Chapter 10. Artifacts in ePortfolios: Moving from a Repository of Assessment to Linkages for Learning

Howard B. Sanborn
Virginia Military Institute

Jenny Ramirez
Mary Baldwin University/James Madison University

As a collection of ***artifacts*** presented and curated by students in the digital space, ePortfolios offer students a means to trace the building of their knowledge and ***showcase*** their talents and abilities. They permit both students and instructors to explore new types of assignments that can be submitted as evidence of student learning over time. Yet, as willing as many students are to try their hand at assignments beyond the typical essay, they often do not have an understanding of how to draw out sophisticated observations about their own learning that take place in and between assignments.

In this chapter, we explore the linkages across ePortfolio artifacts made possible by explicit reflection at discrete moments in the semester. In particular, we demonstrate how the guidance of humanities professors who require assignments with implicit reflection, such as artist statements and the design and creation of artistic pieces, can help professors in the social sciences gain insights into how to teach reflection about other cultures to students. Structured reflection can encourage students to think deeply about their work on a specific artifact. More importantly, it can allow them to more easily connect their artifacts together into a sophisticated narrative about their learning. A statistical data analysis illustrates the improvements students made in analyzing ***artifacts*** in summative reflective essays that they included in their ePortfolios. As students were asked to reflect alongside their artifacts, they gained the skills that allowed them to more competently evaluate the appropriateness of various types of assignments to their learning.

Background and Literature

ePortfolios serve many purposes, from the ***assessment*** of student learning and the marketing of skills to an employer to the showcase of student work and the development of learning in a course or program (Barrett, 2007). In particular, there is great traction gained from the use of ePortfolios in assessment *for* learning (Black & Wiliam, 1998) (see Coleman et al., this collection). Students are not only

DOI: https://doi.org/10.37514/PRA-B.2020.1084.2.10

afforded the space for displaying their work but also given the opportunity to form their ideas once they see an artifact uploaded to their ePortfolio. They may even feel greater facility in solving larger, substantive questions about coursework when given the time to craft an answer while, at the same time, not competing with other students to speak during class time (Black & Wiliam, 2004). These efforts are an attempt to promote learning that "should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas" (Black & Wiliam, 1998, p. 8). ePortfolios, then, are a means to create this space for exploration and reflection, both at discrete moments during the semester and at the culmination of a project or course.

Dellinger et al. (2013) detail the importance of Schön's (1983) reflection-in-action in a study of reflective captions, referred to in their piece as "reflective tags," which students submitted along with their artifacts in an ePortfolio. Explicit contextualization provided students an increased ability to reflect on the lessons learned about cultures and groups. At the end of the semester, students who had curated their ePortfolios with these tags also tended to perform better in their summative assessment: a reflective essay in which they evaluated their processes of learning.

Individuals can engage in increasing levels of reflection, ranging from the simple contrast of what one once knew and what one now knows to the textured and complex insights of an individual weaving empathy, insight, and knowledge into a deeper understanding of a particular topic. Indeed, there have been studies of the quality of reflection in ePortfolios (Dellinger et al., 2013; Parkes et al., 2013), but the evaluation of reflection atomized to the artifact-level is often difficult to conceptualize and measure.

Student artifacts are the elemental components of the ePortfolio. The digital space allows for greater experimentation in the composition of assignments, as well as a means to draw connections that demonstrate learning as it happens (Bhattacharya & Hartnett, 2007). Yet, the practice of effective ***reflection*** can be a challenge for students. Faculty do not often recognize the frustration their students have in learning how to learn (see Polly et al., this collection). They fail to provide them guidance on how to reflect that is tailored to the goals of a particular ePortfolio (Landis et al., 2015). As a result, it becomes easy for students to turn their ePortfolios into digital repositories, akin to a collection of links one might find in Dropbox or Google Drive, rather than spaces for reflection and synthesis of their work.

