

2019-2020 Institutional Assessment of Critical Thinking: Full Report

Report submitted by Jim Taylor, Professor of Philosophy and Lead Assessment Specialist

Submission date: June 9, 2021

I. Overview, Team, Instrument, Data Collection, & Limitations

Overview. Many educators believe critical thinking skills are the most important higher education outcome. Westmont's commitment to critical thinking is represented by our institutional learning outcome that *Westmont graduates will accurately evaluate the strength of evidence in support of a claim*. Our institutional assessment efforts in the 2019-2020 academic year were devoted to investigating how well our students are doing in this area.

Team. The members of the assessment team that agreed to engage in this investigation were Jim Taylor (Philosophy), Lead Assessment Specialist; Steve Contakes (Chemistry), *Assessment Consultant in Natural & Behavioral Sciences*; Angela D'Amour (Student Life), *Assessment Consultant in Student Life*; Edd Noell (Economics & Business), *Assessment Consultant in the Social Sciences*; and Randy VanderMey (English), *Assessment Consultant in the Humanities*. Tatiana Nazarenko (Administration), *Dean of Curriculum & Educational Effectiveness* oversaw, organized, and supported our efforts throughout. All of us, except for Edd, were on the critical thinking ILO assessment team I lead during the 2013-2014 academic year. Edd participated that year in the scoring session (as did the rest of us).

Instrument. We used the same assessment instrument we employed last time: the 15-question, short answer Critical Thinking Assessment Test (the CAT), developed and facilitated by the Center for Assessment & Improvement of Learning at Tennessee Tech University (the CAIL). The test takes students about an hour to complete and requires them to engage in relatively high-level real world critical thinking and problem solving. It also requires them to be able to exercise creativity and to communicate their answers clearly, concretely, and concisely. The skills assessed by means of this test include but go beyond the skill set required to meet our institutional critical thinking student learning outcome (to accurately evaluate the strength of evidence in support of a claim).

We chose the CAT because it focuses on a number of identifiable higher-order critical thinking skills of the sort contained in Bloom's classic taxonomy of cognitive skills, involves thinking about real world problems, requires short essays as answers to most questions (thus revealing students' underlying thought processes), is (ideally) scored in a guided scoring session by Westmont faculty (thus providing us with professional development benefits), has been widely used for over 20 years (by over 350 institutions of higher learning on their campuses and in over 40 NSF projects), is valid and reliable, provides a basis for comparison to national norms, and is relatively inexpensive.

[The CAIL website](#) describes the development, purpose, nature, and value of the CAT as follows: "The Critical-thinking Assessment Test (CAT) was developed with input from

faculty across a wide range of institutions and disciplines, with guidance from colleagues in the cognitive/learning sciences and assessment and with support from the National Science Foundation (NSF). DEVELOPED to assess a broad range of skills that faculty across the country feel are important components of critical thinking and real world problem solving. DESIGNED to emulate real world problems. All questions derived from real world situations with most questions requiring short answer essay responses. ENGAGE faculty in the assessment and improvement of student critical thinking skills and connects faculty to a teaching community.”¹

The specific critical thinking skills covered by the CAT are as follows:²

- Q1 – Summarize the pattern of results in a graph without making inappropriate inferences.
- Q2 – Evaluate how strongly correlational-type data supports a hypothesis.
- Q3 – Provide alternative explanations for a pattern of results that has many possible causes.
- Q4 – Identify additional information needed to evaluate a hypothesis.
- Q5 – Evaluate whether spurious information strongly supports a hypothesis.
- Q6 – Provide alternative explanations for spurious associations.
- Q7 – Identify additional information needed to evaluate a hypothesis.
- Q8 – Determine whether an invited inference is supported by specific information.
- Q9 – Provide relevant alternative interpretations for a specific set of results.
- Q10 – Separate relevant from irrelevant information when solving a real-world problem.
- Q11 – Use and apply relevant information to evaluate a problem.
- Q12 – Use basic mathematical skills to help solve a real-world problem.
- Q13 – Identify suitable solutions for a real-world problem using relevant information.
- Q14 – Identify and explain the best solution for a real-world problem using relevant information.
- Q15 – Explain how changes in a real-world problem situation might affect the solution.

Data Collection. The Westmont CAT results are based on 135 students who identified as seniors and 6 students who identified as juniors in spring 2020 senior seminar classes.³ These classes were from the following departments: Computer Science, Economics & Business, English, Kinesiology, Philosophy, Psychology, Religious Studies, and Sociology. As a result, we were able to test students from eight different departments with at least two departments from each of our three academic divisions (Humanities, Natural & Behavioral Sciences, and Social Sciences).⁴

¹ <https://www.tntech.edu/cat/index.php>

² Questions 5, 8, 10, 12, 13 are objectively scored questions. Scoring students’ answers to the other questions requires subjective judgment, since their answers are given in short-answer or essay form.

³ Only 124 of these students show up in the list of members of the class of 2020. Of course, 6 of these are the students who identified as juniors. Presumably the other 11 had senior status when they took the CAT even though they didn’t graduate in May 2020.

