UNIVERSITY OF TEXAS ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

MEASURING CRITICAL THINKING USING AAC&U VALUE RUBRICS AT THE UNIVERSITY OF TEXAS ARLINGTON

Spring 2024 Report

Measuring Critical Thinking, Spring 2024 Report

Critical thinking is a fundamental skill that enables students to analyze, evaluate, and synthesize information to make reasoned decisions. As a core competency in higher education, it is crucial in preparing students for academic success, professional careers, and informed civic engagement (Facione, 2011). Recognizing its significance, the Texas Higher Education Coordinating Board (THECB) has designated Critical Thinking as one of the six core objectives within the Texas Core Curriculum (TCC), ensuring its integration across all eight Foundational Component Areas (FCAs) (THECB, 2019). The core objective of critical thinking is to encompass essential cognitive processes such as creative thinking, inquiry, problem-solving, and logical reasoning, which are vital for students' intellectual and professional development (Paul & Elder, 2019).

At The University of Texas at Arlington (UTA), the assessment of Critical Thinking is conducted as part of a structured, multi-year cycle evaluating all six TCC core objectives. This systematic approach minimizes the burden on academic departments while maintaining consistent data collection to inform curricular improvements. The assessment process involves the collection of student work samples from embedded assignments within core curriculum courses. These samples are evaluated by faculty and staff using the Critical Thinking VALUE Rubric, a standardized assessment tool developed by the Association of American Colleges and Universities (AAC&U, 2013), ensuring consistency and rigor in measuring student proficiency.

This report presents the findings from the assessment of Critical Thinking conducted in May 2024 at UTA, with a specific focus on student work samples from the Social and Behavioral Sciences FCA courses. The results of this assessment provide valuable insights into student learning and serve as a foundation for faculty-led discussions on improving instructional strategies. Research suggests that developing critical thinking skills is closely linked to academic achievement, workforce readiness, and lifelong learning (Abrami et al., 2015). UTA reaffirms its commitment to fostering an educational environment that equips students with the analytical and reasoning skills necessary for success in an increasingly complex world by prioritizing the assessment and enhancement of critical thinking.

Method

Participants

The project gathered evidence of *Critical Thinking* within a representative sample of students enrolled in Texas Core Curriculum (TCC) courses at UTA and recruited qualified raters to read and score each written student artifact. The student population assessed for Critical Thinking in Spring 2024 consisted of 231 students representing diverse demographic characteristics. In terms of gender distribution, the majority of students were female (76.62%), while male students accounted for 23.38% of the sample.

Regarding racial and ethnic composition, the most prominent representation came from Hispanic/Latino students (31.17%), followed by White students (24.24%), and Black/African American students (19.05%). Asian students comprised 12.99% of the assessed population, while students with multiple ethnicities comprised 6.49%. A smaller proportion of students identified as Foreign (5.19%), Native Hawaiian/Other Pacific Islander (0.43%), and American Indian/Alaska Native (0.43%).

Table 1: Student Demographics

Categorical Information	N	%
Gender		
Female	177	76.62%
Male	54	23.38%
Racial/Ethnic Description		
Hispanic/Latino	72	31.17%
White	56	24.24%
Black/African American	44	19.05%
Asian	30	12.99%
Multiple Ethnicities	15	6.49%
Foreign	12	5.19%
Native Hawaiian/Other Pacific Islander	1	0.43%
American Indian/Alaska Native	1	0.43%
First Generation Student		
Non-First Generation	137	40.69%
First Generation	94	59.31%

The dataset also captured students' first-generation status, a key demographic factor often linked to academic experiences and performance. First-generation students accounted for 40.69% of

the assessed group, whereas non-first-generation students comprised 59.31%. This distinction provides valuable insight into potential variations in critical thinking development based on students' educational backgrounds and access to academic resources. These demographic findings offer an essential context for interpreting the assessment results and identifying potential areas for targeted instructional support and curricular enhancements (see Table 1 for all details).

The assessed student population for Critical Thinking in Spring 2024 included a range of academic levels, enrollment years, and enrollment statuses. In terms of academic standing, the largest group consisted of junior-level students (37.66%), followed by seniors (27.27%) and sophomores (26.84%). A smaller portion of the sample included freshmen (6.93%) and fifth-year students (1.30%).

