2024-2025 Assessment of Critical Thinking Institutional Learning Objective
Final Report (DRAFT)

Assessment Team: Robert Haring-Kaye (Professor of Physics and Lead Assessment Specialist), Blake Kent (Associate Professor of Sociology), Ruth Lin (Music Department Chair and Director of Orchestral Activities), Tatiana Nazarenko (Dean of Curriculum and Educational Effectiveness), and Jim Taylor (Professor of Philosophy) with the assistance of Laura Nichols (Data Analyst and AMS Coordinator)

July 29, 2025

I. Introduction

Westmont College has eight institutional learning outcomes (ILOs) [1], each of which is assessed sequentially in successive academic years. This report summarizes the results, interpretations, and recommendations following the assessment of the Critical Thinking (CT) ILO (*Graduates of Westmont College will be able to accurately evaluate the strength of evidence in support of a claim and apply critical thinking creatively to problem solving*) during the 2024-2025 academic year. Previous assessments of this ILO were performed in 2019–2020 and 2013–2014, though the 2019–2020 assessment was abbreviated due to COVID. Comparisons between the outcomes of these two prior CT ILO assessments and the current one will be discussed in later sections.

The authors of this report, representing the CT ILO Assessment Team, were first assembled during the Spring 2024 semester for the purpose of becoming familiar with the CT assessment and choosing the assessment instrument(s) and analysis methods. Team members were chosen with sensitivity to experience and expertise that would be especially helpful for this assessment work while also providing equal representation among the three major academic divisions. In particular, Jim Taylor was the Lead Assessment Specialist for both the 2013–2014 and 2019–2020 CT ILO assessments, which provided helpful continuity across these three most recent assessments.

After meeting in April 2024, the CT ILO assessment team decided to utilize the Critical-thinking Assessment Test (CAT) [2] as the assessment instrument. This is the same fifteen-question standardized test that was implemented during the 2013–2014 and 2019–2020 CT ILO assessments, which has remained identical in content throughout [3]. The decision to use the CAT again for the 2024–2025 assessment was motivated by the ability to perform a longitudinal assessment, the availability of national norms for comparison, the rigorous statistical analysis (aggregated as desired) provided by the test publisher, and the assessment of student responses using artificial intelligence to provide consistent evaluations across questions and test takers.

Although the individual CAT questions are not made public due to their continual recycling, the critical thinking skills assessed by each question are well known and available publicly. They are:

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.

An additional facet of the 2024–2025 CT assessment cycle that was not implemented in the two previous rounds was the administration of the CAT to a random group of first-year students during the Fall 2024 semester as well as graduating seniors during (primarily) the Spring 2025 semester as an approximate way to gauge CT skill development before and after a Westmont education. Although any comparisons between these two groups must be interpreted with caution since they represent different student populations, we believe the cross-sectional "snapshot" provided by these data yields valuable information about where the strengths and weaknesses of our incoming students lie and how our curriculum can both reinforce the strengths and enhance the areas that we think require growth.

II. Data Collection

The CAT was administered to a total of 399 students during the 2024–2025 academic year, 170 (229) during the Fall 2024 (Spring 2025) semester. Of these, 124 self-identified as first-year students and 212 as seniors. The former represents 41% of incoming first-year students, not including transfer students, at the beginning of the Fall 2024 semester. The latter constitutes 73% of students who graduated in Spring and Mayterm of 2025. Since we are primarily interested in comparing the results of the first-year students with those of the seniors, we disaggregated those students from other test-takers when performing the score analyses that follow (see Sec. III).

We attempted to obtain a random sample of students for both data sets (primarily first-year students in Fall 2024 and seniors in Spring 2025) but were limited by course availability and instructor willingness to participate in the study. In preparation for the Fall 2024 CAT administration, we invited several instructors who teach courses with historically large fractional enrollments of first-year students in the fall to participate. This resulted in the assessment of four unique courses in four different departments spanning two academic divisions. (Humanities was not represented in this sample.) Similarly, the instructors of every senior seminar course offered in the Spring 2025 semester (or an alternative course designated by the associated department chair) were invited to participate in the Spring 2025 CAT administration. Of these, fourteen such courses representing twelve departments and all three academic divisions were represented in this data set (including one senior seminar course that was offered during the Fall 2024 semester). Before taking the test, each student was asked to submit demographic information

including gender, class standing, age, proficiency with the English language, race, and ethnicity. A "local code" was also correlated with each student's response so the results could be organized according to the course in which the CAT was taken. Although too extensive to include in this summary report, all of the data and their various categorizations are available upon request [4].