⁴ All the tests were administered on campus in senior seminar classes before the transition to remote instruction due to the COVID-19 pandemic—except for the tests given to the sociology students. The sociology students were given an opportunity to take the test online. Only one sociology student appears in the list of students who took the test. There is also a student who took the test who is classified as a communication studies major. These two students’ results are included in the “All Students” and “Transfer and Non-Transfer Students” reports but not in the “Division” and “Department” reports. Also, there is a biology major report even though the biology senior seminar did not participate and even though only two biology majors took the test.

Unfortunately, our faculty and library staff were not able to score the tests as planned because of COVID-19 physical distancing restrictions. So, Tatiana hired some student workers to enter each of the students' answers into an online form provided by the CAIL. Then the CAIL scored each of the tests by an automated process involving machine learning.

Limitations. This study is subject to the usual limitations of an assessment administered to a proper subset of a population only once. Appropriate caution must be taken when drawing conclusions about the entire population on the basis of the performance of the sample.

There is good reason to think the sample of 141 students whose CATs we scored is relatively *random*. Since nearly all majors have a senior seminar or capstone course of some kind, and since the test was administered only to students in senior seminar or capstone courses, almost any graduating Westmont senior could have wound up being part of the sample.

Moreover, since there were three departments represented from the NBS division (Biology, Mathematics/Computer Science/Data Analytics, Kinesiology), three from the HUM division (English, Philosophy, Religious Studies), and two from the SS division (Economics & Business and Sociology⁵), the students who participated came from a wide range of disciplines. However, the distribution of students across the three divisions was not equal in number: HUM (n=32), NBS (n=67), and SS (n=40).⁶

There is also good reason for thinking the sample of 141 is *representative* of the class of 2020. Tim Loomer, Director of Research, Planning, and Implementation, performed some statistical tests (with data provided by Records System Specialist Anna Darby) and concluded that, based on analysis of SAT, ACT, and GPA data, the members of the class of 2020 (n=124) who took the CAT do not differ significantly from the entire class (n=305).^{7,8}

⁵ But see footnote 4 about the Sociology department.

⁶ According to Tim Loomer (see the next paragraph in the report), the percentage of students in the graduating class of 2020 who graduated with a major in each division is as follows: Humanities: 26.2%; Natural and Behavioral Sciences: 40.8%; Social Science: 33.0%. Students with two or three majors could be in multiple categories in this data (there were 336 majors for 309 graduating students). From the CAT the numbers are Humanities: 32 students (23.0%); Natural and Behavioral Sciences: 67 students (48.2%); and Social Science: 40 students (28.8%). So, students from the Natural and Behavioral Sciences were a bit over-represented and Social Science a bit under-represented in the CAT group but there were certainly a good number of students from each division represented in the CAT and the percentages are reasonably similar to that of the graduating class of 2020.

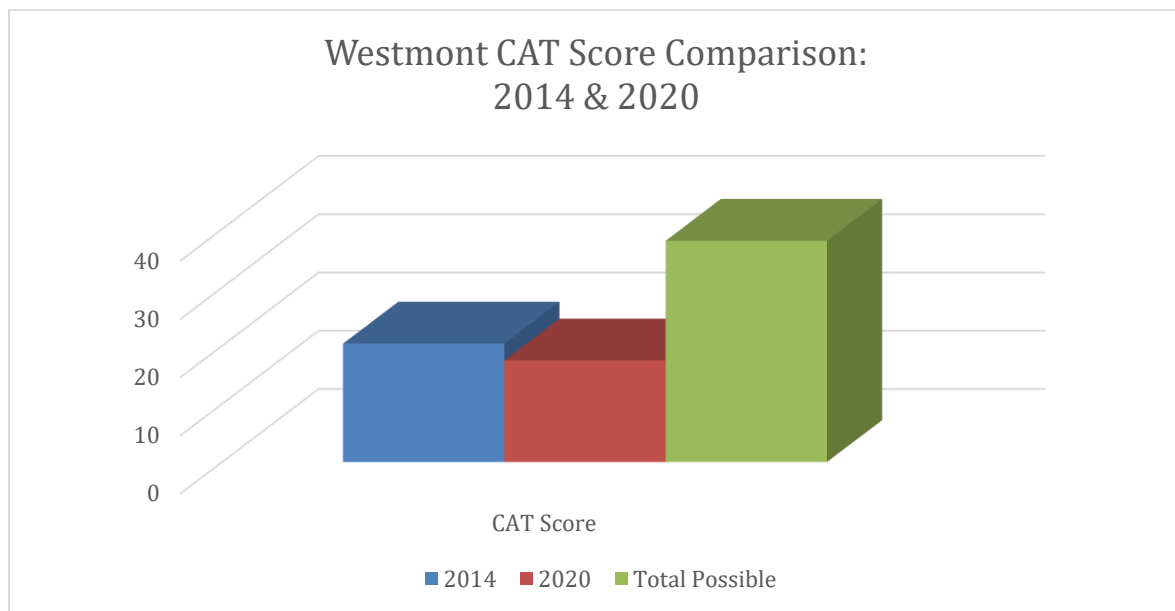
⁷ The class of 2020 CAT sample has a mean GPA of 3.284 and the class of 2020 as a whole has a mean GPA of 3.353. The class of 2020 CAT sample has a mean SAT score of 1199 and the class of 2020 as a whole has a mean SAT score of 1185. Finally, the class of 2020 CAT sample has a mean ACT score of 26.76 and the class of 2020 as a whole has a mean ACT score of 26.88.

