

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

Report on UMass Amherst's Participation in the PILOT

Valid Assessment of Learning in Undergraduate Education (VALUE) Initiative, 2015-2016

In 2015-2016, UMass Amherst participated in the multi-state, multi-institutional VALUE Initiative to assess student learning outcomes. In this report, we provide background on the national project and the decision to participate in it, briefly highlight findings and then explain in more detail the nature of our participation and lessons learned.

Background

The VALUE Initiative is a multi-state project coordinated by the Association of American Colleges and Universities (AAC&U), the State Higher Education Executive Officers (SHEEO), and the Multi-State Collaborative to Advance Quality Student Learning (MSC). As a multi-state, multi-institutional project, it aims to provide the means for a valid assessment of student learning on key aspects of students learning (Critical Thinking, Writing, and Quantitative Analysis) in ways that allow for comparison of performance by student demographics and, in time, sector and state comparisons.

What distinguishes the VALUE Initiative from standardized approaches (e.g. the Collegiate Learning Assessment instrument) is that it:

1. uses actual student work, collected from actual course contexts, as the source of evidence of student performance;
2. uses rubrics developed in 2008-2009 by national teams of faculty for key learning objectives (AAC&U VALUE Rubric Development Project); and
3. emphasizes the formative side of assessment (for example, encouraging campus-based involvement in student learning assessment, conversations about how to better teach these skills, cross disciplinary conversations about teaching).

It is these qualities that led UMass Amherst to participate in 2015-2016, the second year of the MSC VALUE Initiative. Focusing on student performance is a natural progression from the focus on indirect measures of student learning (student surveys/self-reports, satisfaction measures) that have been a feature of our Strategic Planning/Unit Planning efforts so far. The assessment process also echoes the direct assessment process the campus has used to assess student reflective and integrative thinking in the Integrative Experience (IE).

Highlights of UMass Amherst Participation

Per the MSC VALUE Project requirements:

- We focused on upper division—300 or higher—courses and asked for volunteer instructors to submit student work from their courses that they felt met the VALUE rubric Critical Thinking criteria. (See Appendix A for VALUE Critical Thinking Rubric.)
- Our target sample size for the VALUE project was 100 pieces of student work from students who had earned at least 75% of the credits for graduation.
- To select the sample, we used a stratified random sample by course and then checked to ensure the sample of students was generally representative of the race/ethnicity and gender of the UMass Amherst population as a whole. (See Appendix B for sample/population comparisons by selected demographics).

We focused only on Critical Thinking, since that is a learning objective that all Gen Ed courses share, and that is arguably a priority objective for most courses on campus.

In addition to submitting the work for national scoring, we had a team of UMass Amherst faculty score the same work to see how our scoring compared to that of the national instructor population and, equally important, to involve them in evaluating the assessment process and rubric.

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

Key Findings from Participating

- Participating in the MSC VALUE Initiative was worthwhile both for encouraging the campus to participate in an assessment process at the local level and for providing a framework for doing so. More specifically, the common rubric and the process of reading student work from UMass Amherst courses in a cross-disciplinary group opened up opportunities to talk together about definitions of critical thinking, expectations of students, and pedagogy. Certainly, it engaged those instructors in talking about teaching and learning in ways that a standardized test cannot.
- While the process was worthwhile, the results themselves for this first year are less useful in terms of what they mean for overall student performance at UMass Amherst. First, work from 100 students cannot be considered representative of the senior student population at UMass Amherst. Second, there is a lot of “noise” in the data: we learned that the rubric needs fine-tuning, and we need to collect assignments that are better matched to the rubric (that is, less variation in length and nature of the assignments).
- The participating instructors were interested to learn of the results of the national scoring of UMass Amherst student work, but they were wary of publishing institutional results given the provisional nature of the data. They also expressed more general concerns about the validity of comparisons across contexts.

Further Detail on the Assessment Process at UMass Amherst

- 17 faculty across 14 departments submitted student work and the corresponding assignment. The highest concentration of participating courses came from Humanities & Fine Arts (4), with a tie for second highest in Engineering, Isenberg School of Management, and Public Health and Health Sciences (3).
- All told, we collected 386 pieces of student work and randomly selected from among those pieces to identify 132 to submit to the national project. We also selected 100 of those for internal scoring.
- For internal scoring we recruited 12 UMass Amherst faculty and 1 representative each from Student Affairs and the Institute for Teaching Excellence and Faculty Development. All participated in two 2 ½ hour norming sessions to familiarize themselves with the rubric and discuss their assessments of sample papers. The sample papers were drawn from the ones used for the national MSC norming sessions. The scorers then worked independently to score up to 13 student papers apiece, with two readers for each paper. In the case of discrepancies (scores differing by more than one point on two or more criteria), a third reader was used¹. (See Appendix C for more detail on scoring agreements and discrepancies.) The UMass Amherst scoring was done during Summer 2016. (Details on the procedures for the MSC national scoring are provided in AAC&U’s publication *On Solid Ground: VALUE Report 2017*.)