III. Results

Figure 1 shows a comparison between the average total score of all Westmont students who took the CAT by the end of the indicated spring semester along with the corresponding national average. We emphasize that the results should be viewed with caution since there are several factors that potentially cloud their interpretation including differences among the student populations and the fact that the 2014 (2020 and 2025) Westmont student responses were assessed internally (externally). We also note that the national data for this assessment are typically coming from universities and private institutions where all students are required to take the assessment, which is not the case at Westmont. Nevertheless, the overall downward trend in the average scores of Westmont students that was noted in the 2020 CT ILO assessment report continues here. This downward trend appears to be statistically significant based on the standard deviation of the mean of the data sets, which represent the statistical uncertainty in the average values [5]. Including this uncertainty, the 2025 average score of 16.2 ± 0.3 does not overlap with the other average scores assuming that they also have similar (or smaller) uncertainties.

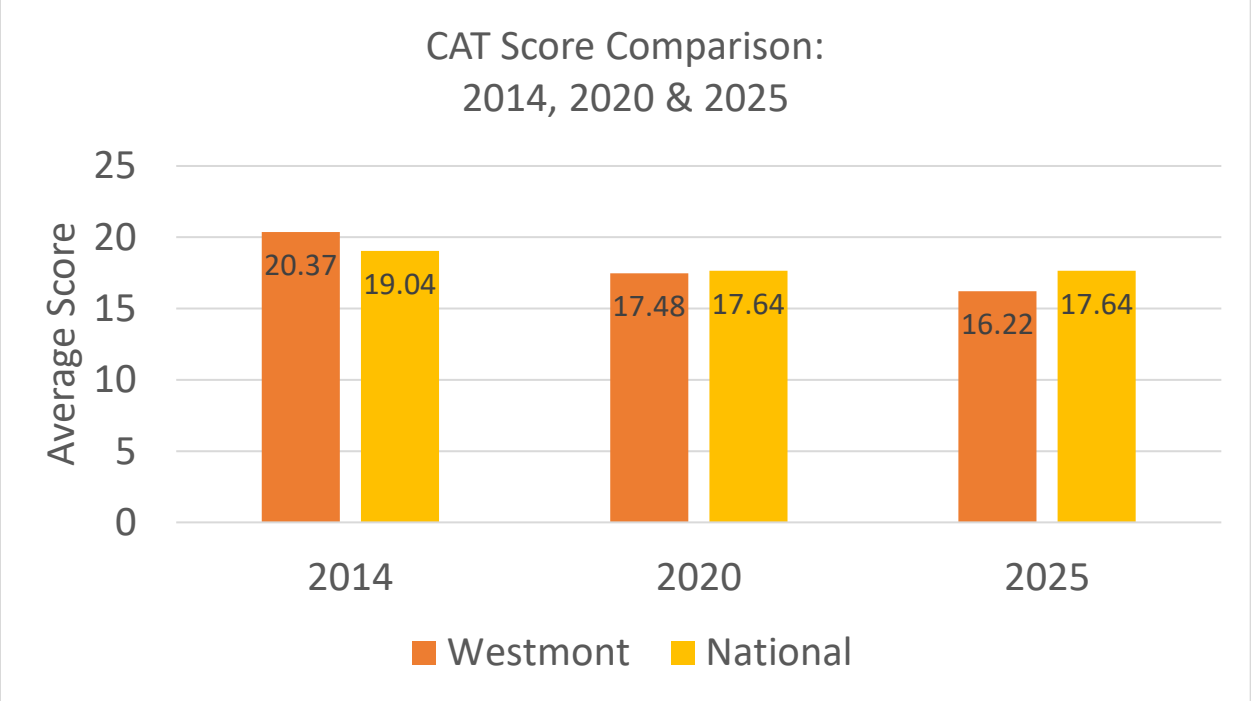

Figure 1. Comparison of average CAT scores obtained by all Westmont students by the end of the indicated spring semester along with the corresponding national average. The maximum possible score is 38.

Interestingly, the total average score obtained by first-year Westmont students (collected during the Fall 2024 semester), 16.1 ± 0.5, was only about 1% less than the senior average score and

agreed with that score within their respective uncertainties. Compared with the national average for first-year test takers (15.55), our students performed quite favorably.

Senior scores sorted by the academic division of their major are displayed in Fig. 2. As indicated in the figure, students in the Natural and Behavioral Sciences had the highest average score (17.29), followed by those in the Social Sciences (15.51) and the Humanities (15.09). A similar trend was followed in the 2020 CAT administration, although the average scores in the three divisions were higher at that time (17.88, 16.98, and 16.81, respectively).