⁸ Tim also found that cumulative Westmont GPA is a significant but not strong predictor of CAT scores. This finding could indicate that the ability to think critically in the ways assessed by the CAT are only a small component of the learning reflected in course grades (roughly 7%). If so, it will be interesting to see going forward whether this percentage increases in the case of students who take courses taught by instructors who incorporate assignments requiring CAT skills into

In sum, the students who took the CAT in 2020 constitute a random and representative sample of the entire class of 2020. Consequently, while keeping the limitations mentioned above in mind, it is reasonable to draw conclusions about the whole class on the basis of the performance of the members of that class who took the CAT.

II. What We Learned

The overall 2020 Westmont average total CAT score is 17.48 (out of a possible 38). This score is lower than the overall 2014 Westmont average total score, which was 20.37. So, our students, on average, scored lower than the 2014 group. But a comparison of the two scores should take into account that (1) the number of students who took the test in 2020 (141) is almost exactly double the number who took it in 2014 (71); (2) the standard deviation was 5.33 in 2014 and 4.77 in 2020, so the difference in overall score may be less significant than it appears to be;⁹ and (3) in 2020, students from five additional majors (Biology, Data Analytics, English, Kinesiology, and Mathematics) participated along with students from the six departments whose seniors took the CAT in 2014 (Economics & Business, Computer Science, Philosophy, Psychology, Religious Studies, and Sociology—although only one student represented Sociology in 2020). So, there were not only substantially more students who took the test in 2020 but also students from a greater diversity of disciplines.



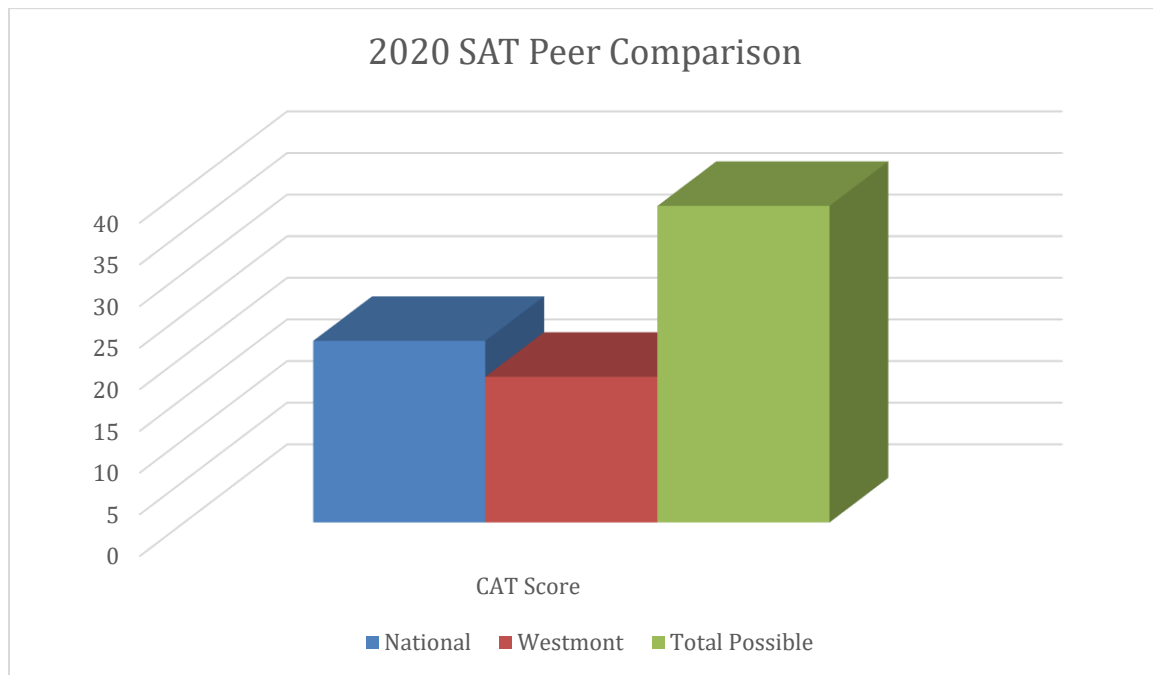
In spite of the 2020 Westmont overall average total being less than the 2014 Westmont overall average total, Westmont students did better collectively on questions 1, 10, 13, & 14 in 2020 than in 2014 (but they did collectively worse on all the other questions—though some of these differences, both positive and negative, may not be statistically significant).

the course and who base the final grade to some extent on the students' performance on those assignments. But the more important indicator will be our students' performance relative to the average national CAT score of students with the same SAT and/or ACT scores.

⁹ Another difference between the overall results of the two years concerns the range of scores. In 2020 scores ranged from 4 to 28 while in 2014 scores ranged from 8 to 36.