Table 2: *Student Status at UT Arlington*

Categorical Information	N	%
Academic Level		
Junior	87	37.66%
Senior	63	27.27%
Sophomore	62	26.84%
Freshman	16	6.93%
Fifth Year	3	1.30%
Enrollment Year		
2021 - 2022	84	36.36%
2020 - 2021	68	29.44%
2022 - 2023	37	16.02%
2019 - 2020	27	11.69%
2018 and Prior	15	6.49%
Academic Load		
Enrolled Full-Time	200	86.58%
Enrolled Part-Time	31	13.42%
Transfer Student		
Non-Transferred	175	75.76%
Transferred	56	24.24%

*Eligibility as of Spring 2024

Examining students' enrollment years, the majority entered UT Arlington between 2021 and 2022 (36.36%), with another 29.44% enrolling between 2020 and 2021. A smaller proportion of students began their studies in 2022–2023 (16.02%), 2019–2020 (11.69%), or 2018 and prior (6.49%). Regarding academic load, most students were enrolled full-time (86.58%), while a smaller

group, 13.42%, attended part-time. Additionally, the student population included transfer students (24.24%), though the majority, 75.76%, had completed all their coursework at UT Arlington. These enrollment characteristics provide essential context for assessing Critical Thinking, as student status and academic experience may influence engagement with coursework, exposure to critical thinking instruction, and overall performance on assessment measures.

The Spring 2024 Critical Thinking assessment student sample represented a diverse range of academic disciplines across UT Arlington. The College of Nursing & Health Innovation had the highest representation, comprising 31.17% (72 students) of the assessed population. The College of Liberal Arts followed with 19.48% (45 students), reflecting a strong presence of students from humanities and social sciences disciplines.

Table 3: *Students by Colleges/Schools*

College/School	No. of Students	Percentage
College of Nursing & Health Innovation	72	31.17%
College of Liberal Arts	45	19.48%
College of Science	26	11.26%
College of Business	26	11.26%
College of Engineering	25	10.82%
College of Architecture, Planning & Public Affairs	12	5.19%
Division of Student Success	10	4.33%
School of Social Work	9	3.90%
College of Education	6	2.60%

Other colleges with significant representation included the College of Science (11.26%), College of Business (11.26%), and College of Engineering (10.82%), suggesting a balanced distribution of students from both STEM and non-STEM fields. Smaller proportions of students came from the College of Architecture, Planning & Public Affairs (5.19%), Division of Student Success (4.33%), School of Social Work (3.90%), and College of Education (2.60%).

The varied academic backgrounds of students provide insight into how critical thinking skills are developed and applied across disciplines. This distribution highlights the broad integration of critical thinking instruction within the core curriculum, ensuring that students from multiple fields engage with analytical reasoning and problem-solving skills as part of their academic experience.

Procedure

As part of the assessment process, students enrolled in core courses completed assignments designed to evaluate critical thinking skills. Written student work samples were obtained from multiple sections of WOMS 2310 Intro Women's & Gender Studies and GLOBAL 2301 Intro to Global Issues courses offered in spring 2022, spring 2023, and summer 2023 semesters. In these courses, students were tasked with writing a paper that required them to examine an issue from multiple viewpoints.

All submitted work samples underwent a preparation phase before scoring to ensure a fair and unbiased evaluation process. Each paper was assigned a coded tracking number, and all personally identifiable information—including student names and faculty instructor names—was removed. This anonymization process helped mitigate potential rater bias during the structured Scoring Day activities.

Assessment Instrument

The assessment of student work samples was conducted using the AAC&U's Critical Thinking VALUE Rubric (AAC&U, 2019). Developed by a multidisciplinary team of faculty experts with support from the Lumina Foundation, this rubric provides a structured framework for evaluating critical thinking proficiency (see Appendix A). It consists of five key dimensions that define critical thinking: 1) Explanation of Issues, 2) Evidence, 3) Influence of Context & Assumptions, 4) Student's Position (Perspective, Thesis, or Hypothesis), and 5) Conclusions & Related Outcomes (Implications and Consequences).