Several comparisons available from the demographic data are shown in Table I. In general, ethnic minority students tend to underperform their white counterparts, with the exception of senior Asian students. Another comparison worth noting is the divergence between the average point scores (indicated as a percentage in Table I) of senior females $(15.9 \pm 0.5, N = 127)$ and senior males $(16.8 \pm 0.6, N = 84)$, a difference neither observed nationally (17.64 for both genders) nor in 2020 for Westmont students (17.50 for men and 17.41 for women). On the other hand, a reversal in the difference between the average scores of both genders was observed for first-year Westmont students in the Fall 2024 data, with females $(16.3 \pm 0.6, N = 75)$ outperforming males. Both genders of first-year students also had average scores above the national average for students of the same class level (15.55 for both genders).

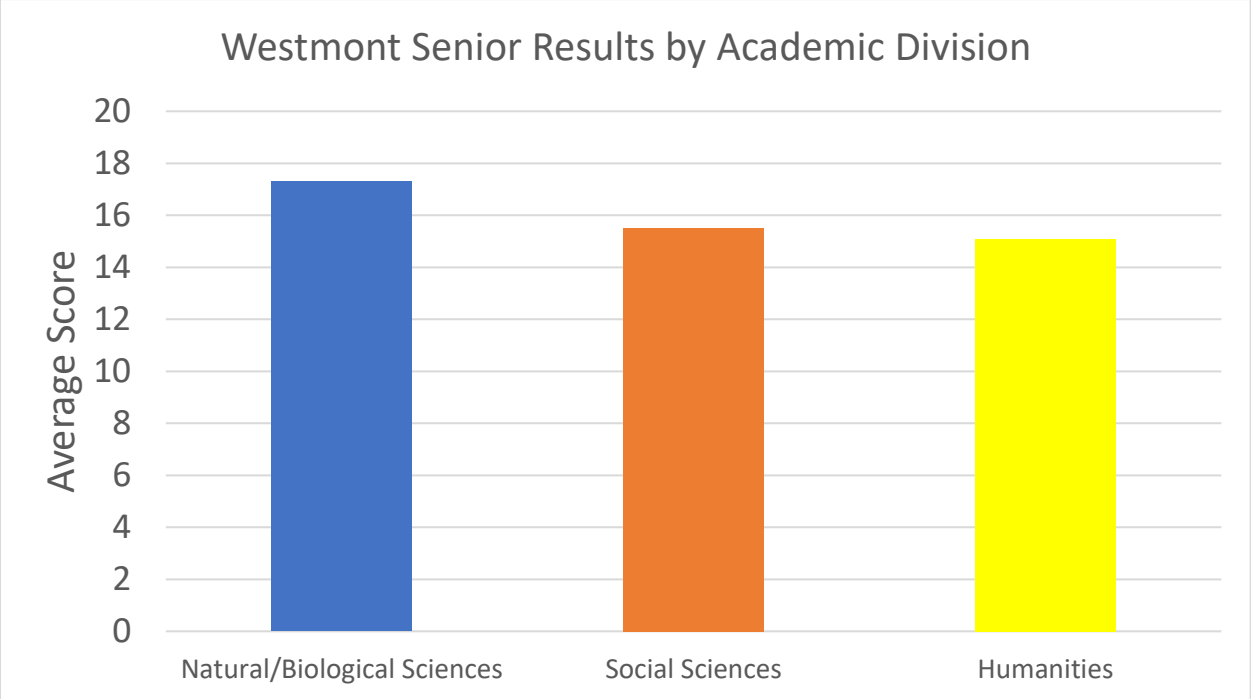

Figure 2. A comparison of the average CAT scores for 2025 Westmont seniors according to the academic division of their majors.

Various strengths in the CT skills of our students were highlighted by the results of the CAT based on the average fraction of attainable points earned on a given question (not all CAT questions are worth the same number of points) and on score comparisons with the respective national averages. Figure 3 indicates the top five questions that seniors earned the highest average fraction of attainable points. (Refer to Sec. I for the CT skill assessed by each question.)