Another comparison between the 2020 and 2014 results concerns how our students did relative to the national CAT score average. In 2014, Westmont’s mean score (20.37) was greater than the national average (19.04). However, when the Westmont 2014 CAT score average is compared to the average national CAT score achieved by upper-division students at 4-year undergraduate institutions with the same average SAT score (1199), the Westmont 2014 CAT score (20.37) was only 93.4% of the national SAT-peer CAT score (21.81).

In 2020, Westmont’s overall mean score (17.48) was slightly lower than (99% of) the national average (17.64). But when compared to their national SAT peers (average 1199¹⁰—as in 2014), the Westmont average CAT score (17.48) falls far shorter than the national average CAT score (21.79). The Westmont score is only 80% of the national SAT-peer score. So, our students did not perform as well overall in 2020 as they did in 2014—relative to the national average score achieved by students with the same entering SAT scores. A comparison focusing on ACT scores yields almost exactly the same result.^{11,12} The Critical Thinking Assessment Team considers this outcome a reason to be concerned about the effectiveness of Westmont’s current teaching of critical thinking skills.



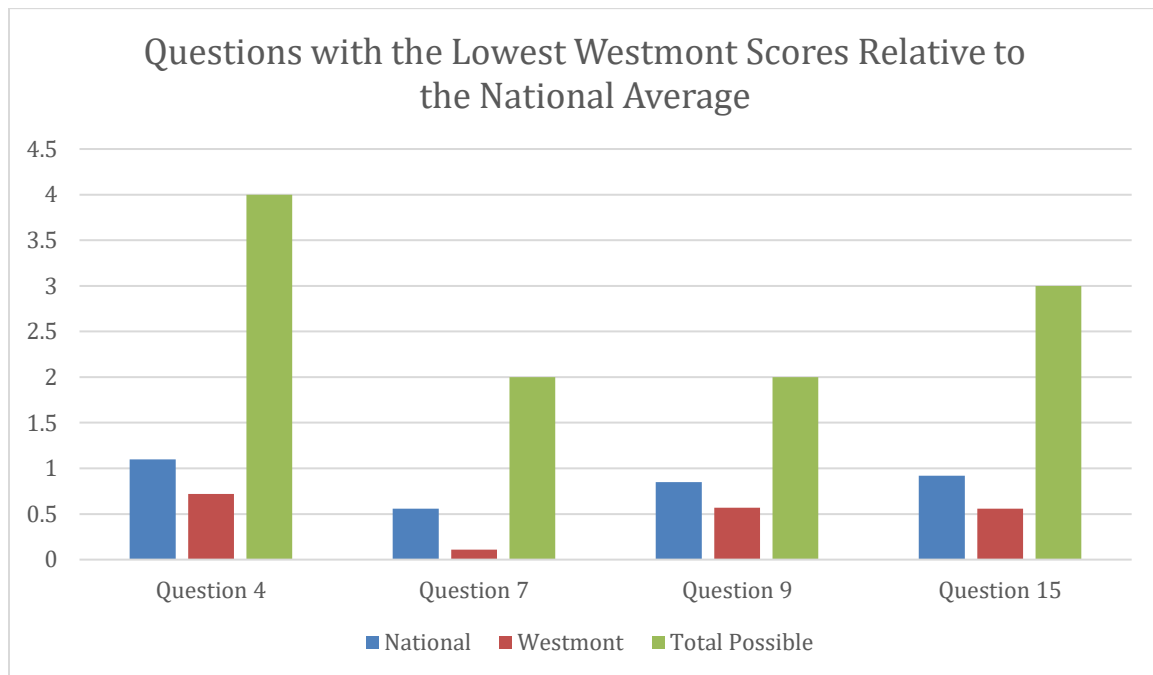
¹⁰ The number of students in the class of 2020 who had SAT scores and took the CAT is 87.

¹¹ The mean ACT score of 2020 Westmont students who took that test and the CAT (n=54) is 26.76. Nationally, students with that ACT score had an average CAT score of 22.

¹² A CAIL representative said they adjust their national norms every five years to reflect changes in their student user population. After their last adjustment in December 2019, they saw a decrease in the mean score in the four-year college senior-level norm category. They believe most of this shift is due to a large increase in students in their data set which increases the variability and diversity of the student data. The national average CAT score has been adjusted from 19.04 (in 2014) to 17.64 (in 2020)—to reflect a larger and more diverse pool of students.

But there were questions on which Westmont students did significantly better than the national average in each of those years. In 2014, our students did better than the national average on questions 1, 2, 5, 6, 8. And in 2020, our students did better than the national average on questions 1, 5, 6, 8, 13, 14 (notice the overlap of questions 1, 5, 6, and 8).