Evaluation of the Assessment Tool and Processes: Selected Findings

In September 2016 when initial results from the national scoring were available, as well as our internal results, we conducted focus groups with the on-campus scorers. The discussion focused on the norming sessions, the scoring process (including the nature of the student work that they scored, and the Critical Thinking rubric), the overall goals of the VALUE Initiative, and applications to teaching and learning in their own classes and departments.

We also interviewed a number of the instructors who provided student work. As part of these interviews, we provided each instructor with two of the student papers from her/his class (one scored high and one low) and discussed that instructor’s evaluation of each in relation to the rubric. This encouraged the instructors to look

¹ An important methodological note: as discussed earlier, for our internal process each paper was read at least twice. The double scoring of work is a recommended practice for qualitative assessment, a practice the national project fully supports. However, due to resource constraints, the national project was only able to score each paper once for reporting purposes.

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

closely at the rubric in relationship to their aims for critical thinking in their courses and their expectations for students.

- Scorers had difficulty because of the wide variety in the student work submitted, from short, low stakes papers to longer semester projects; work based heavily on secondary sources or original research to work based on observation; work that was more reflective or interpretive to work that was more argument focused. (A statistical analysis also shows a correlation of scores to assignment length and number of secondary sources. See Appendix D.)
- “Critical thinking” is a multi-faceted construct the components of which can vary by discipline and assignment. The VALUE Critical Thinking rubric does not encompass all aspects or definitions of critical thinking. Still, the scorers agreed that it does assess generally valued traits of critical thinking and does work for evaluating student work across disciplines although scorers mentioned the difficulty of assessing work in disciplines where they had no expertise (e.g. lack of knowledge of content, scholarship, or specialized analytic methods). This awareness prompted discussion of whether scoring should be based in disciplines or pair a disciplinary expert with an “outside” reader. On balance, scorers still felt it feasible in most instances to assess work according to the rubric criteria, using two readers, and that doing so has the added benefit of prompting cross-disciplinary conversations.
- Also, all involved with the scoring and the instructors with whom we spoke about their submitted work felt that the rubric needs fine-tuning in order to be suitable for a somewhat broader range of student work. There were also suggestions for edits to the rubric to replace some confusing descriptors that posed challenges to scorers.
- While only a small sample of UMass Amherst student work was assessed, it was still interesting to compare our campus scoring with the national results. Overall external and internal scores were fairly similar, although when comparing mean scores our internal scores were a bit lower than those for the external scorers.
- The VALUE Initiative rubric and on-campus assessment process is a useful formative tool that opens up important opportunities for instructors to talk about expectations regarding critical thinking, students’ performance, and teaching for critical thinking. What made the conversations with scorers and instructors who submitted work so generative is that they were considering the rubric in relation to actual student work from UMass Amherst courses.
- Scorers enjoyed the opportunity to talk about critical thinking standards and expectations across disciplines and to see the range of assignments across disciplines as well.
- The discussion about critical thinking, the VALUE rubric, the review of student work from other disciplines, and the consideration of assignments triggered scorers to think about how they define critical thinking, the ways they teach for critical thinking, how effective their assignments are, and what their standards are and should be.
- The process and follow-up made all think more about how to communicate to students about expectations and definitions related to critical thinking. Some shared the rubric with students as a way to help students understand specific aspects of critical thinking.
- The assessment process we followed could be useful not only as a cross-disciplinary project, but also as a departmental one. In fact, in the focus group discussions, some of the scorers mentioned seeing the value of doing so in their departments.

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update
Office of Academic Planning & Assessment (OAPA)

Pilot Student Learning Assessment: Results of Scoring

The results of the national assessment are available in *On Solid Ground: VALUE Report 2017*.

Next Steps

UMass Amherst will continue a second year of the VALUE Initiative, participating in the MSC national assessment and again conduct a corresponding local assessment for Critical Thinking. For this coming round, we are on track to have 21 faculty members across 17 departments submitting student work, about half of whom (10) come from Humanities & Fine Arts, 5 from Social & Behavioral Sciences, and 2 each from Engineering, Public Health/Health Sciences, and Natural Sciences. Based on our experience last year, we are being more precise in describing the nature of the student work to be submitted and the Critical Thinking rubric is being fine-tuned.

Departments have also been invited to participate in a departmental level assessment that uses either the Critical Thinking or the Written Communication rubric.