These results suggest that our seniors are proficient overall in fundamental graphical interpretation (Q1), mathematical operations (Q12), and drawing appropriate inferences and conclusions from both relevant and irrelevant data (Q5, Q8, and Q10). Question 1 also showed one of the highest average score increases above the associated national average (0.81 points compared to 0.70). Questions 2 and 6 also showed statistically significant increases in average score above the national averages (1.56 vs. 1.20 for Q2 and 1.77 vs. 1.53 for Q6). As shown in Fig. 4, our first-year students demonstrate similar strengths, with Q1, 5, 10, and 12 ranking among the top five highest average scores (some of which are slightly higher than the corresponding scores for seniors). Moreover, Q1, 5, 6, and 13 all ranked significantly higher statistically than the associated national averages based on their average points (0.79 vs. 0.62, 0.75 vs. 0.64, 1.80 vs. 1.33, and 1.17 vs. 0.92, respectively). Of the top five strengths that were different between the two groups (Q6 for first-year students and Q8 for seniors), the average Q6 scores were much more similar (60% for first-years vs. 59% for seniors) than the corresponding Q8 scores (53% for first-years vs. 65% for seniors).

Table I. Average fraction of attainable points (in percent) on selected questions of the CAT exam along with total average scores obtained by the indicated demographic groups sorted by first-year (FY) and senior (SR) students. Scores given in green (red) indicate values with effect sizes that are significantly higher (lower) than the corresponding national average. Demographic groups indicated with an asterisk do not have their scores compared with national averages since the sample size was too small.

Group	Q1 (%)	Q7 (%)	Q12 (%)	Q15 (%)	Overall (%)
FY females	81	5	81	12	43
SR females	81	5	79	13	42
FY males	76	10	86	8	42
SR males	80	8	83	12	44
FY African American*	71	7	86	0	35
SR African American	77	4	69	8	35
FY Asian	70	10	85	10	44
SR Asian	95	10	75	13	48
FY Other Race	77	0	77	0	35
SR Other Race	65	3	76	18	40
FY White	81	8	82	11	43
SR White	80	7	83	13	44

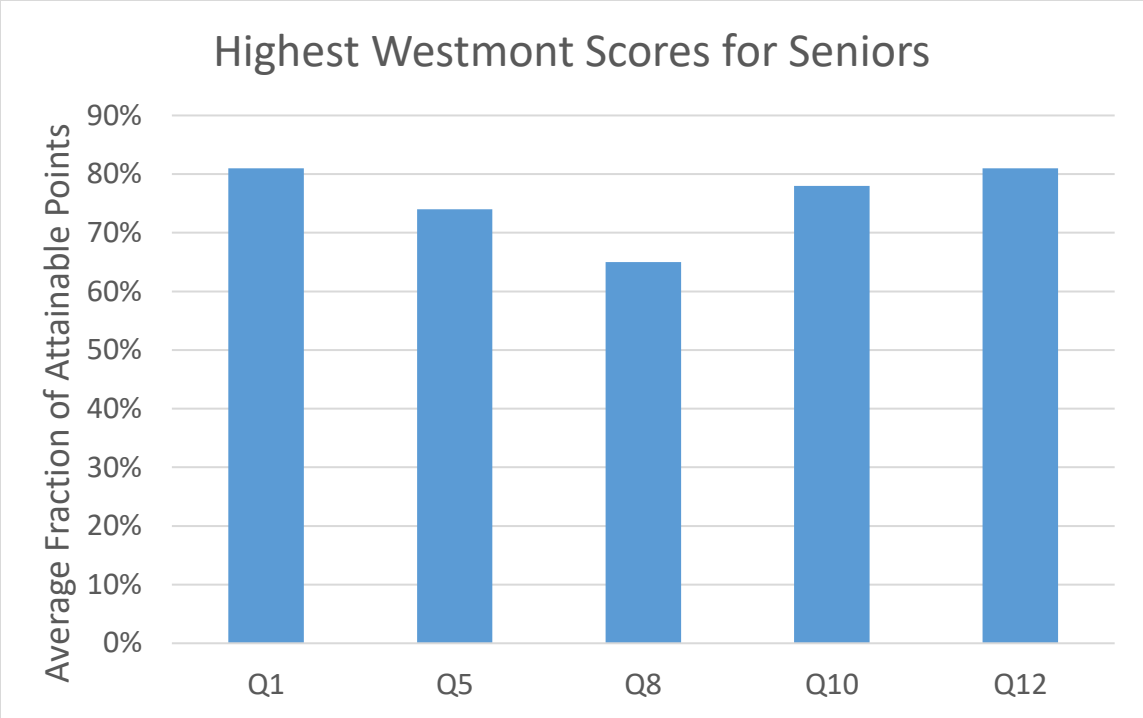

Figure 3. The top five highest average scores by CAT question obtained by seniors.

