UNIVERSITY OF
TEXAS
ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

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

Spring 2025 Report

Measuring Critical Thinking, Spring 2025 Report

Critical thinking is an essential skill enabling students to analyze, evaluate, and synthesize information. As a key competency in higher education, critical thinking significantly contributes to academic success, career readiness, and active civic participation (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). Research consistently demonstrates that strong critical thinking skills correlate with higher academic achievement, enhanced workforce readiness, and lifelong learning capabilities (Abrami et al., 2015; Tiruneh et al., 2018).

At The University of Texas at Arlington (UTA), the assessment of Critical Thinking is conducted as part of a structured, two-year cycle evaluating all six TCC core objectives. 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.

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. A total of 299 undergraduate student work samples were analyzed in the assessment of the Critical Thinking core objective during the Spring 2025 semester. The gender distribution of the assessed students was relatively balanced, with 52.17% identifying as female and 47.83% as male.

In terms of racial and ethnic background, nearly half of the students (47.49%) identified as Hispanic/Latino, followed by White (16.05%), Asian (12.37%), and Black/African American (12.04%) students. Smaller proportions of students identified as Foreign (7.69%), Multiple Ethnicities

(3.01%), or did not specify a racial/ethnic identity (1.34%). The diverse demographic profile provides valuable context for interpreting the assessment results and underscores the importance of ensuring inclusive and equitable instructional practices across the core curriculum. A majority of the students (57.86%) were identified as first-generation college students, indicating that they are the first in their families to attend college. Non-first-generation students comprised 38.13% of the sample, while first-generation status data was unavailable for 4.01% of the students (see Table 1 for details).

Table 1: Student Demographics

Categorical Information	N	%
Gender		
Female	156	52.17%
Male	143	47.83%
Racial/Ethnic Description		
Hispanic/Latino	142	47.49%
White	48	16.05%
Asian	37	12.37%
Black/African American	36	12.04%
Foreign	23	7.69%
Multiple Ethnicities	9	3.01%
Not Specified	4	1.34%
First Generation Student		
First Generation	173	57.86%
Non-First Generation	114	38.13%
First Generation Data Unavailable	12	4.01%

The sample for the Spring 2025 assessment of Critical Thinking has a distribution across various academic levels. Sophomores constituted the largest group at 32.44%, followed closely by juniors at 30.10%. Freshmen represented 19.40% of the sample, while seniors accounted for 17.06%. A small percentage (1.00%) were classified as fifth-year students. In terms of enrollment year, the majority of students began their studies at UT Arlington in either the 2022–2023 (44.15%) or 2021–2022 (22.41%) academic years. An additional 16.72% enrolled most recently in 2023–2024, with smaller groups having started in earlier years (2020–2021: 9.70%, 2019 and prior: 7.02%). Most students (86.29%) were enrolled full-time, while 13.71% were attending part-time. Most of the sample (71.91%) consisted of non-transfer students, with 28.09% having transferred from another institution. This distribution highlights that the assessment sample primarily includes students who are early-to mid-stage in their undergraduate careers, with a strong representation of traditional full-time, non-

transfer students. These characteristics provide important context for interpreting performance trends related to the development of critical thinking skills (see Table 2 for details).

Table 2: *Student Status at UT Arlington*

Categorical Information	N	%
Academic Level		
Sophomore	97	32.44%
Junior	90	30.10%
Freshman	58	19.40%
Senior	51	17.06%
Fifth Year	3	1.00%
Enrollment Year		
2022 - 2023	132	44.15%
2021 - 2022	67	22.41%
2023 - 2024	50	16.72%
2020 - 2021	29	9.70%
2019 and Prior	21	7.02%
Academic Load		
Enrolled Full-Time	258	86.29%
Enrolled Part-Time	41	13.71%
Transfer Student		
Non-Transferred	215	71.91%
Transferred	84	28.09%

The student work assessed for Critical Thinking in Spring 2025 came from a broad range of academic disciplines across UTA.