But our 2020 students did significantly worse than the national average on four questions (4, 7, 9, & 15) and our 2014 students did significantly worse than the national average on only one question (7). The questions on which our 2020 students did significantly worse than the national average (4, 7, 9, & 15) all involve problem solving, creative thinking, and effective communication (in the case of 4, 7, & 15) and creative thinking and effective communication (in the case of 9). None of them involve the general skills of evaluating and interpreting information (which eight of the other questions required). Two of these (4 & 7) involve the same specific skill (“additional information needed to evaluate a hypothesis”), one involves the specific skill of “providing relevant alternative interpretations for a specific set of results” (9), and the fourth (15) the specific skill of explaining how changes in a real-world problem situation might affect the solution of a problem. Our students did worst on 7 relative to the national average (-.89) and roughly the same on the other three (-.41, -.44, & -.41).¹³



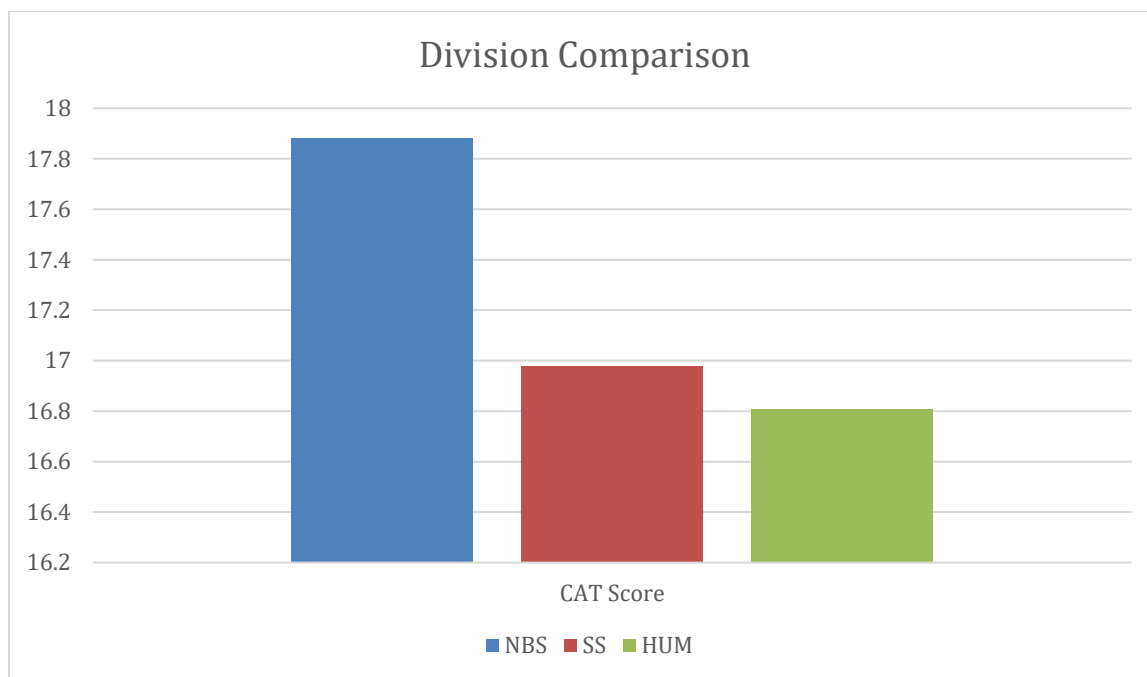
What each of the skills with which our students are deficient seem to have in common is *a creative use of the intellect and/or imagination* to think of evaluations, interpretations, or solutions that go beyond what is explicitly (and perhaps even implicitly) given in the test. The skill involved in answering questions 4 and 7 well requires *additional* information, the skill involved in answering question 9 well requires relevant *alternative* interpretations, and the skill involved in answering question 15 well requires explaining how *changes* in a real-world problem situation might affect a solution.

¹³ The negative numbers in parentheses are the “effect size” (mean difference between the Westmont and national average score divided by pooled group standard deviation).

The members of the Critical Thinking Assessment Team suggested a few possible explanations of our students' collective relatively poor performance on these questions that require similar skills: Our students go after relatively simple answers rather than pushing further to engage in deeper and more complex thinking; when they settle on one evaluation, interpretation, or solution, they do not (or do not consider) looking for alternatives; their thinking is so narrow or conventional that they do not easily think "outside the box"; they are not motivated to think more widely and independently; they come from backgrounds emphasizing accepting what they are taught over exploring multiple viewpoints; they are so burdened by anxiety or stress that they are not free to engage in challenging higher thinking; they are more disposed to please others than they are to think in ways that go against the prevailing accepted ways of thinking; and they have a need for security that prevents them from going beyond black and white thinking.