Each dimension includes a detailed narrative description outlining the expected quality of responses. The rubric functions as a matrix that provides narrative descriptions of expected work quality and corresponding point values for scoring the five measures. The point values range from 1 to 4, with 1 indicating baseline performance (Benchmark-1), 2 indicating approaching milestone (Milestone-2), 3 indicating achieved milestone (Milestone-3), and 4 indicating the highest mastery (Capstone-4) of Critical Thinking. This rubric ensured consistency in evaluation and facilitated a comprehensive analysis of students' analytical and reasoning skills. The attainment target (numerical ratings) was set at a score of 2 (Milestone-2). The attainment target was set above the benchmark following recommendations from AAC&U research (Greenhoot & Bernstein, 2012) and standard acceptance criteria in the assessment community.

All raters assigned a score to each of the five dimensions in the rubric for each student's work sample. Higher values indicate more evidence of Critical Thinking in student work and vice versa. Raters were advised to use zero per AAC&U recommendations if any dimension is absent in student work. Typically, in student samples, the five dimensions are adequately represented in the narrative.

Raters, Rater Calibration, and Scoring

Raters scored the student writing samples during a scheduled scoring day, and each paper was reviewed twice (by two separate raters) in a group setting. A third "tiebreaker" rating was obtained when ratings diverged by more than one rating interval on a single dimension. In these cases, the mean score of the three ratings was used as a final score. The rating group included twelve faculty members and professional staff with advanced degrees.

The scoring day began with an orientation and description of the rating process. A qualified UTA facilitator led the raters through reviewing the rubric and discussing the rating dimensions and scale designed to calibrate the raters' understanding and use of the rubric in the rating process. Then, the entire group read and rated one practice anchor paper, which was chosen beforehand by the facilitator. Following the sample paper review, the facilitator led a discussion among all raters using the anchor paper to reach a common understanding of the Critical Thinking dimensions and to find exemplar indicators within the paper for the rubric levels of mastery. Following the completion of the calibration activity, formal review and rating of the de-identified student papers began. During the formal review and rating of papers, raters read each paper and assigned scores for each dimension on the rubric using the four-point scale (plus the available "zero" rating). If the values of the skill measure scores for a paper from the two raters were identical or within a point difference, then the two scores were considered in agreement and averaged. For example, if Rater A scored the Content Development measure with a value of 2 and Rater B scored the same measure with a value of 3, then the rating was considered in agreement, and scores for that dimension were averaged, resulting in a score value of 2.5. If the scores from the two raters differed by over two points, a third rater was assigned the paper then the three scores were averaged together to determine the final score in such cases. For example, if Rater A scored the Content Development measure with a value of 1 and Rater B scored the same measure with a value of 3, the rating was not in agreement, and a third rater was asked to read and score the paper.

Analysis and Results

Inter-rater Agreement and reliability

Inter-rater agreement was analyzed across the five rubric dimensions to ensure reliability in the assessment of Critical Thinking. The agreement percentage was calculated by determining the proportion of ratings in which two raters assigned scores within one rating point of each other. Once each paper had been rated twice, the IER staff collected the rating sheets, entered the rating scores into a spreadsheet, and analyzed them to determine agreement. Each score was calculated as the average of the two raters' scores if the values assigned by the raters differed by one point or less. The agreement percentages among raters across different dimensions of the Critical Thinking VALUE Rubric demonstrate high consistency in scoring.

The results indicate a high level of agreement among raters, with all dimensions exceeding 90% agreement. Specifically, Evidence and Influence of Context & Assumptions demonstrated the highest agreement at 94%, followed by Conclusions & Related Outcomes at 92%. Student's Position showed 91% agreement, while Explanation of Issues had the lowest but still strong agreement at 90%. These results suggest that the assessment process maintained strong consistency across raters, reinforcing the reliability of the rubric and the scoring process. The high levels of agreement indicate that faculty raters applied the rubric criteria consistently when evaluating student work.

Inter-rater agreement analyses assessed whether the rater scores corresponded to each other for a particular student paper. The intraclass correlation coefficient (ICC) determined levels of agreement. High ICC values (Cronbach's Alpha) indicate more agreement between rater scores (Fleiss, 1986; Shrout & Fleiss, 1979). For this sample, values indicated a trend of good agreement (see Table 3). These high values give confidence to proceed with analyses involving student attainment.

Table 4: *Agreement Percentages Among Raters*

Dimension (Critical Thinking VALUE Rubric)	Percentages
Explanation of Issues	90%
Evidence	94%
Influence of Context & Assumptions	94%
Student's Position	91%
Conclusions & Related Outcomes	92%

Note: If values assigned by the raters differed by the rating interval of one point or less, it was counted as agreement. The agreement percentage was computed by dividing the number of agreements by the total number of ratings.