Table 3: *Students by Colleges/Schools*

College/School	No. of Students	Percentage
College of Architecture, Planning & Public Affairs	104	34.78%
College of Engineering	46	15.38%
College of Science	44	14.72%
College of Nursing & Health Innovation	31	10.37%
College of Liberal Arts	31	10.37%
College of Business	16	5.35%
Division of Student Success	16	5.35%
College of Education	6	2.01%
School of Social Work	5	1.67%

The College of Architecture, Planning & Public Affairs contributed the largest share of student

samples, accounting for 34.78% of the total. This was followed by the College of Engineering (15.38%), College of Science (14.72%), and both the College of Nursing & Health Innovation and College of Liberal Arts, each contributing 10.37%. Smaller representations came from the College of Business and the Division of Student Success, each with 5.35%, followed by the College of Education (2.01%) and the School of Social Work (1.67%) (see table 3 for details).

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 ANTH 1306 Introduction to Anthropology, ANTH 2322 Global Culture, and MATH 1303 Trigonometry courses. 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 (related to students, courses, and instructors) was removed. This anonymization process helped mitigate potential rater bias during the scoring process.

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, the rubric provides a structured framework for evaluating critical thinking proficiency (see Appendix A). The rubric 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 student's 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 for each student 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.

Raters, Rater Calibration, and Scoring

Raters scored the student writing samples during a scheduled scoring day, and each paper was reviewed twice (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 three ratings was used as a final score. The rater group included fourteen 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 rater's 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 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 one 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 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 rater 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.

Among the five rubric dimensions, Influence of Context & Assumptions showed the highest agreement at 91%, indicating strong consistency in how raters assessed students' consideration of contextual factors and underlying assumptions in their analysis. Evidence and Conclusions & Related Outcomes followed closely, with agreement levels of 88% each. Explanation of Issues and Student's Position both achieved agreement rates of 86% (see Table 4 for details).

Table 4: Agreement Percentages Among Raters

Dimension (Critical Thinking VALUE Rubric)	Percentages
Explanation of Issues	86%
Evidence	88%
Influence of Context & Assumptions	91%
Student's Position	86%
Conclusions & Related Outcomes	88%

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, work experience and attended the 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.

To check the consistency level of the inter-rater agreement, the Intraclass Correlation Coefficient (ICC) was computed. 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* = 299
Explanation of Issues	0.55
Evidence	0.54
Influence of Context & Assumptions	0.56
Student's Position	0.62
Conclusions & Related Outcomes	0.54

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.

Table 5 presents the ICC values calculated for each dimension of the Critical Thinking VALUE Rubric based on 299 student work samples. ICC values quantify the reliability or level of agreement among raters in scoring student work. According to the interpretation criteria, ICC values below 0.40 indicate poor agreement, values from 0.40 to 0.74 indicate fair to good agreement, and values above 0.74 represent excellent agreement. In this assessment, ICC values ranged between 0.54 and 0.62 across all rubric dimensions, signifying fair to good agreement among raters. The highest reliability was observed in the dimension "Student's Position" (ICC = 0.62), indicating relatively stronger consensus among raters in evaluating how students articulated and supported their positions. The lowest ICC values were reported for "Evidence" and "Conclusions & Related Outcomes" (ICC = 0.54 each), suggesting slightly less consistency among raters in scoring these aspects of student

critical thinking. Dimensions "Explanation of Issues" and "Influence of Context & Assumptions" had ICC values of 0.55 and 0.56 respectively, demonstrating a moderate level of rater agreement. It is important to note (see Table 5 notes) that ICC values were calculated using a two-way random effects model, which typically produces lower reliability coefficients than models with fixed rater pairings, thus contextualizing the observed ICC scores within expected norms. Overall, the ICC values suggest reasonable levels of inter-rater reliability, affirming the robustness and dependability of the assessment process in evaluating students' critical thinking skills.

Students Performance

The mean scores and standard deviations for each dimension of critical thinking were computer (see Table 6 for details). The assessment included 299 student work samples, evaluated across five rubric dimensions. The institutional attainment target for student performance was set at a mean score of 2.00 or above, representing a benchmark for acceptable proficiency in critical thinking.