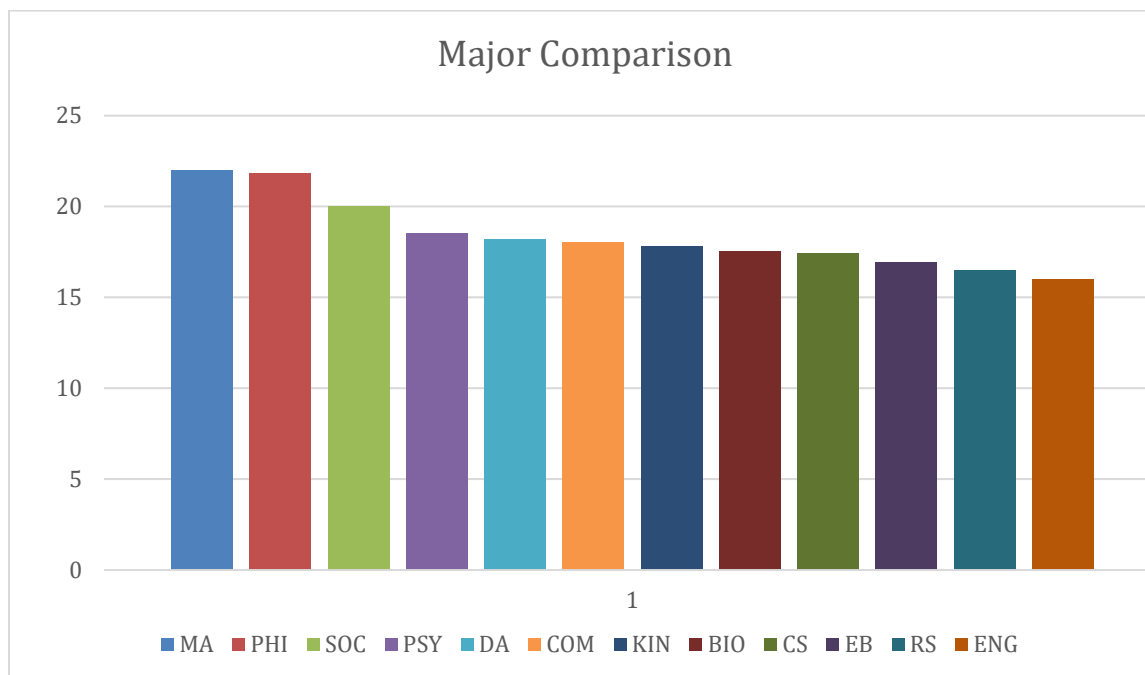
Some other interesting observations about the 2020 results have to do with various sub-groups of students. We have results by academic division of the senior seminar class in which the students took the test as well as results by student-reported major department, gender, racial identity, non-resident status, (non-)transfer status, and whether the student is a first-generation student.

- The 67 NBS students did the best overall (17.88) followed by the 40 SS students (16.98) and the 32 HUM students (16.81).



- By self-reported major department, the average total scores are as follows (from highest to lowest):
 - o Mathematics (n=1)—22
 - o Philosophy (n=4)—21.8

- Sociology (n=1)—20
- Psychology (n=24)—18.5¹⁴
- Data Analytics (n=6)—18.2
- Communication Studies (n=1)—18
- Kinesiology (n=26)—17.8
- Biology (n=2)—17.5
- Computer Science (n=10)—17.4
- Economics & Business (n=39)—16.9
- Religious Studies (n=14)—16.5
- English (n=13)—16

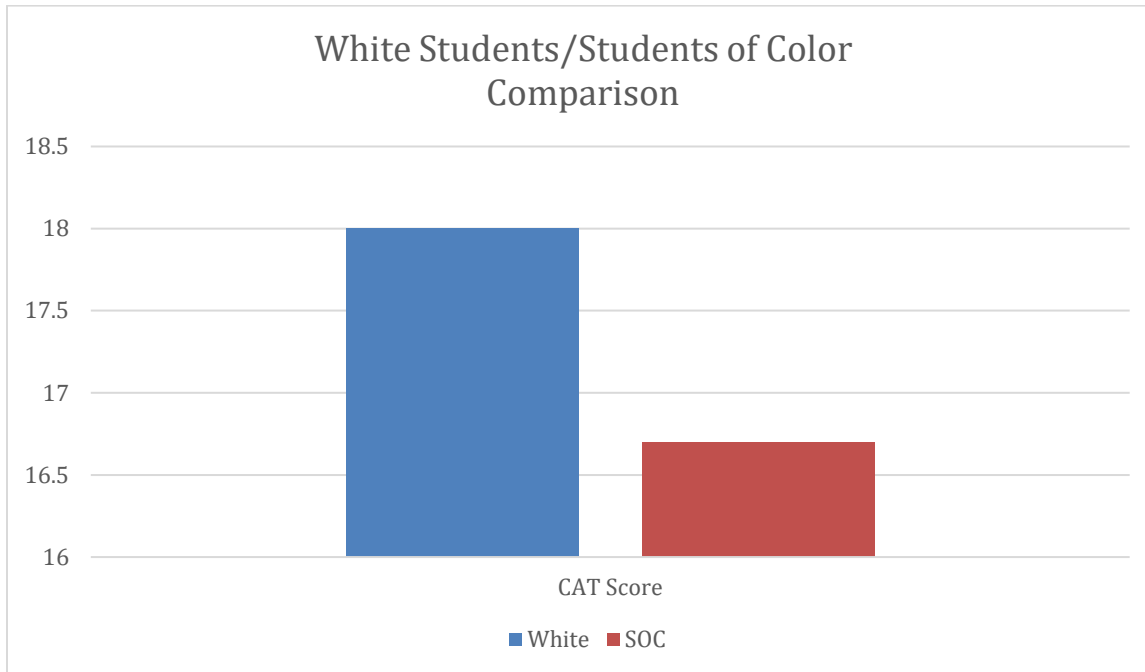


- As for gender, the average total scores are nearly the same:
 - Female (n=78)—17.41
 - Male (n=62¹⁵)—17.5
- Average scores by racial categories selected by students from a list given to them to choose from are as follows (from highest to lowest):¹⁶
 - White (n=83)—18
 - Latinx (n=22)—17.6
 - Students of Color (n=46)—16.7
 - Asian (n=15)—14.5

¹⁴ Of the five departments that participated in both 2014 and 2020 with more than one student (CS, EB, PHI, PSY, & RS), only Psychology has a higher overall average score in 2020.

