

Critical Thinking General Education Assessment: AY 2019-2020

Background: General Education at Community College of Philadelphia

The definition of and rubric for Critical Thinking was created by the all-faculty Critical Thinking subcommittee of the General Education Task Force in 2010. It was based on the then-current Middle States Commission for Higher Education (MSCHE) Standards for Affiliation. In order to fulfill the Critical Thinking requirement and earn an associate degree, each Community College of Philadelphia (College) student must take a course tagged as Interpretive Studies in the College catalog. Faculty in 2010 created a rubric that outlined six dimensions of critical thinking. Each dimension included a brief description “Beginning,” “Developing,” “Competent,” and “Accomplished” student work.

Between 2010 and 2019 Critical Thinking was directly and indirectly assessed multiple times. Past evidence indicated that faculty and students believed that students’ critical thinking skills improved as they took courses at the College. Other evidence pointed to the need to clearly state critical thinking outcomes in Interpretive Studies course syllabi.

Assessment and analysis of General Education outcomes at the College is performed and reported by the Office of Assessment and Evaluation (OAE). Once the data is analyzed and a report is written, the General Education report is shared with various groups. OAE then edits the report and recommendations before it is finalized and released publicly.

Faculty, staff, and administrators collaboratively discuss data, analysis, and recommendations at each semester’s Assessment Tuesday. Edits, recommendations, and “loop closing” ideas are added to the report. The finalized General Education Assessment report is then posted on the College website, and a link circulated to the entire College community. Monitors throughout campus display results and tips to students, staff, and faculty as well.

The 2019 Information Literacy assessment included a recommendation to implement a plan to address students’ continued deficiency in documenting sources. Because the Critical Thinking rubric contained a dimension on sources, this was an opportunity to facilitate that discussion. Further, informally collected evidence by OAE staff pointed to the need for assessment-focused collaborative professional development in order to fully implement assessment as a College policy.^{1 2}

¹ “Information Literacy General Education Assessment,” Community College of Philadelphia, Spring 2019. Staff observations and faculty and administrative comments during Academic Program Reviews comprise the informal evidence.

² “Research Brief VI: Changing Teacher Beliefs and Instructional Practices,” Center for Leadership in Education, University of Washington.

To assess critical thinking in 2019, OAE utilized a mixed-methods “assessment summit” originally developed in 2010 by the General Education Assessment Committee at Old Dominion University.³ This method allowed the College to collect evidence and data required for accountability, assess students’ Critical Thinking skills, and create space for collaborative “loop closing.” This new-to-the-College method had unfortunate trade-off of being unable to directly compare results with previous assessments. Staff decided the imperative for collaboration justified the change.

Research Questions

- Does one Interpretive Studies Course equate to student learning in the area of Critical Thinking?
- How have student outcomes in critical thinking changed over time?
- To what extent does the faculty body have a general agreement on what constitutes competent critical thinking by students?
- Does a sample of student work adequately assess Critical Thinking across the College?
- What additional information can collaborative outcomes assessment yield?

Theoretical Framework

This Critical Thinking outcomes assessment was planned and implemented at the intersection of three research areas:

A. Mixed methods research

Mixed-methods research combines qualitative and quantitative data and analyses in order to combine rigorous and standardized quantitative data analyses with a pattern-oriented qualitative analysis.⁴

B. Defining, teaching, and assessing critical thinking

The definition of Critical Thinking, and the rubric used in the assessment were created by faculty during the 2010 General Education revision.⁵ Additional research on employers and critical thinking and lack of consensus in assessing critical thinking was consulted in the design of the assessment, but was not brought into the rating session in order to avoid biasing the participants.⁶

³ " Old Dominion University, General Education Assessment," accessed June 11, 2020, <https://www.odu.edu/facultystaff/assessreport/assessment/competencies>

⁴Secolsky, Charles, and D. Brian Denison. 2018. *Handbook on measurement, assessment, and evaluation in higher education*.

⁵ <https://www.mycpp.online/office-assessment-and-evaluation/institution-level-assessment>

⁶ AACU, “Key Findings from Employer Research.” 2018.