Apart from the simple percentage agreements, researchers widely measure the reliability of rating agreements between different raters to eliminate chance agreements. All raters who participated in the scoring process had advanced degrees, teaching experience, and had attended the scoring training just before the scoring session. Hence, the probability of chance agreement was very low, but inter-rater agreement was computed to follow best research practices. Inter-rater reliability is the consistency among raters when scoring the same subjects independently. The extent to which different raters agree on their judgments establishes the validity and credibility of measurements or ratings.

The inter-rater agreement was determined to check the consistency level of the rating by calculating the Intraclass Correlation Coefficient (ICC). ICC) values reflect the reliability of scores assigned by different raters across various dimensions of the Written Communication VALUE Rubric. ICC is commonly used to assess the degree of agreement among raters beyond chance. High ICC values indicate more reliability between rater scores. Commonly accepted guidelines were used to interpret the ICC results. These suggest that the range of 0.40 to 0.74 is considered fair to good inter-rater agreement, with results above 0.74 classified as excellent inter-rater agreement and results lower than 0.40 considered poor inter-rater agreement (Fleiss, 1986; Shrout & Fleiss, 1979).

Table 5: *Intraclass Correlation Coefficient for Critical Thinking Dimensions*

Critical Thinking VALUE Rubric Dimension	*n* = 231
Explanation of Issues	0.58
Evidence	0.57
Influence of Context & Assumptions	0.60
Student's Position	0.53
Conclusions & Related Outcomes	0.56

Note 1: *less than 0.40 = poor agreement; between .40 and .74 = fair to good agreement; greater than .74 = excellent agreement.*

Note 2: The intra-class correlation coefficient (ICC) was calculated as a two-way random effects model. Values in this model type with random rater pairings are typically expected to be lower than those where rater pairings are fixed throughout the rating day.

Results indicate that all five dimensions demonstrated fair to good agreement, with ICC values ranging from 0.53 to 0.60. The highest reliability was observed in the Influence of Context & Assumptions (0.60), followed by the Explanation of Issues (0.58) and Evidence (0.57). The Conclusions & Related Outcomes dimension had an ICC of 0.56, while Student's Position exhibited the lowest reliability at 0.53. These findings align with expectations for a one-way random effects

model, where random pairings of raters can lead to slightly lower ICC values than models with fixed rater pairings. Although the agreement levels fall within the fair to good range, the variation in scores suggests potential differences in the interpretation of rubric criteria among raters. Future assessments may benefit from additional rater calibration sessions to enhance scoring consistency.

Students Performance

The mean scores for the Critical Thinking dimensions provide insight into students' performance across the five rubric categories (Table 6). The attainment target was established at a mean score of 2.0 or higher, indicating milestone-level proficiency. Students exceeded the attainment target in only one dimension, Explanation of Issues (M = 2.11, SD = 0.73). This indicates students were generally proficient in identifying and clearly articulating the main issues addressed in their assignments. However, students fell below the attainment target in all other dimensions. Scores were notably lower in Influence of Context and Assumptions (M = 1.68, SD = 0.68), Conclusions and Related Outcomes (M = 1.61, SD = 0.66), Evidence (M = 1.56, SD = 0.67), and Student's Position (M = 1.56, SD = 0.65). These lower scores highlight significant challenges students faced in evaluating context, formulating logical conclusions, integrating and supporting evidence, and clearly articulating their positions or arguments.

Table 6
Means for Critical Thinking Measure Scores

Measurement Dimensions	N	Mean	SD
Explanation of Issues	231	2.11	0.73
Evidence	231	1.56	0.67
Influence of Context and Assumptions	231	1.68	0.68
Student's Position	231	1.56	0.65
Conclusions and Related Outcomes	231	1.61	0.66

Overall, while students demonstrated milestone proficiency in explaining issues, targeted instructional efforts are required to improve student performance in the other critical thinking dimensions, moving students toward meeting the established attainment target.