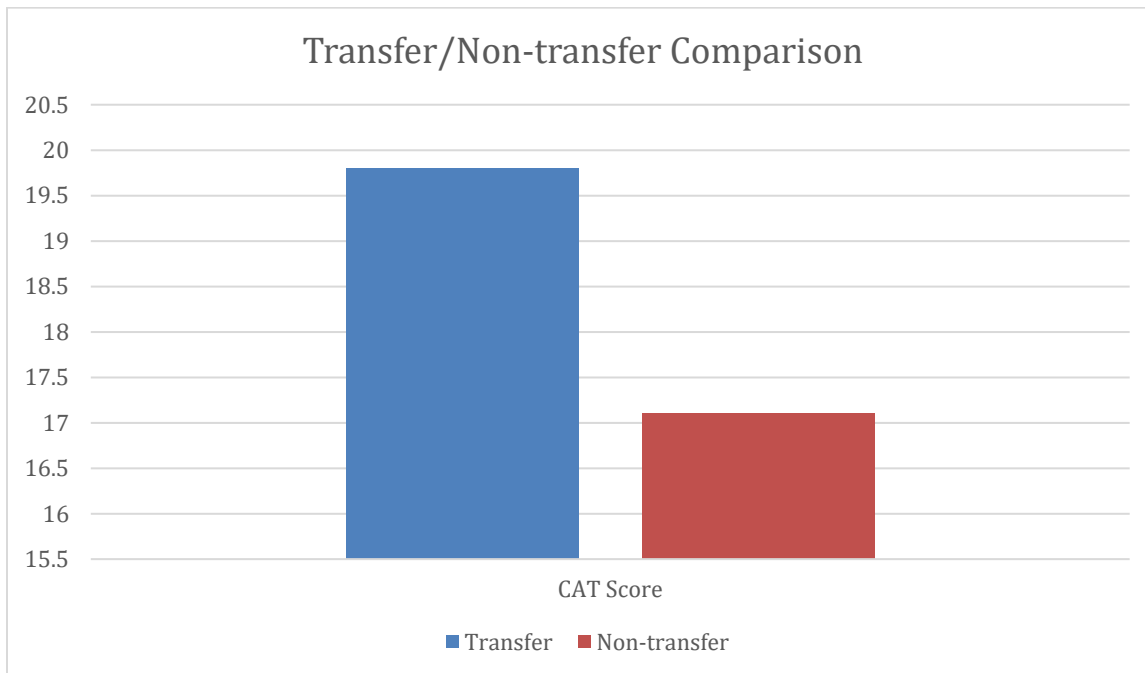
¹⁵ The total number of female and male students adds up to 140 rather than 141 because one student (with a traditionally male first name) chose not to specify gender.

¹⁶ The total number adds up to more than 141, because some students selected more than one racial category. It's not clear why the sum of white students and students of color is 129.

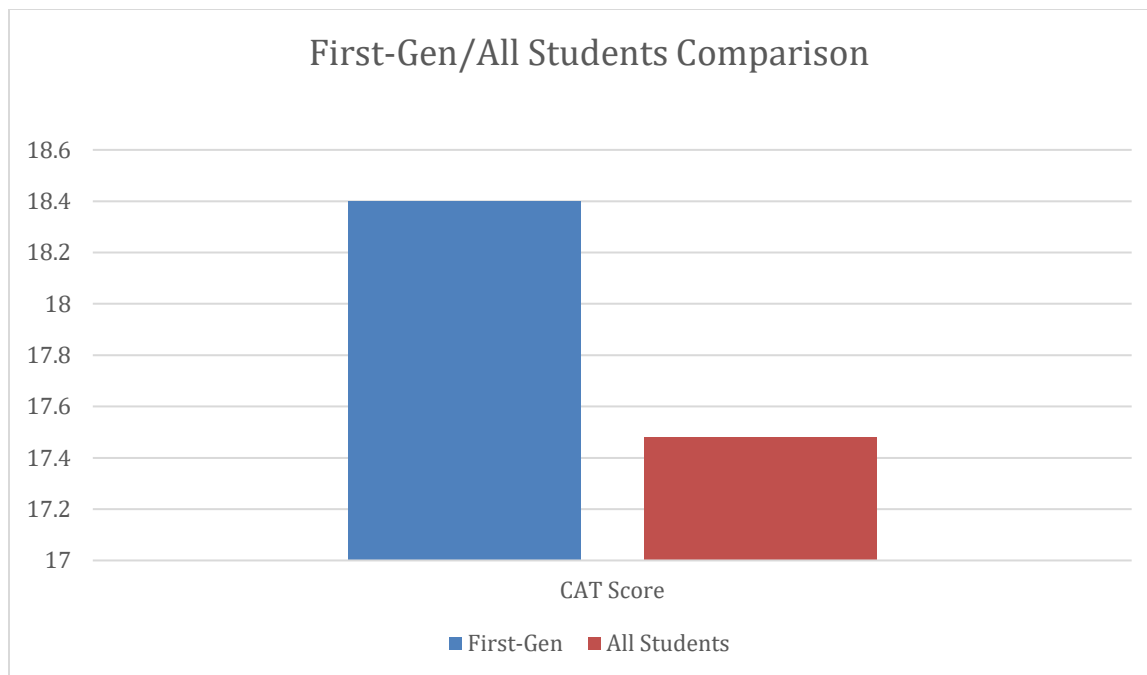


- Average score achieved by students who are not citizens of the United States:
 - o Non-resident (n=4)—17