Another related pitfall is the competing goal of implementing both formative and summative assessments in student coursework (Barrett & Carney, 2005). ePortfolios should be a collection of "unique linkages, connections, and reflections among multiple experiences and artifacts in ways that would not otherwise be possible with a traditional paper portfolio" (Parkes et al., 2013, p. 101). Yet, many professors ask students to produce a final product for the class that can be

used as a means to evaluate whether one has met program leaning outcomes. The linking of artifacts provides a thread that holds the ePortfolio together, but this attempt to scaffold reflection throughout a course can be lost in the desire for a more straightforward assessment.

In sum, ePortfolios can offer students an effective means to synthesize the lessons of a course or program by not only asking them to showcase their work but also to draw out the connections between artifacts in an effort to reinforce their learning and, perhaps, even spur on the creation of knowledge. Unfortunately, students are not often prepared to link ePortfolio artifacts together in a manner that encourages the reflection that results in deep learning desired by professors. As a result, professors must make efforts to incentivize reflection from students as they submit artifacts so as to draw out the "linkages, connections, and reflections" that will cultivate a richer engagement with the lessons from the classroom.

Case Study: Artifact Analysis in the Study of Asia

Students were asked to take two classes as part of a general education requirement (see Terry & Whillock and Carpenter & Labissiere, this collection) to teach them about the "civilizations and cultures" (C&C) of the world. For each class, their guidance was the same; they were to upload artifacts, tagged with reflective annotations, to their ePortfolios alongside a summative, reflective essay. This reflective essay was subsequently assessed by a team of faculty for student achievement in three categories: Cultural Understanding, Reflection, and Artifact Analysis.[1]

Two of the classes included in the C&C program were drawn from the Department of International Studies and Political Science: Politics in East Asia and Politics in China. These two courses were junior-level seminars without pre-requisites and were open to students of all majors. The subject matter of the two courses centered on government institutions and political processes. In the East Asia course, these elements were compared across Japan, China, Taiwan, and the Koreas. In the China course, students learned about the post-Mao reforms that took place during the 1980s and how the authoritarian government led by the Communist Party has evolved since it came to power in 1949.

Initially, students were asked to complete response essays as part of the requirements for these classes, in line with the typical assignments for a regional political science survey course. In these essays, students were asked to relate current events to the lessons of the class. These assignments formed the bulk of the artifacts from which a student could choose for their ePortfolio and write about in their summative reflective essay. However, even though students performed well in lessons about government and politics, their assessment scores in cultural understanding and reflection consistently fell below benchmark, indicating that

1. The first assessment instrument was a ***holistic rubric***, which was later replaced by this three-category scale.

they did not seem to adequately grasp lessons about culture that prompted the creation of the civilizations and cultures component of the core curriculum. This poor performance was seen in C&C classes drawn from across the curriculum, particularly in the natural and social sciences, as well as in engineering courses.

Consequently, in 2012, professors from humanities courses held workshops designed to promote the functionality of ePortfolios, particularly the ability to create projects that were not possible outside of the digital space. In both the Politics in East Asia and Politics in China courses, the professor introduced one such project: a propaganda poster. Students could be creative in using software to design posters or scan in handmade drawings, while illustrating the messaging techniques of government propaganda and marketing bureaucracies in places like China, using materials purchased from the Shanghai Propaganda Poster Art Center as guides, and North Korea (Myers 2010). Students in these courses were also given the opportunity to upload podcasts or videos as descriptions of the cultures they were learning about; one assignment asked students to create a tourism ad encouraging visitors from the West to visit an East Asian country using lessons about culture from the class. However, despite these efforts to encourage students to complete a variety of ***multimodal*** assignments, students still did not appear to grasp the lessons on and nuances of similarity and difference in cultures across the region. Results from the 2013 assessment of these courses largely supports this conclusion.

It is not surprising, then, that these "creative" artifacts did not produce a sophisticated understanding of the course lessons on democracy and governance in Asia. Students composed creative assignments without context—without much thought on how their work at this one moment connected to others across the entirety of the course. Asia, in particular, serves as a challenging topic for study; as important and increasingly relevant as the region is, it is often quite "foreign" to Western students (Bahree, 1986). In terms of government, students are often limited by their personal conceptualizations of ideas like democracy and culture, defaulting to their own perspectives and even stereotypes to answer questions about places like China and Japan. This lacuna between showcasing academic work and drawing connections between artifacts and beyond the subject matter revealed the shortcomings of summative reflection. Indeed, reflection at discrete moments in the semester can help students to lay their views bare and contrast them with the views of citizens of other countries (see Balthazor et al., this collection).