C. Assessment as cross-disciplinary collaboration and professional development

The rating summit method demonstrates how collaboratively assessing General Education outcomes can foster cross-disciplinary changes practice or policy in response to assessment results.⁷ In this model, cross-disciplinary discussions, disagreement, and eventual consensus—or uneasy but respectful compromise—are crucial processes in faculty professional development.⁸

Methodology

In Fall 2019 there were 198 sections of courses tagged as “Interpretive Studies. In order to achieve a 95% level of confidence that the sampled student work products were representative of the total body of student work in courses labeled “Interpretive Studies,” at least 20 student work products from 25 sections each were required. Staff used a random number generator to select 25 sections. Out of the 25, three faculty had two of their sections selected. One section for each of them was eliminated so as not to overburden or over-represent any particular faculty member or discipline. Faculty members were informed of their selection via email. They were asked to submit a copy of a critical thinking assignment by September 30, 2019 and to submit their students’ work (artifacts) on the assignment by the end of the Fall 2019 grading period.

All randomly selected faculty members supplied OAE with student work, for a total of 417 artifacts. These were numbered sequentially. Using the same random number generator, 125 artifacts were selected for use at the rating summit. Staff removed names, course and section numbers, and other identifying details. The anonymized artifacts were duplicated. They were then organized into packets of eight artifacts for rating. Every rater would receive eight artifacts. Each artifact would be scored by two different raters. Of the remaining 292 artifacts, three were chosen for the collaborative rubric norming session.

On the January 5, 2020—Assessment Tuesday, 28 faculty members participated as raters in the rubric norming session. This session began with a brief explanation of the purpose and nature of the assessment summit methodology. A standard rubric norming and calibration procedure was followed.⁹ Raters were asked to individually rate an artifact on the first dimension of the Critical Thinking rubric. The ratings were publically tallied and raters shared their justifications for the tallies. Discussion and consensus-building followed. One OAE staff member facilitated the discussion while the other took notes. When participants were comfortable moving to the next dimension, comments were summarized back. Raters then applied the next dimension of

⁷ Jankowski, “Closing the Loop: Using Assessment Results to Enhance Student Learning.”

⁸ Hill, L., Kim, S. L., & Lagueux, “Faculty Collaboration as Professional Development.”

⁹ Kevin Schoepp, Maurice Danaher, and Ashley Ater Kranov, “An Effective Rubric Norming Process,” *Practical Assessment, Research, and Evaluation* 23 (2018): |PAGE|, doi:<https://doi.org/10.7275/erf8-ca22>

the rubric to the same artifact. This was followed by another facilitated discussion. This procedure was repeated for the remainder of the dimensions. After two hours and two more artifacts, some consensus in interpretation of the rubric was achieved. Time constraints required that the discussion end.

14 participants left during a mid-day break. The remaining 14 raters were given sets of 8 artifacts. Some raters also agreed to rate artifacts after the conclusion of the session, resulting in a total of 112 usable artifact ratings.

Data Collection

Qualitative data was collected from the following sources:

- email and in-person conversations
- small group and whole-group discussions at the ratings summit
- written comments on the student artifact ratings forms
- comments from feedback forms

These data amounted to approximately 100 discrete comments. Comments that could be personally identifiable were eliminated. The rest were then categorized by content.¹⁰ The categories that emerged were:

1. References to an individual disciplinary areas or discipline-specific arguments and assertions
2. Professed lack of trust of the goal of current assessment practices at the College.
3. Professed lack of understanding of the goal of current assessment practices at the College
4. Concern that students' privacy was being violated in this process.
5. Concern that faculty members' privacy was being violated, or that this process was otherwise unfair to faculty.
6. Arguments with the methodology of this Critical Thinking assessment
7. Comments and questions about sources
8. Comments and questions about citations
9. Comments about students' writing separate from the content of the work sample
10. Comments about classroom practice
11. Disagreement that the work sample and assignment indicated that the assignment adequately assessed critical thinking
12. Disagreement over what supports are appropriate for students

These categories were then applied to the original research question they addressed or to a catchall category of "data that answers questions we didn't ask."¹¹ Each research question had

¹⁰ Smith Tuhiwai, *Decolonizing Methodologies: Research and Indigenous Peoples*.