References:

Association of American Colleges & Universities (AAC&U). (2017). *On Solid Ground: VALUE Report 2017*. Retrieved from <https://www.aacu.org/sites/default/files/files/FINALFORPUBLICATIONRELEASEONSOLIDGROUND.pdf>

_____. (2010). VALUE Rubric Development Project. Retrieved from <https://www.aacu.org/value/rubrics>

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update
Office of Academic Planning & Assessment (OAPA)

Appendix A. VALUE Critical Thinking Rubric

CRITICAL THINKING VALUE RUBRIC

for more information, please contact valu@aaau.org



The VALUE rubrics were developed by teams of faculty experts representing colleges and universities across the United States through a process that examined many existing campus rubrics and related documents for each learning outcome and incorporated additional feedback from faculty. The rubrics articulate fundamental criteria for each learning outcome, with performance descriptors demonstrating progressively more sophisticated levels of attainment. The rubrics are intended for institutional-level use in evaluating and discussing student learning, not for grading. The core expectations articulated in all 15 of the VALUE rubrics can and should be translated into the language of individual campuses, disciplines, and even courses. The utility of the VALUE rubrics is to position learning at all undergraduate levels within a basic framework of expectations such that evidence of learning can be shared nationally through a common dialog and understanding of student success.

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.

Framing Language

This rubric is designed to be transdisciplinary, reflecting the recognition that success in all disciplines requires habits of inquiry and analysis that share common attributes. Further, research suggests that successful critical thinkers from all disciplines increasingly need to be able to apply those habits in various and changing situations encountered in all walks of life.

This rubric is designed for use with many different types of assignments and the suggestions here are not an exhaustive list of possibilities. Critical thinking can be demonstrated in assignments that require students to complete analyses of text, data, or issues. Assignments that cut across presentation mode might be especially useful in some fields. If insight into the process components of critical thinking (e.g., how information sources were evaluated regardless of whether they were included in the product) is important, assignments focused on student reflection might be especially illuminating.

Glossary

The definitions that follow were developed to clarify terms and concepts used in this rubric only.

- **Ambiguity:** Information that may be interpreted in more than one way.
- **Assumptions:** Ideas, conditions, or beliefs (often implicit or unstated) that are "taken for granted or accepted as true without proof." (quoted from www.dictionary.reference.com/browse/assumptions)
- **Context:** The historical, ethical, political, cultural, environmental, or circumstantial settings or conditions that influence and complicate the consideration of any issues, ideas, artifacts, and events.
- **Literal meaning:** Interpretation of information exactly as stated. For example, "she was green with envy" would be interpreted to mean that her skin was green.
- **Metaphor:** Information that is (intended to be) interpreted in a non-literal way. For example, "she was green with envy" is intended to convey an intensity of emotion, not a skin color.

CRITICAL THINKING VALUE RUBRIC

for more information, please contact valu@aaau.org



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 (all one) level performance.

	Capstone 4	Milestones 3	2	Benchmark 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 <i>Selecting and using information to investigate a point of view or conclusion</i>	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.

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

Appendix B: Campus Representativeness

Our sample size is too small to make any claims about representativeness of the data, and we urge caution in using these results to generalize about learning on campus. This appendix is primarily to demonstrate the distribution of our sample across Colleges and Departments, as well as to compare the demographics of our sample to those of the undergraduate population. These demographics are for the final sample submitted for external scoring (n=132) and includes students from courses using group work.

Table 3. Select Comparison Demographic \geq 75% of Credits for Graduation Earned: Comparison of student with submitted work sample and UMass Amherst students as a whole.

School/College	Sample	UMass	URM	Sample	UMass
Education	0%	0.3%	URM	19.2	11.4
PHHS	17.5	10.0	Asian	9.2	11.4
CNS	9.9	28.5	International	5.8	3.0
HFA	18.3	9.0	White	65.8	67.5
Engineering	12.2	10.0	PELL Eligible		
CICS	0	3.9	Yes	33.6	21.9
SBS	14.5	19.0	No	66.4	77.9
ISOM	23.7	17.0	Transfer		
Nursing	0	1.8	Transfer	22.9	24.1
Other-ROTC	3.8	1.5	Frosh	77.1	75.7
Gender			First Gen		
Female	55.0	46.9	Yes	34.4	26.0
Male	45.0	53.1	No	65.6	73.7

Table 4. Distribution of the sample across School/College and Department

School/College	School/College Represented in sample*	# Departments in School/College	# Departments in sample
College of Education		3	0
School of Public Health and Health Sciences	*	7	2
College of Natural Sciences	*	15	1
College of Humanities and Fine Arts	*	14	3
College of Engineering	*	5	2
College of Information and Computer Sciences		1	0
College of Social and Behavioral Sciences	*	15	2
Isenberg School of Management	*	7	3
College of Nursing		1	0
Other--ROTC	*	1	1