Previous research has detailed the efforts to teach political science students more effectively about Asia by borrowing lessons from the humanities (Sanborn & Ramirez, 2017). As many of the multimodal assignments for the Politics in East Asia and Politics in China courses were drawn from ideas spurred on by professors from the humanities, so, too, were the solutions on how to contextualize lessons about democracy and Asian politics with reflection at discrete moments. In C&C art history classes, for example, students created artifacts that asked them

to envision "sacred spaces" based on lectures and discussions about spirituality in Asian art. While these projects often involved creative expression in the form of graphic design, they also positioned the learner at the center of the lessons. Students would discuss their process of creating these unique assignments and reflect on their choices. In doing so, students personalized their learning of complex subjects, drawing links across artifacts in a more sophisticated way in their summative reflective essays.[2]

Thus, in revising the politics courses, the focus became less on creating multimodal assignments and more on incorporating reflection into students' composition of propaganda posters and podcasts. The students in the courses even participated in Skype exchanges with students from Hong Kong who were protesting for greater democracy in the city. However, these artifacts were bounded by specific requirements to reflect on a concept, such as democracy or accountability, from their own point-of-view. Then, after completing the assignment, the students were asked about what they perceived as different, or how their views had changed, and then they were charged with tracking that gain of knowledge in their reflective essay later in the semester.

For example, in 2015, students were asked to create a free form blog post about interactions with their own governments, followed by a second blog post, informed by readings and class discussions, about how citizens of China hold their government accountable. Then, they engaged in their video conference with students in Hong Kong, asking questions about democracy, armed with their baseline views of what democracy meant to them and what they thought democracy meant to individuals in Asia. Based on this discussion, students were asked to explain what the necessary components of democratic governance were, with the idea that they could glean the similarities and differences of conceptions of this regime type having both talked with students in Hong Kong about democracy and having reflected on their own conceptions of accountability, liberty, and equality.

In addition, students were still given the opportunity to create propaganda posters. However, they were directed to reflect explicitly on their process of creation, rather than simply summarize what they hoped to achieve in their messaging. As a result, students could make connections among artifacts in a more sophisticated manner than they had in earlier iterations of the course.

To test the effectiveness of this approach, we drew summative reflective essays from ePortfolios for each of the classes included in this study: the 2013 pre-revision versions of Politics in East Asia and Politics in China, the 2014 post-revision version of Politics in East Asia, and the 2015 post-revision version of Politics in China. We also assessed essays from two art history courses offered in 2013 on which the revisions were based: History of Asian Art and Chinese Art & Culture. These

2. For more information on the types of assignments discussed here, see Ramirez and Sanborn (2015) and Sanborn and Ramirez (2017).

essays were scored by two raters using an analytical ***rubric*** created for the C&C program (Figure 10.1); classes had ended months before and the points given by the raters did not count in students' grades. An essay was scored from 1 to 5 along the Cultural Understanding, Reflection, and Artifact Analysis scales, respectively, and these points were summed for a rating, out of a total of 15 points. If the total score awarded by each of the two raters for an essay differed by more than three points, the raters discussed their allocation of points for the essay before coming to a consensus. The scores of the two raters were summed for a final overall score out of 30 points, including a score out of 10 points for artifact analysis.

	ANALYSIS OF ePORTFOLIO ARTIFACTS
"5" RATING	Presents a complex, insightful analysis of a selection of substantive, varied, and revealing artifacts which fully support and develop the essay's thesis/focus.
"4" RATING	Presents an effective analysis of a selection of relevant and varied artifacts which effectively support and develop the essay's thesis/focus.
"3" RATING	Presents a clear analysis of a selection of appropriate and varied artifacts, which loosely support and develop the essay's thesis/focus.
"2" RATING	A weak analysis, which may be attributed to a poor selection of artifacts either in terms of relevance, diversity, or quality.
"1" RATING	Fails to include or discuss artifacts.