¹¹ Erickson, "Qualitative Methods on Research in Teaching."

at least ten comments that helped to answer it. Representative comments have been inserted in the analysis.¹² The catchall was further divided into “Further Professional Development Opportunities,” and “Areas for Organizational Culture Change.”

Quantitative data was collected from 112 artifact ratings. Mean scores for all elements was below 3 (Competent). There were no statistically significant differences between group means as determined by one-way ANOVA ($F(4,479) = .182, p = .95$). Percent agreement was calculated between each pair of artifact ratings; average percent agreement for the 14 participants was 25.4%.

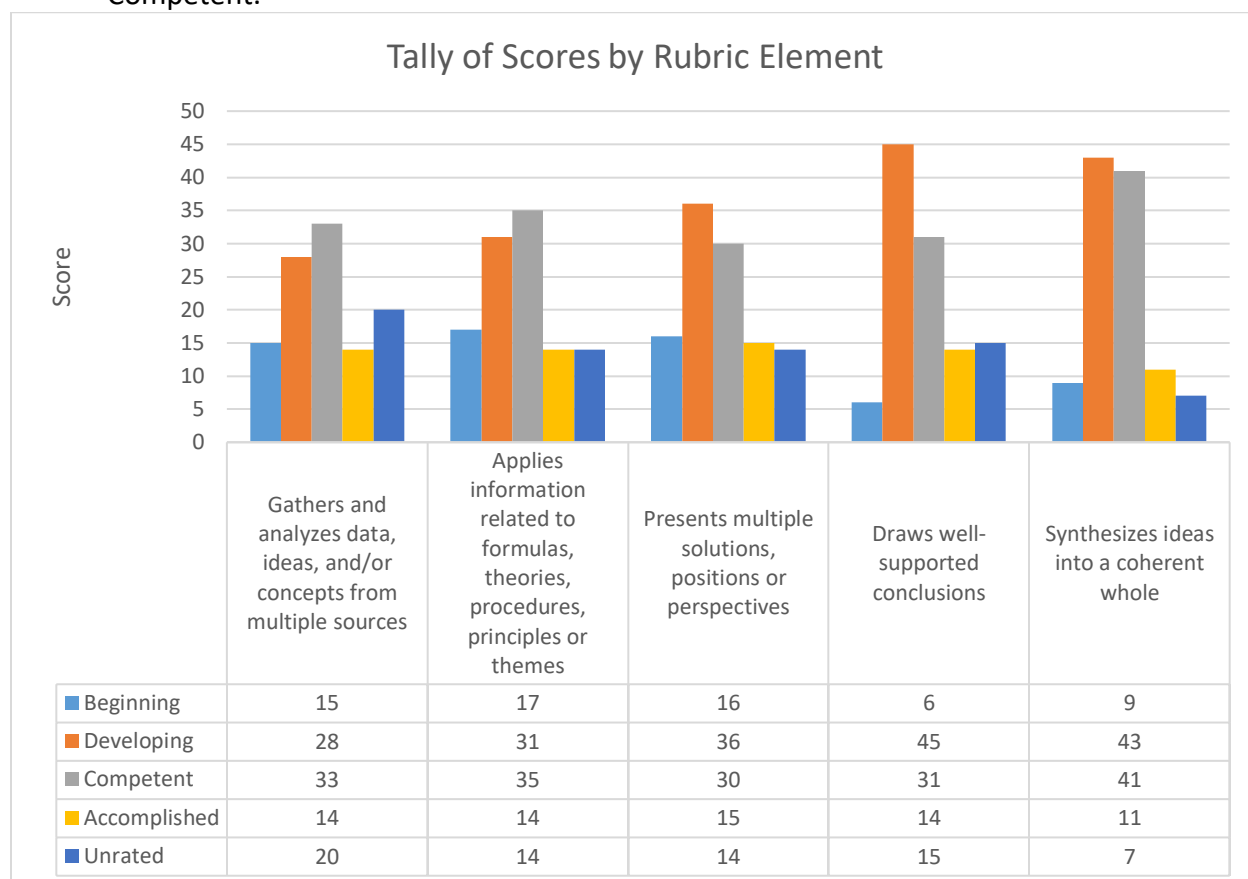
	Mean Score N = 112	Standard Deviation
Student gathers and analyzes data, ideas, and/or concepts from multiple sources.	2.51	0.95
Student applies information related to formulas, theories, procedures, principles or themes.	2.47	0.95
Student presents multiple solutions, positions, or perspectives.	2.45	0.94
Student draws well-supported conclusions.	2.33	0.82
Student synthesizes ideas into a coherent whole.	2.52	0.80