*indicates participation from this School/College

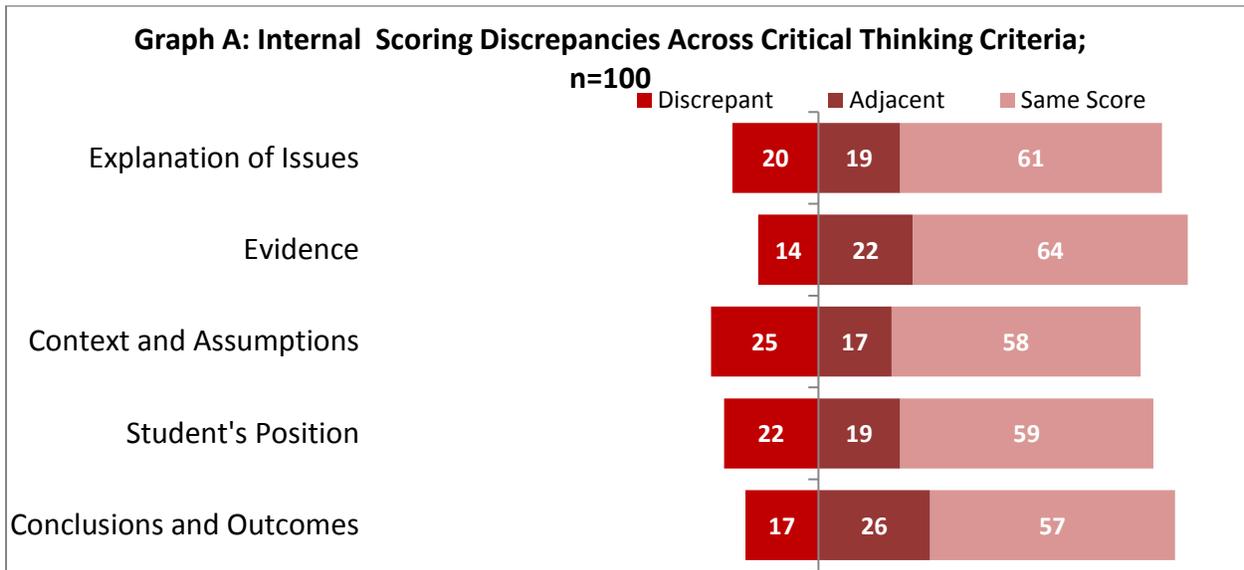
UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

Appendix C: UMass Internal Scoring Reliability

There were two UMass Amherst readers for each paper. Scores on each criterion were considered acceptable if they were the same or differed by only one level (for example, a score of 2 and 3). When the two readers differed by two or more levels on the rubric (for example, a 2 and a 4), their scores were considered “discrepant.” When a paper received discrepant scores on two or more criteria for a given paper, a third reader was used. In all cases, whether a paper was scored by two or three readers, the reader scores were averaged to yield the final score for the paper.

As Graph A illustrates, just under two two-thirds of all scores were the same, and about 20 percent of the scores on each criterion were “discrepant” using our definitions. The number of discrepant scores was somewhat higher for the Context and Assumptions criterion, supporting the scorers’ view that this criterion was among the rubric criteria needing revision.



Of the 100 papers scored, 28 required a third reading; that is, the initial two readers’ scores differed by 2 or more score points on at least 2 of the traits.

UMass Amherst Multi-States Collaborative (MSC) PILOT Participation Update

Office of Academic Planning & Assessment (OAPA)

Appendix D: Assignment Alignment

The scoring process revealed that two paper characteristics had a relationship with average score: length and use of external sources.

Table 1 provides descriptive information about these two variables for the 100 papers that were scored by both national and internal scorers.

Table 1. Descriptives for each group of papers (N=100)

	Overall Average	Overall Median	Overall Range
Number of Pages	14.4	9	2-102
Number of References	9.92	8	0-67

(NOTE: This includes both the Group Work assignments and the Single Author assignments)

Table 2 shows that there are moderately strong positive correlations between a paper's page length and number of references and the paper's overall external and internal scores. Longer papers and those using external sources tended to score higher on the rubric's Critical Thinking criteria than papers not sharing those characteristics.

Table 2. Correlation values for Papers Scored by National Group and Locally (n=100)

	National Average Score	Local Average Score
Number of Pages	.427**	.426**
Number of References	.478**	.454**

***Correlation is significant at the .01 level*