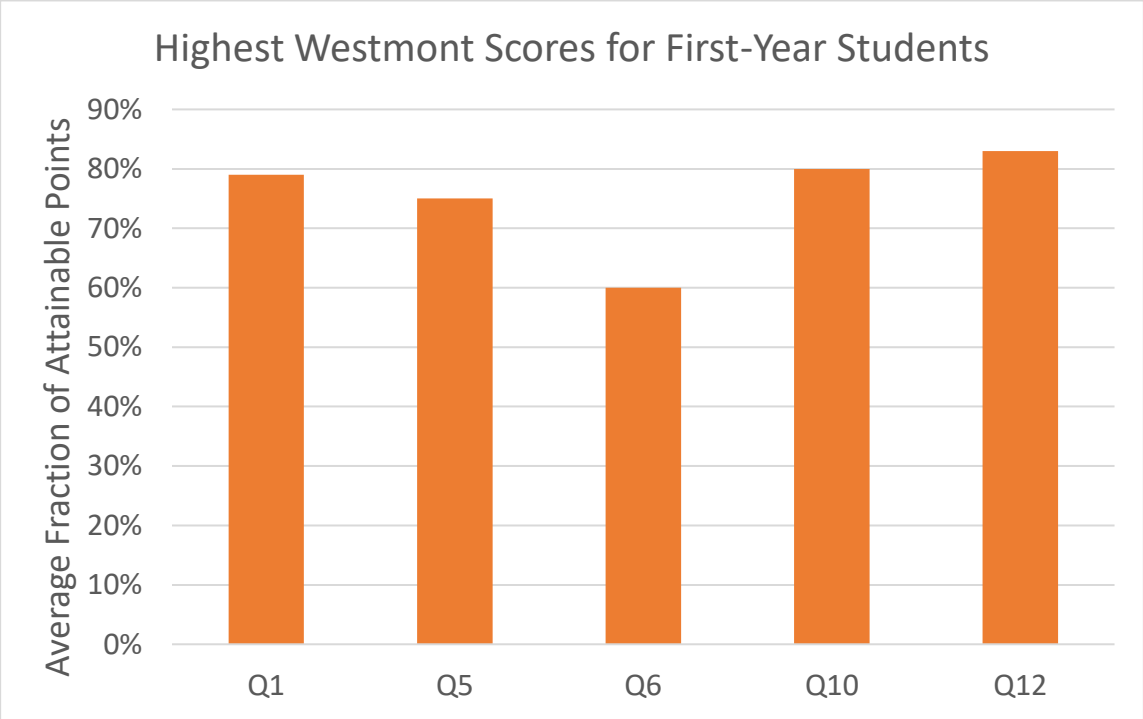

Figure 4. The top five highest average scores by CAT question obtained by first-year students.

Of course, with strengths also come opportunities for growth. According to the CAT assessment, the CT skills that our students struggle with most (based on the lowest average scores) are indicated in Figs. 5 (seniors) and 6 (first-year students). As with the CT skill strengths discussed above, strong similarities are observed in the areas of weakness between the two groups. For the

seniors, several questions were also below the respective national averages in a statistically significant way (Q4, 7, 9, 11, 14, and 15) based on the calculated effect size [6]. Consequently, the overall average CAT score for Westmont seniors (16.29) is not in agreement with the national average score (17.64) within the expected statistical fluctuations based on the p score [6] results. We also note that most of the same questions (Q4, 7, 9, and 15) were targeted as areas of weakness in the 2019–2020 CT ILO summary report [7]. As discussed in that report, these questions involve problem solving, creative thinking, and effective communication (Q 4, 7, and 15) as well as creative thinking and effective communication (Q9), pointing to a persistent deficit in these areas for our students.

The new vantage point of first-year student performance provided by this study reveals that our students appear to arrive at Westmont with similar CT skill challenges (see Fig. 6). However, only the average Q7 and 15 scores for these students were determined to be significantly lower than the corresponding national averages based on the effect sizes.