¹² Miles, Matthew B. Huberman, A. Michael. Saldana, *Qualitative Analysis: A Methods Sourcebook*.

Results

Does one Interpretive Studies Course equate to student learning in the area of Critical Thinking?

No, with caveats. The quantitative data collected during the rating summit was drawn from a double-blind, representative cluster sample of written student work completed in courses labeled “Interpretive Studies.” That work was rated, on average, below Competent.



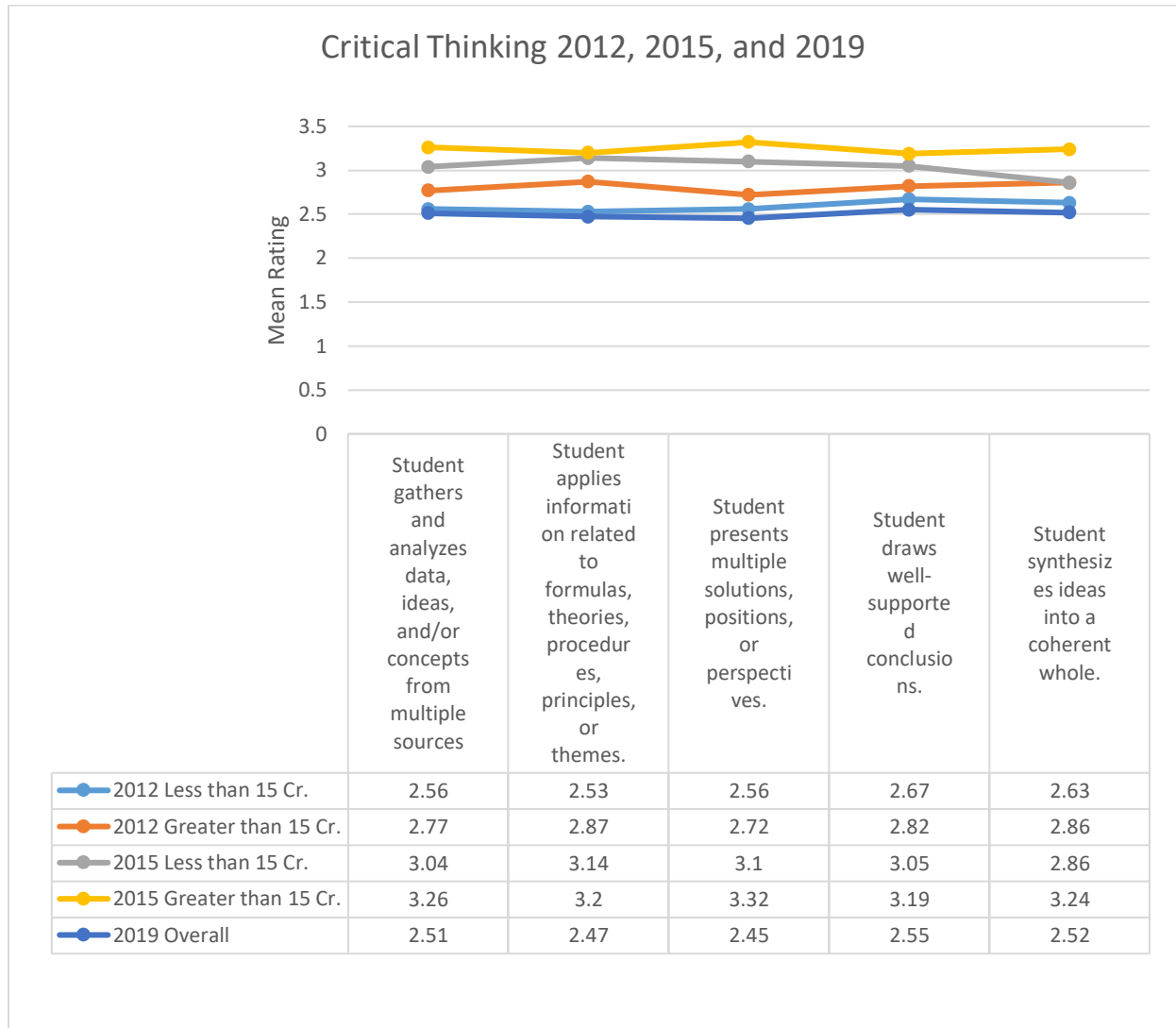
The addition of qualitative data offers a different lens. The range and tenor of comments indicate that CCP faculty differ on granular elements of the rubric, for example what constitutes a source or what constitutes “well supported conclusions.”

