who took one or more college remediation courses and students that did not take college remediation courses. *Students who did not take college remediation courses had significantly higher mean critical thinking scores than students who took one or more college remediation courses.*

Enrollment Status

To determine if there is a difference in enrollment status on mean critical thinking scores (0-20) an independent samples *t*-test was conducted between transfer ($N = 166, M = 10.93, SD = 3.10$) and native ($N = 265, M = 11.24, SD = 3.20$) students. There was not a statistically significant difference between the two groups $t_{(429)} = .970, p = .332$. These results do not indicate a significant difference in mean critical thinking scores between transfer and native students.

Class Status

A one-way analysis of variance (ANOVA) was conducted to compare mean critical thinking scores (0-20) across class status (freshmen, sophomore, junior, senior). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p = .360$). The results show that critical thinking score means are significantly different across class status groups, $F_{(3, 435)} = 3.908, p = .009$, partial $\eta^2 = .03$. Post-hoc comparisons further show that *senior* ($M = 11.65, SD = 3.21$) students scored significantly higher than *junior* ($M = 10.56, SD = 3.19$) students ($p = .009$).

Table 4. Class Status Means and Standard Deviations on Total Critical Thinking Scores

Class Status	<i>N</i>	<i>M</i>	<i>SD</i>
Freshmen	29	11.83	3.68
Sophomores	73	10.92	3.50
Juniors	162	10.56	2.91
Seniors	175	11.65	3.19

College

A one-way analysis of variance (ANOVA) was conducted to compare mean critical thinking scores (0-20) across eight colleges (College of Agriculture, College of Behavioral and Social Sciences, College of Business, College of Communication and Education, College of Engineering Computer Science and Construction Management, College of Humanities and Fine Arts, College of Natural Sciences, and Undergraduate Education). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p = .867$). Mean critical thinking scores were not significantly different across colleges groups, $F_{(7, 431)} = 1.856, p = .075$, partial $\eta^2 = .03$. These results do not indicate a significant difference in mean critical thinking scores between eight colleges.

Table 5. College Means and Standard Deviations on Total Critical Thinking Scores

College	<i>N</i>	<i>M</i>	<i>SD</i>
Agriculture	15	10.60	3.11
Behavioral & Social Science	154	10.79	3.17
Business	64	11.48	3.09
Communication & Education	55	11.14	3.34
Engineering Computer Science & Construction Management	42	12.00	3.01
Humanities & Fine Arts	44	10.84	3.00
Natural Sciences	54	10.94	3.55
Undergraduate Education	11	13.54	2.77

Ethnicity

A one-way analysis of variance (ANOVA) was conducted to compare mean critical thinking scores (0-20) across ethnic groups (American Indian/Alaskan Native, Asian, Black/African American, Decline to State, Hispanic/Latino, Nonresident Aliens, Two or More Races/Ethnicities, White). There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p=.801$). The results show that critical thinking score means are significantly different across ethnic groups, $F_{(7,431)} = 5.784, p < .000$, partial $\eta^2 = .09$. Post-hoc comparisons further show that White ($M = 11.78, SD = 3.01, p < .000$), Two or more Races/Ethnicities ($M = 11.12, SD = 3.29, p = .021$), Hispanic/Latino ($M = 10.50, SD = 3.22, p = .017$), and Decline to State ($M = 11.63, SD = 3.22, p < .000$) students scored significantly higher than Asian ($M = 8.31, SD = 2.83$) students. Post-hoc comparisons further show that White ($M = 11.78, SD = 3.01$) students scored significantly higher ($p = .009$) than Hispanic/Latino students ($M = 10.50, SD = 3.22$).

Table 6. Ethnic Group Means and Standard Deviations on Total Critical Thinking Scores

Ethnic Group	<i>N</i>	<i>M</i>	<i>SD</i>
American Indian/Alaskan Native	5	12.00	1.87
Asian	29	8.31	2.83
Black/African American	8	11.25	2.49
Decline to State	38	11.63	3.22
Hispanic/Latino	114	10.50	3.22
Nonresident Aliens	6	10.00	4.47
Two or more Ethnicities/Races	25	11.12	3.29
White	214	11.78	3.01

Conclusions

General

- Assessment of Critical Thinking remains elusive due to a lack of consensus on its meaning and the difficulty of measuring a complex cognitive phenomenon like CT with available methods.
- Faculty are generally not enthusiastic participants in assessment and seemed particularly reluctant to engage in assessment of CT.
- Multiple methods present clear advantages over any one approach to CT given the measurement challenges.
- Assessing CT using “signature assignments” requires a more explicit commitment to linking pedagogy and assessment than is currently the case on our campus.
 - Faculty assignments are apparently designed to fulfill a number of instructor objectives.
 - Signature assignments require explicit linkage of the design of assignments to particular outcomes being assessed, in this case CT.