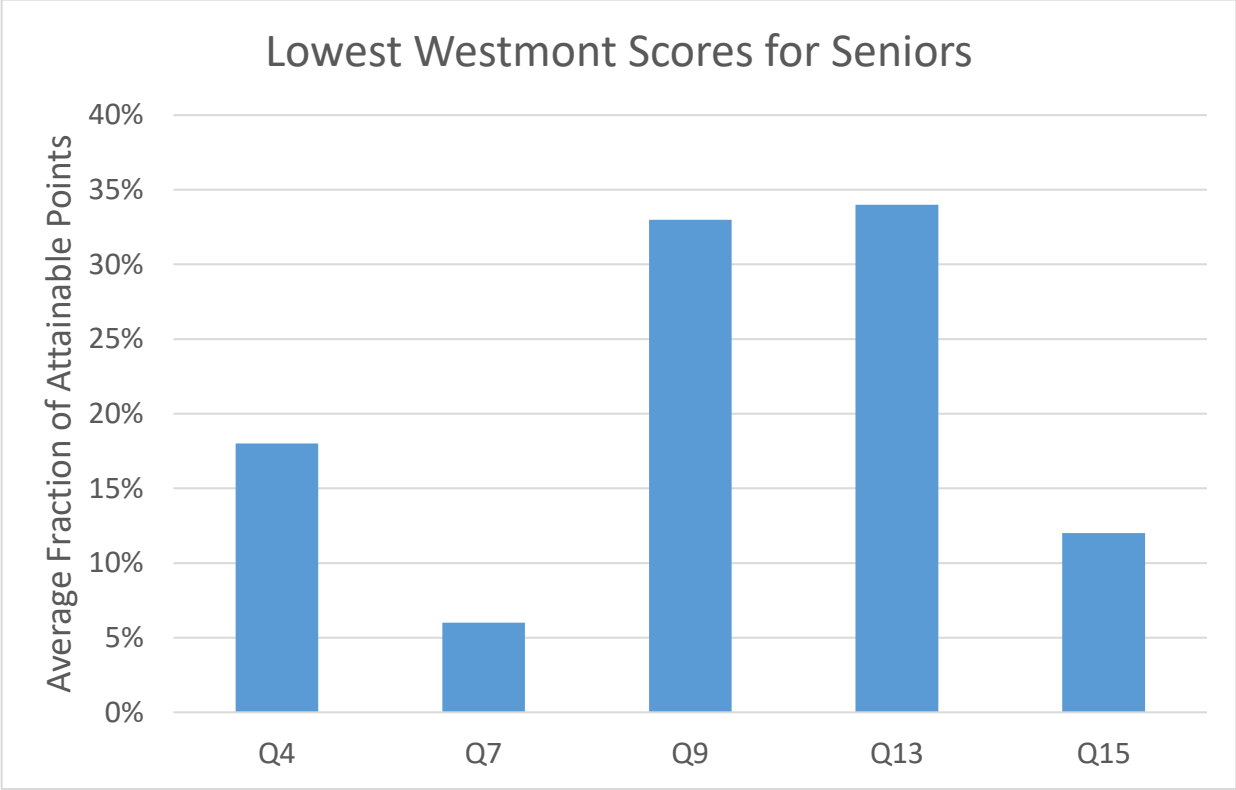

Figure 5. The five lowest average scores by CAT question obtained by seniors.