How have student outcomes in critical thinking changed since the last two Critical Thinking assessments?

A direct comparison of mean ratings from 2012 and 2015 to 2019 is complicated by the differences in assessment methodology between various cycles. In 2012 and 2015, faculty were asked to apply the Critical Thinking rubric to their overall impression of a sample of students and their bodies of work during their classes. The double-blind

artifact rating methodology used in the 2019 assessment may have corrected for potential bias from faculty rating their own students, but may have introduced additional difficulties in determining ratings by limiting the application of the rubric to one artifact per student.

As such, while the mean ratings for all elements of the Critical Thinking rubric fell below those in 2012 and 2015, it is not possible to attribute that decline solely to a decline in student or institutional performance in the attainment of competent Critical Thinking skills.



To what extent does the faculty body have a general agreement on what constitutes competent critical thinking by students?

To a low extent. Twenty-five percent of faculty members submitting an assignment and student work expressed the concern or expectation that their colleagues would disagree that the assignment assessed critical thinking. These included concerns over “submitting

one piece of a larger project,” that “every assignment in the course assesses critical thinking,” and that out-of-field colleagues would not recognize field-specific critical thinking.

These points were also expressed among the 28 colleagues at the rating summit, either in the norming discussion, written on the rating forms, or written on session feedback forms. One member noted concerns that it was unfair to assess the work of students taught by a colleague; a few asserted that they felt uncomfortable assessing a piece of a project; many written comments disagreed with or doubted whether the student work products assignments actually captured critical thinking.

Quantitative data supports this conclusion. The average percent agreement for the 14 artifact raters was 25.4%, indicating that, even after participating in a rubric-norming session and extensive discussion during the early part of the rating summit, raters only agreed on the scores for their artifacts on an average of one in four ratings. Likewise, standard deviations of nearly a full point for each element of the Critical Thinking rubric indicate that raters tended to differ by about one point in either direction. Because raters and artifacts were deliberately anonymized, it is not possible to analyze whether this discrepancy occurred more often in cross-disciplinary pairs. Pairing this information with the discomfort expressed by faculty in verbal comments, however, it is possible to deduce that faculty seem to have a breadth of views on what constitutes evidence of competent critical thinking by students.

Does a sample of student work adequately assess Critical Thinking across the College?

Faculty members do not believe so. Despite following and sharing openly the professionally-recognized assessment and mixed-methods practices OAE followed in planning and implementing this assessment, qualitative comments disagreed with the assessment model.

The results of a one-way ANOVA ($F(4,479) = .182, p = .95$) indicated no statistically significant differences between the means for each of the elements of the rubric. There are many possible explanations for this.

Additional Conclusions

Areas for Further Collaborative Professional Development

The language of the Critical Thinking rubric in combination with the results of the Information Literacy ratings summit yielded a discussion during the rubric norming session. In Spring 2019 CCP students had taken the Standardized Assessment of Information Literacy (SAILS). Their lowest score had been in the Citations category.