- The use of a multiple choice “quiz” for assessing CT strikes many as inadequate as the questions lack context, are not linked explicitly to course material and have to be answered without the opportunity for much reflection or any dialog with others.
 - The use of a multiple choice quiz does, however, isolate quality of thinking from quality of writing.
 - Assessing the validity of the CT Quiz is challenging.
 - The lack of nationally normed results vitiates the interpretation of results in the current instance.

The two different assessments involved somewhat different interpretations of “critical thinking”; the writing assignment focused on students’ ability to make strong written arguments and draw conclusions based on evidence, while recognizing other conditions and issues that may impact their conclusions. The multiple-choice quiz was more directly focused on using proper logic and reasoning to form sound conclusions. While these two styles of measuring “critical thinking” aren’t necessarily mutually exclusive, there should be no surprise if the different data sets led to different results.

Based on the Writing Assignments

- The ability of our upper class students (juniors/seniors) to correctly use their critical thinking skills to make written arguments is wide-ranging but, on average, adequate (based on their mean scores). However, that ability does not significantly improve as the students move through their studies.
- Interestingly, students who took one or more remedial courses in college had a higher average score on their writing assignments. This difference was statistically significant at the 95% confidence level. It could be explained by students in remedial courses (if it’s remedial ENGL) having done more writing in small class settings. (If they took remedial MATH and not ENGL that explanation doesn’t make sense)
- No significant differences in students’ abilities to use their critical thinking skills to make written arguments were found based on students’ gender, transfer status, college of study, or ethnicity.

Based on the Multiple Choice Critical Thinking Quiz

- There was a significant drop in the average scores in students’ results in the Spring 2016 quiz (55.6%) vs. the combined results of students in the Fall 2012/13 quizzes (66.1%). *Note: the Spring 2016 quiz was 20 questions, comprised of 9 of the questions from the Fall 2012 quiz and 11 of the questions from the Fall 2013 quiz.* The Fall 2012 and 2013 quizzes were administered to students near the end of their A3 course (primarily PHIL 102), where much direct attention is given to analyzing the types of arguments present in the CT quiz. The enrollment in PHIL 102 consists of 75% freshman and 20% sophomores (data from Sections 1, 2, and 3 of PHIL 102 in F13). The Spring 2016 quiz was given to students in upper-division GE courses, which primarily (77%) consisted of juniors and seniors. Most of these upper-division students would have been several years removed from PHIL 102; it stands to reason that their ability to answer questions based on systematic use of logic and reasoning had faded. Perhaps this indicates a need to

reinforce the particulars of logic and reasoning in courses beyond A3, if the campus deems developing that ability to be worthwhile.

- The quiz results also had the highest score among the freshman. While the number of freshman tested was small (6.6% of the sample), there was a statistically significant difference in their score compared to the juniors (at the 95% CL). This again reinforces the conclusion that, having more recently completed PHIL 102, they are better at answering questions involving analyzing logical and illogical reasoning than those further removed from that class. Scores were lower for sophomores and lower still for juniors, but improved significantly in seniors. This may indicate that some further instruction in logic and reasoning is taking place in senior classes or that the ability to properly use logic and reasoning is being indirectly reinforced in those classes.
- Based on the data in Table 6, scores for Asian and Hispanic/Latino students were significant lower than those for other ethnic groups. Because this quiz involved interpretation of nuanced language which must be read very precisely, the lower scores could be explained in part by some of the Asian and Hispanic/Latino students not having English as their primary language.
- There were no differences in scores based on transfer status or college of study.
- There was a statistically significant difference in scores based on these categories:
 - Male students scored 4% higher than female students.
 - Non-first-generation students scored 6% higher than first-generation students.
 - Students who did not take remedial courses scored 11% higher than students who did take remedial courses.

The results of the assessments carried out in AY 15-16 point to the need for more explicit consideration of critical thinking in the curriculum and pedagogical approaches designed specifically to strengthen and reinforce critical thinking and elicit its practice in a variety of assignments across the curriculum throughout students' careers.

A further recommendation based on the results of this assessment was to share with faculty teaching GE courses the list of their courses' SLO's, the definitions of those SLO's, and, where available, rubrics for them to consider with designing assignments meant to assess their students' competencies in those SLO's.