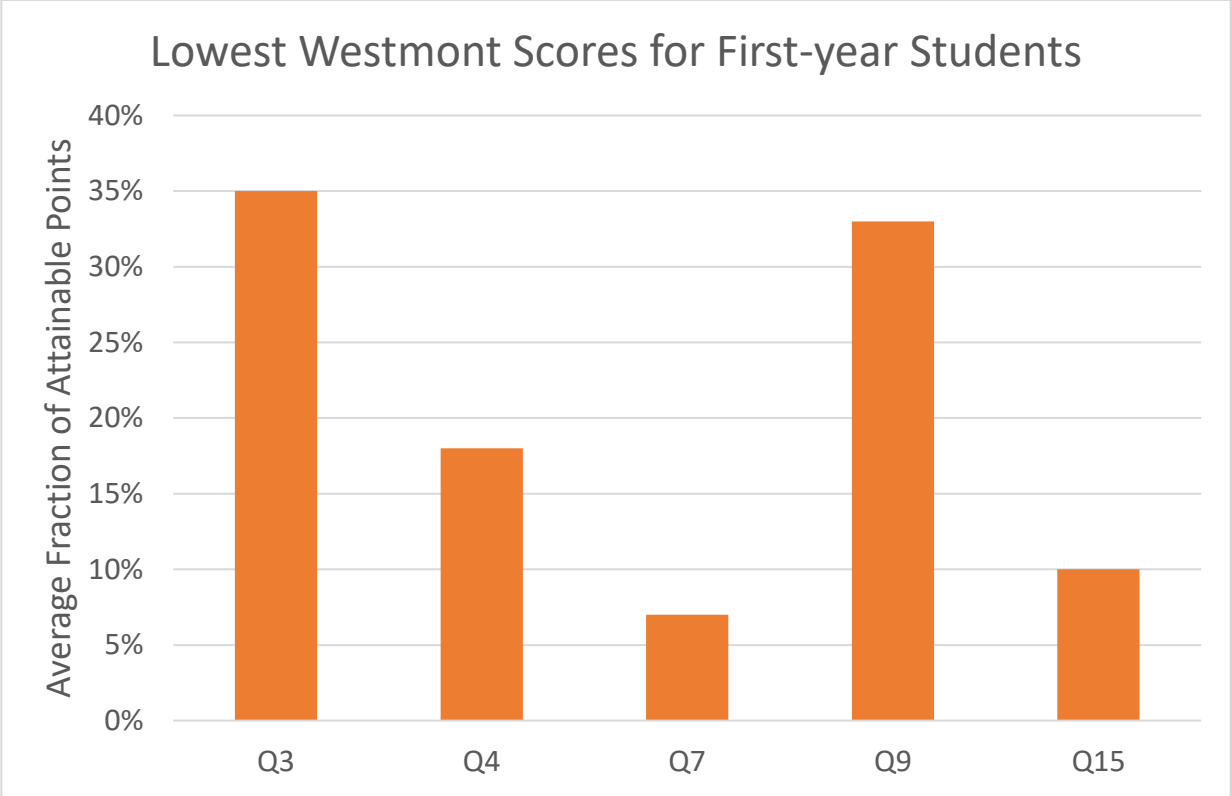

Figure 6. The five lowest average scores by CAT question obtained by first-year students.

IV. Discussion

Overall performance summary. Considering all test takers, independent of class level, Westmont students performed slightly below the national average (see Fig. 1), with a small effect size of -0.25. The spread in scores (see Sec. I) was moderate, indicating variability in critical thinking skill levels.

Skill area performance (selected highlights). Each CAT question (see Sec. I) maps to skills in critical thinking, problem solving, creative thinking, or effective communication. Table I indicates some skill-specific insights.

Results that correlate to evaluation and interpretation skills suggest that our students perform relatively well at summarizing data and interpreting correlations. However, they struggle to identify information needed to evaluate or test hypotheses. In terms of problem solving, these same students show strong ability to distinguish relevant information and use basic mathematics. On the other hand, there is a need to improve application of information to problem solving. Students generally perform better in recognizing flaws in reasoning or considering alternative explanations based on the creative thinking and reasoning category of questions. Communication of reasoning, especially adapting explanations to changing contexts, is a major growth area.

Table I. Westmont student performance on CAT questions associated with the indicated skills.

🧠 **Evaluate & Interpret Information**

Skill	Mean Score	% of Max Points	Comparison to National	Effect Size
Q1 – Summarize graph	0.80	80%	Higher	**+0.27**
Q2 – Evaluate correlation	1.53	51%	Higher	**+0.29**
Q4 – Info needed for hypothesis	0.69	17%	Lower	**-0.41**
Q7 – Add'l info for hypothesis	0.13	6%	Much lower	**-0.85**

🔍 **Problem Solving**

Skill	Mean Score	% of Max Points	Comparison to National	Effect Size
Q10 – Relevant vs. irrelevant info	3.13	78%	On par	0.00
Q11 – Apply relevant info	0.76	38%	Lower	**-0.32**
Q12 – Math skills	0.81	81%	Equal	0.00

🎨 **Creative Thinking & Reasoning**

Skill	Mean Score	% of Max Points	Comparison to National	Effect Size
Q6 – Spurious associations	1.76	59%	Higher	**+0.27**
Q3 – Alt. explanations for patterns	1.21	40%	Slightly higher	0.06

📣 **Communication**

Skill	Mean Score	% of Max Points	Comparison to National	Effect Size
Q14 – Best solution & explanation	1.94	39%	Lower	**-0.15**
Q15 – Impact of problem changes	0.36	12%	Much lower	**-0.65**

Subgroup insights: Seniors vs. first-year students. Westmont's seniors slightly outperformed the national norms in some areas while underperforming in others. It is difficult to compare the senior and first-year student results, though, since we do not know the extent that the current first-year students entered Westmont with similar CT skills as the current seniors had when they were first-year students. As mentioned earlier, there are several factors (including the different

students incorporated in each group, sample size, randomness of each sample, previous academic preparation, etc.) that cloud the interpretation of the results and the comparisons between them. Although there is a systematic decline in the overall CAT performance by Westmont seniors since 2013–2014, the reasons for that decline are also unclear for similar reasons. However, we can say with certainty that the same CT skill growth areas identified in 2019–2020 remain so in this assessment cycle. In particular, effect sizes on key questions like Q4, Q7, Q15 were significantly negative for seniors, suggesting weak development in advanced evaluative reasoning and adaptability.