Summary, Observations, and Limitations

This assessment of Critical Thinking was conducted in Spring 2024 at The University of Texas at Arlington, focused specifically on student work from the Social and Behavioral Sciences

FCA. Faculty raters evaluated student work using the AAC&U Critical Thinking VALUE Rubric, assessing five dimensions: Explanation of Issues, Evidence, Influence of Context and Assumptions, Student's Position, and Conclusions and Related Outcomes. Results indicated that students achieved the attainment target (mean ≥ 2.0) only in the dimension Explanation of Issues (M = 2.11), reflecting a strength in identifying and articulating core issues.

However, students did not reach the attainment target in the other four dimensions. The lowest performance occurred in Evidence (M = 1.56) and Student's Position (M = 1.56), demonstrating that students faced significant difficulties in effectively integrating evidence and clearly articulating their arguments. Lower scores in Influence of Context and Assumptions (M = 1.68) and Conclusions and Related Outcomes (M = 1.61) further indicate that students struggled with critical analysis and drawing reasoned conclusions. These findings suggest a clear need for instructional interventions emphasizing evidence-based reasoning, critical analysis of assumptions, and logical argumentation.

Rater agreement percentages remained high (90%–94%), demonstrating strong consistency among raters. Reliability scores, measured by the Intraclass Correlation Coefficient (ICC), fell in the fair-to-good range (0.53–0.60), typical for assessments using random rater pairings.

Several limitations should be considered when interpreting these results. The assessment was limited to a single FCA (Social and Behavioral Sciences), reducing generalizability across the broader curriculum. Future assessments should expand to include multiple FCAs to gain a more comprehensive understanding of critical thinking development. Variability in assignment prompts across different courses likely influenced performance, suggesting that aligning prompts explicitly with rubric dimensions could improve consistency and student outcomes.

Overall, these findings offer valuable insights into student proficiency in critical thinking and indicate areas requiring targeted instructional support and assignment refinement to achieve the established milestone-level attainment target across all rubric dimensions.

References

Association of American Colleges and Universities (2019). *VALUE Rubrics*. Retrieved from https://www.aacu.org/value-rubrics/

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. Review of Educational Research, 85(2), 275–314. https://doi.org/10.3102/0034654314558493

National Association of Colleges and Employers. (2018). *Job Outlook 2016*. Bethlehem, PA.

Rhodes, T. (Ed.). (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Association of American Colleges and Universities. (2013). *VALUE rubrics*: Critical Thinking. Retrieved from https://www.aacu.org/value-rubrics

Fleiss J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley & Sons.

Facione, P. A. (2011). Critical thinking: What it is and why it counts. Insight Assessment.

Paul, R., & Elder, L. (2019). Critical thinking: Tools for taking charge of your learning and your life (4th ed.). Pearson.

Texas Higher Education Coordinating Board. (2019). Texas core curriculum [Report]. http://board.thecb.state.tx.us/apps/TCC/

Shrout, P., & Fleiss, J. L. (1979). Intraclass correlation: uses in assessing rater reliability. Psychological Bulletin, 86(2), 420 - 428.

Appendix A: Critical Thinking VALUE Rubric

CRITICAL THINKING VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

Definition

Critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone	**Milestones**		**Benchmark**
	4	3	2	1
Explanation of issues	Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.	Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.	Issue/problem to be considered critically is stated but description leaves some terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.	Issue/problem to be considered critically is stated without clarification or description.
Evidence *Selecting and using information to investigate a point of view or conclusion*	Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis. Viewpoints of experts are questioned thoroughly.	Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis. Viewpoints of experts are subject to questioning.	Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis. Viewpoints of experts are taken as mostly fact, with little questioning.	Information is taken from source(s) without any interpretation/evaluation. Viewpoints of experts are taken as fact, without question.
Influence of context and assumptions	Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.	Identifies own and others' assumptions and several relevant contexts when presenting a position.	Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).	Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.
Student's position (perspective, thesis/hypothesis)	Specific position (perspective, thesis/hypothesis) is imaginative, taking into account the complexities of an issue. Limits of position (perspective, thesis/hypothesis) are acknowledged. Others' points of view are synthesized within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) takes into account the complexities of an issue. Others' points of view are acknowledged within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) acknowledges different sides of an issue.	Specific position (perspective, thesis/hypothesis) is stated, but is simplistic and obvious.
Conclusions and related outcomes (implications and consequences)	Conclusions and related outcomes (consequences and implications) are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.	Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes (consequences and implications) are identified clearly.	Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes (consequences and implications) are identified clearly.	Conclusion is inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.