Table 6: Means for Critical Thinking Measure Scores

Measurement Dimensions	N	Mean	SD
Explanation of Issues	299	2.14	0.69
Evidence	299	1.64	0.68
Influence of Context and Assumptions	299	1.59	0.63
Student's Position	299	1.54	0.72
Conclusions and Related Outcomes	299	1.61	0.68

Among the dimensions assessed, students successfully met the attainment target only in the dimension "Explanation of Issues," with a mean score of 2.14 (SD = 0.69). This indicates students demonstrated adequate proficiency in clearly identifying and articulating the central issue or problem. However, students did not reach the attainment goal in the remaining four dimensions. The lowest performance was noted in the dimension "Student's Position," with a mean of 1.54 (SD = 0.72), highlighting difficulty in articulating and justifying personal viewpoints. Similarly, the mean scores in dimensions "Influence of Context and Assumptions" (M = 1.59, SD = 0.63), "Evidence" (M = 1.64, SD = 0.68), and "Conclusions and Related Outcomes" (M = 1.61, SD = 0.68) were all below the institutional benchmark. These findings suggest students encountered notable challenges in effectively incorporating contextual awareness, providing robust supporting evidence, and drawing logical conclusions based on their analyses. Overall, the results indicate that while students showed

competence in clearly identifying core issues, substantial improvement is needed in areas related to evaluating evidence, recognizing contextual influences, developing a reasoned position, and drawing coherent conclusions. These insights will serve as critical focal points for instructional enhancement efforts moving forward.

Summary, Observations, and Limitations

This report presented findings from the Spring 2025 assessment of Critical Thinking at UTA. The evaluation involved analyzing written student assignments explicitly designed to assess critical thinking skills across multiple sections of three core curriculum courses: ANTH 1306 (Introduction to Anthropology), ANTH 2322 (Global Culture), and MATH 1303 (Trigonometry). A total of 299 student work samples were scored by faculty and staff raters using the AAC&U Critical Thinking VALUE Rubric.

Students scored highest in the Explanation of Issues dimension (M = 2.14), indicating that they are generally able to identify and articulate the problem or issue being addressed in their assignments with some level of proficiency. All other rubric dimensions—Evidence, Influence of Context and Assumptions, Student's Position, and Conclusions and Related Outcomes—had mean scores below the milestone threshold of 2.00, suggesting that many students are still developing foundational critical thinking skills in these areas. The lowest performance was observed in Student's Position (M = 1.54), which reflects challenges in formulating a clear, well-supported position or thesis.

Despite modest performance outcomes, rater agreement was relatively strong, with percentages ranging from 86% to 91%, indicating consistent scoring among raters. Intraclass Correlation Coefficients (ICCs) ranged from 0.54 to 0.62, placing them in the "fair to good" reliability category. This suggests that while agreement between raters was high, the degree of consistency in exact scoring varied a little across dimensions. While interrater agreement percentages were high, ICCs reflect only moderate reliability. This may be attributed to the use of a two-way random effects model and the randomized pairing of raters, which tends to yield lower ICCs than fixed rater pairings.

The average scores in four of the five rubric categories fell below the milestone level, which may indicate inconsistencies in assignment design, student preparation, or instructional emphasis on critical thinking across courses. As with past assessments, variability in assignment types and

prompts across different colleges and departments may have influenced the distribution of scores, particularly in dimensions requiring nuanced reasoning or thesis development. The data represents a snapshot from one scoring cycle and may not fully capture longitudinal growth or course-level improvements in student critical thinking.

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. (2015, November). Job Outlook 2016. Bethlehem, PA: National Association of Colleges and Employers.

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.

Greenhoot, A. F., & Bernstein, D. J. (2012). Using the VALUE rubrics for improvement of learning and authentic assessment. In T. Rhodes (Ed.), *Assessing outcomes and improving achievement: Tips and tools for using rubrics* (pp. 29–33). Association of American Colleges and Universities.

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/

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2018). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. International Journal of Science and Mathematics Education, 16(3), 527-549.

Shrout, P., & Fleiss, J. L. (1979). Intraclass correlations: 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.