Notable strengths.

- Data interpretation and mathematical reasoning
- Basic identification of relevant vs. irrelevant information
- Recognition of misleading or spurious associations

Key growth areas.

- Evaluating hypotheses and identifying missing information (Q4, Q7)
- Explaining and adapting problem solutions in dynamic contexts (Q15)
- Synthesizing and articulating complex solutions (Q14)

V. Recommendations

Based on the CAT assessment data alone, the CT Assessment Team offers the following recommendations to improve outcomes. Although many (if not all) of these approaches are already being followed to some extent in our curriculum, perhaps a more systematic (and intentional) implementation across courses and disciplines may be needed.

1. Embed hypothesis evaluation practice in assignments and discussions across disciplines.
2. Use case-based learning that requires identifying gaps in data and proposing investigative approaches.
3. Incorporate metacognitive prompts to encourage students to explain how and why a solution might change with different constraints.
4. Improve writing-to-learn strategies to develop articulation of problem-solving and reasoning. Writing courses that naturally pair with quantitative reasoning ones (such as laboratories in the natural and life sciences) might be strong candidates to help develop these skills.
5. Use CAT-style questions in courses to simulate real-world critical thinking scenarios. Instructors could use these questions as a way to frame the CT skills that they wish to emphasize in their courses. A similar point was also raised in the 2019–2020 CT ILO assessment report.
6. Administer the CAT again in four years so the first-year student population assessed in this study can be evaluated as seniors.

We also recognize that there are other, more institution-specific, needs and preferences that should be addressed. To that end, we invited department chairs, instructors who participated in this CAT assessment, and other faculty with previous experience and investment in ILO assessment to participate in a CAT data discussion on May 14. During the meeting, most of the

data summarizing the CAT results were presented, and a rich discussion of the implications and potential next steps followed. Several key points were raised and led to the following set of additional recommendations based on this preliminary discussion:

1. Target specific courses within the Exploring Physical and Life Sciences general education (GE) curricula that could serve as a conduit for deepening CT competencies.
2. Bridge CT skills into students' majors at the department level to make the skill transfers explicit, meaningful, and relevant while maximizing student engagement and motivation.
3. Further focused discussions by the broader faculty community should be encouraged to discern whether the CT skills associated with our areas of greatest weakness (Q4, 7, 9, 11, 14, and 15) are being addressed in our curriculum and if we even want them to be.
4. In a broader context, our community may benefit by having a college-wide conversation about what we mean by critical thinking skills. Discussions could explore whether we are assessing exactly what we mean by critical thinking skills.
5. These CAT results could be presented at an upcoming Department Chair meeting. Consider asking chairs to discuss in their department how CT skills are used in their disciplines. Specifically, what assignments help their students strengthen these skills? How can faculty support discipline-specific development of CT competencies?
6. Departments could then have follow-up conversations and clearly identify which CT questions they could reasonably address within the department and in which course, then share that data. This might actually help us understand what areas our students should be excelling at because instructors are clearly trying to teach those skills.
7. Identify other critical thinking assessment instruments and their viability for the CT ILO assessment next time. One possibility may be the Watson-Glaser critical thinking assessment tool [8]. However, in the opinion of the current CT ILO Assessment Team, this assessment instrument focuses on evaluating individual CT skills while ILO assessment is more concerned with seeing patterns of student development and performance. Perhaps this instrument is better suited for use in selected individual courses.

References

[1] https://www.westmont.edu/sites/default/files/InstitutionalLearningOutcomes_07_24_2017.pdf
[2] https://www.tntech.edu/cat/
[3] Kevin Possin, *Informal Logic* **40(3)**, 489 (2020).
[4] Contact Bob Haring-Kaye (rharingkaye@westmont.edu) for further information.
[5] John R. Taylor, *An Introduction to Error Analysis*, Second Ed. (University Science Books, Sausalito, 1997).
[6] Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed. (Routledge, 1988).
[7] Jim Taylor, Steve Contakes, Angela D'Amour, Edd Noell, Randy VanderMey, and Tatiana Nazarenko, 2019–2020 Institutional Assessment of Critical Thinking: Summary Report.
[8] https://www.talentlens.com/recruitment/assessments/watson-glaser-critical-thinking-appraisal.html