The Critical Thinking Rubric required that a student “Gathers and analyzes data, ideas, and/or concepts from multiple sources.” At the summit the rubric norming discussion on this rubric element centered around two points. Faculty learned that they have different expectations of competent standards of requiring, teaching, and assessing citations. Faculty also discussed how to apply the “multiple sources” requirement in the source dimension of the rubric. They found that source and citation requirements differed across the College.

The conversation was an important step in the assessment cycle by having faculty share with each other their expectations and departmental or divisional policies. It can be followed by concrete action—such as creating and putting in syllabi a division- or department-wide citation policy on citations—as an act of “Closing the Loop.”

Numerous participants expressed discomfort in rating student’s work that was assigned by another faculty member. This indicates that multi-reader grading or rating is unfamiliar or uncomfortable for the faculty who made these comments. Given that multi-reader rating is a staple of nearly all standardized written assessments, from the A.P. exams to this college’s own English department’s exams, this may indicate a potential area of opportunity for faculty growth.

Areas for Organizational Culture Change

Comments from faculty that expressed disdain (as opposed to academically-appropriate critique) for their colleagues’ work, fear of such disdain, comments that directly or indirectly addressed past labor complaints, discomfort with rating student work assigned by a colleague, and comments that reflect practices and policies—but not rules—at the College, were grouped under “**Organizational Culture.**”

Indirect Assessment

In 2019 Institutional Research administered the Community College Survey of Student Engagement (CCSSE). Students were asked several questions that indirectly measured their experience in practicing the critical thinking skills from the College rubric. This chart shows the percentage of students who responded “Very Often” on the four-point Likert scale.

	CCP 2016	CCP 2019	National CCSSE Cohort 2019
Worked on a paper or project that required integrating ideas or information from various sources	59.9%	63.5%	66.9%
Analyzing the basic elements of an idea,	73.2%	74.1%	69.9%

experience, or theory			
Applying theories or concepts to practical problems or new solutions	54.2%	56.9%	56.2%

Recommendations

While this work was occurring, so was a second important collaboration. The faculty General Education Task Force (GETF) was revising the College's General Education standards based on revised MSCHE Standards for Affiliation. These were approved in Spring 2020 and echo MSCHE Standard III.¹³ There is no standalone Critical Thinking component of General Education at CCP. This report makes no recommendation regarding the content of the retired standard or rubric.

1. Create consensus and document where critical thinking will be taught in the College curricula

Action: Beginning Summer 2020 and before Essential Skills are implemented in Fall 2021, faculty should identify within Essential Skills rubrics or other curricular documentation where and how all elements of critical thinking will be taught and assessed. This report has shown that faculty do not believe the former curricular model, in which critical thinking was embedded in Interpretive Studies courses, adequately demonstrated student outcomes in critical thinking. The new model should explicitly address this.

Who should do this: Either GETF or Program faculty members writing Essential Skills rubrics or submitting course revision documents. Committees and Program faculty should hold each other accountable. OAE should be available to assist.

Who will track this: The Vice President of Academic and Student Success should designate the appropriate unit or persons to track this. After the Essential Skills are implemented, department heads should track integrity of implementation through syllabi checks or other methods they deem appropriate.

¹³ Standard III. <https://www.msche.org/standards/>

2. **Research and implement transparent ways to incentivize faculty who create and facilitate critical thinking/competence, sourcing, and citations workshops during upcoming Professional Development Weeks.**

With faculty pulled in many different directions, some local research should be conducted to ascertain what additional incentives or returns on investment (ROI) would compel more faculty to participate in creating and/or facilitating professional development sessions in response to student learning outcomes.

Action: Uncover 3-4 implementable “returns on investment,” that will encourage faculty to create and facilitate high quality outcomes-focused professional development on critical thinking or sources and citation for their peers.

Who should do this: Institutional Research will research the question and report results to the Cabinet for discussion and implementation.