- A comparison of transfer and non-transfer student scores:
 - o Transfer students (n=20)—19.8
 - o Non-transfer students (n=121)—17.1



- Average score achieved by students whose parents did not complete a four-year college or university degree:
 - o First-generation students (n=21)—18.4



The assessment team discussed possible reasons both transfer students and first-generation students did so much better on average than the overall Westmont average (17.48) and better than the average achieved by most of the other sub-groups aside from specific majors (with students identifying as white coming in third). Some suggested that students in these two groups are more likely than the others to engage in “out of the box” thinking and “own their education more” due to having to be more proactive and adult in their pursuit of an education. That is, these students’ need to take more responsibility for their learning may have motivated them to do more independent thinking. Another suggestion is that these groups have more resolve and persistence than students in the other groups as they have navigated through various additional trials in making it to their senior year.

III. Recommendations

The Critical Thinking Assessment Team recommends the Westmont faculty discuss possible pedagogical and curricular changes that will facilitate our students’ improvement in the three skill areas involved in the four CAT questions on which our students did significantly worse than the national average:

- Identify additional information needed to evaluate a hypothesis. (Q4 & Q7)
- Provide relevant alternative interpretations for a specific set of results. (Q9)
- Explain how changes in a real-world problem situation might affect the solution. (Q15)

There are at least three alternative general approaches we could take: (1) Add a Critical Thinking requirement to our GE program, (2) simply encourage all faculty to focus more on

critical thinking in their relevant courses, or (3) implement a targeted but less comprehensive approach. Option 1 seems unreasonable, since our GE program already takes up a lot (some think too much) of our students' units. And option 2 is unlikely to lead to much significant positive change. But option 3 seems feasible. It's the alternative the assessment team recommended in 2014. But it ended up having the effect of option 2, since we didn't follow through with our specific plan, which I will now restate.

The CAIL recommended that, now that we have an idea where our students are as they are exiting the college, we identify some courses in which to focus on the skills we've chosen to target and use the CAT to do pre-testing and post-testing in those courses to see whether our teaching of those skills is effective. Then we invite faculty members who teach these courses to consider which of the skills we target they would like to provide instruction for in one of their classes (again, using pre- and post-tests). According to the CAIL, it would not be necessary or helpful to test incoming freshmen to get a benchmark to compare with our senior class results. Instead, it would be better for us to focus on using the CAT at the individual course level (using a pre-test and post-test design) going forward. Then we could compare not only individual student's results on these two tests but also use as a benchmark the national average score of students who have the same SAT and/or ACT scores (if these scores are available).

In light of their recommendations, the Critical Thinking Assessment Team recommends that the Academic Senate:

- (1) Identify courses in which focused instruction could be given for the improvement of the three skills listed above (the Team recommends at least one GE course from each of the three divisions such as PHI-012, CHM-005, and SOC-001—but ideally more);
- (2) Secure faculty members willing to teach those skills in those courses;
- (3) Provide those faculty members with the financial and pedagogical resources they need to implement this skill instruction effectively;
- (4) Arrange to use CATs for pre-tests and post-tests in these courses;¹⁷
- (5) Set as a benchmark the average overall national CAT score of SAT-peers at comparable institutions; and
- (6) Consider changing the ILO to include a broader range of critical thinking skills.

Also, in response to a CUPA Team suggestion, the CT Team also recommends the Senate

- (7) Change the language of the Critical Thinking ILO. The new ILO reads, "*Westmont graduates will demonstrate sound judgment and creative thinking when evaluating the strength of evidence in support of a claim.*"

Another of the CAIL's recommendations for individual departments is to (1) select the skills tested by the CAT that are especially important in their discipline and then (2) develop discipline-specific analogs to the CAT questions that test students on these skills ("CAT Applications" or "CAT Apps"). Though students' responses to these discipline-specific analogue questions will not be assessable by means of the standard CAT scoring process, departments can formulate their own rubrics as tools to evaluate the tests they construct out

¹⁷ The documents entitled "Effective Practices for Improving Students' Critical Thinking and Problem Solving" and "Effectively Using the CAT for Assessment" in the CAT Training Manual will be especially useful as resources for the faculty who agree to teach a class that targets specific CAT critical thinking skills.

of their discipline-specific questions. We recommend offering a workshop for the faculty who volunteer to incorporate CAT-related assignments in their courses (and for other interested faculty as well).

See the separate appendices containing the following documents:

- 1: CAT Institutional Reports from the CAIL
- 2: Effective Practices for Improving Students' Critical Thinking and Problem Solving
- 3: Effectively Using the CAT for Assessment