Who will track this: Institutional Research

3. **Focus use of contracted Professional Development time on cross-disciplinary collaborative analysis of General Education student outcomes and creation of action plans.**

Faculty contracts include three days of professional development at the start of one semester and four at the start of the other each academic year. Attendance is not taken. Assessment Tuesday has been set aside for assessment-focused development. Academic deans, department heads, faculty members, IR, and OAE should collaboratively plan Assessment Tuesday in order to capitalize on this paid time.

Action: Increase faculty attendance at Assessment Tuesday by 50% from the baseline of Spring 2020 by Fall 2022.

Who should do this: OAE staff should coordinate the effort among their office, IR, department heads, deans, and faculty facilitators.

Who will track this: The Office of Professional Development

4. **The College community must align to make positive, research- and experience-based changes to better serve our students.**

These data indicate that a shift in organizational culture of the College is necessary to create consensus in assessment and accountability.

Action: By June 2021, 80% of academic and administrative units will report an, outcomes-informed research- or experience-based change that they made to their work. The AES or AUR (Department Head) reports should have a question to that effect.

The report will also include all other required elements. The unit should track the effectiveness of the change in the same manner as all other goals.

Who should do this: The Cabinet or IEC should determine this

Who will track this: These efforts will be tracked by AES and AUR Department Head reports by default.

5) Curate an outcomes action resource list

Action: By December 2020 OAE and library faculty will curate and release a list of research resources faculty and staff can read, apply, and cite in their AES and AUR. The resource list should be searchable by common outcomes, and learning taxonomy verb, so that faculty or staff looking for a solution to a specific problem are guided to 2-3 non-specialist resources on the subject. Resources should include relevant research and practice from K-12, as many current higher education accountability practices were drawn from there and some research is appropriately transferable to a community college setting.

Who: OAE and Library, in consultation with the GETF.

Who will track this: OAE

Appendix: Critical Thinking Rubric

Critical Thinking Rubric

Skills	Beginning	Developing	Competent	Accomplished
Gathers and analyzes data, ideas, and/or concepts from multiple sources	Copies information, often inaccurately, incompletely, or omits relevant information Uses only one source	Reports information - paraphrases. May have minor inaccuracies, irrelevancies or omissions	Presents information accurately and appropriately. Uses multiple sources.	Blends multiple sources. Interprets accurately, appropriately and in depth in new contexts.
Applies information related to formulas, theories, procedures, principles or themes	Labels formulas, theories, procedures, principles or themes inappropriately, inaccurately or omits them	Uses appropriate formulas, theories, procedures, principles or themes with minor inaccuracies	Applies formulas, theories, procedures, principles or themes appropriately and accurately in familiar context	Employs formulas, theories, procedures, principles or themes accurately, appropriately and or creatively in new contexts
Presents multiple solutions, positions or perspectives	Identifies a single solution or resolution or fails to present one	Identifies simple solutions or perspectives with minor inaccuracies	Describes two or more solutions, resolutions, positions or perspectives accurately	Explains accurately and thoroughly, multiple solutions, resolutions, positions, or perspectives that balance opposing views of an issue
Draws well-supported conclusions	Proposes a conclusion or resolution that is inconsistent with evidence, illogical or omits conclusion altogether	Offers an abbreviated or overly simple conclusion that is mostly consistent with evidence and has minor inaccuracies	Organizes a conclusion that is complete, logical and consistent with evidence	Creates an independent judgment that is reflected in the conclusion or solution. Well supported by evidence and logic.
Synthesizes ideas into a coherent whole	Lists ideas or expresses solutions in a disjointed manner, no order	Uses a simple pattern to organize solutions	Connects ideas or develops solutions in a clear and orderly manner	Integrates ideas into solutions that are clear and cohesive.
Participates in the self reflection/ assessment process	Unable to identify major strengths and weaknesses in work Does not seek and/or resists feedback on work	Attempts to identify strengths and weaknesses in work Accepts feedback to improve work	Identifies strengths and weaknesses in work Seeks assistance when needed to improve work	Self-identifies strengths /weaknesses in work and makes efforts to improve Uses feedback to increase self awareness, improve overall research methods, and enhance learning