Figure 10.1 Artifact Analysis Scores. Source: Virginia Military Institute (VMI) Core Curriculum Oversight Committee, guidance for Civilizations and Cultures summative reflective essays.

This analysis of quantitative data is limited by the constraints of a rubric that attempts to quantify reflection and analytical abilities. In addition, there are a small number of observations for each of the classes included in this study, limiting generalizability. However, with clear calls for empiricism in the study of effective practices for the use of ePortfolios by faculty and administrators (Bryant & Chittum 2013; Rhodes et al., 2014), we proceed with this analysis, aided by the use of statistical techniques designed to detect significant differences across small samples.

Table 10.1 features a summary of data on the analysis of artifacts by students enrolled in six courses from 2013–2015. In 2013, one can clearly see the difference in the evaluation of ePortfolio artifacts for both pan-Asia and China-specific courses, by discipline. Students in the Asian and Chinese Art classes scored, on average, a 7.3 out of 10 points. The benchmark for the C&C program, for the sake of comparison, was a total score of 18 points, or 6 points per category. At the same time, the Asian (3 of 10) and Chinese (4.8) politics classes fell well below the standard for the program.

Table 10.1. Analysis of student ePortfolio artifacts in summative reflective essays, 2013–2015

	Artifact Analysis (out of 10)	Standard Deviation	Number of Essays
Pan-Asia courses			
2013 History of Asian Art	7.3	1.30	20
2013 Politics in East Asia	3.0	1.41	14
2014 Politics in East Asia	7.1	1.39	15
p-value of difference between politics courses	**0.0000**		
China courses			
2013 Chinese Art and Culture	7.3	1.35	16
2013 Politics in China	4.8	2.24	17
2015 Politics in China	6.0	1.56	15
p-value of difference between politics courses	**0.0497**		

Notes: To test for the significance of the difference of the scores across politics classes, a two-sample t-test is conducted with equal variances assumed. One-tailed statistical significance ($p < .05$).

In 2014 and 2015, after the revisions to the politics coursework described above, the scores improved significantly. Students in the revamped Politics in East Asia course scored a 7.1 out of 10, on average, while students in the updated Politics in China class scored a 6 of 10. To test whether this improvement in scores met the threshold for statistical significance, we conducted a two-sample t-test comparing the scores of the 2013 politics classes to their later counterparts. In both cases, we were able to reject the null hypothesis of no difference between the two sets of classes; the roughly 4-point average improvement in scores for the Politics in East Asia course was significant at the .001 level, while the 1.2-point improvement in the Politics in China course just crossed the threshold of significance at the .05 level. Students, thus, performed better when the artifacts were explicitly tied to reflection, rather than simply a collection of multi-modal assignments without explicit analysis of their purpose in the student learning process.

Conclusion

The initial motivation for the revision of these politics courses, based on lessons from the humanities, was to promote in students a greater appreciation for culture and sophisticated reflection on how they came to know what they now knew (Sanborn & Ramirez, 2017). A somewhat intended, but certainly fortuitous, outcome was the thoughtful evaluation of artifacts that they included in their ePortfolios. Students drew on different artifacts to make the case for their learning in

the summative reflective essays as they had for years before. They linked together artifacts, however, in a more persuasive narrative, tied to the reflection and learning they developed over the course of the semester.

A simple, perhaps obvious, lesson of this study is that professors should encourage reflection as part of the artifact-selection process. The digital space affords a flexibility and creativity in assignments that many educators are willing to explore with their students. It is the reflection that occurs alongside this exploration that produces deep learning and allows students to develop their thinking about complex subjects as it happens and sort out the lessons of a course upon reflection at the end of the term.

References

Bahree, P. (1986). Teaching about Asia. *Western European Education, 18*(4), 31-59.

Barrett, H. C. (2007). Researching electronic portfolios and learner engagement: The REFLECT initiative. *Journal of Adolescent & Adult Literacy, 50*(6), 436-449.

Barrett, H. C., & Carney, J. (2005). Conflicting paradigms and competing purposes in electronic portfolio development. http://electronicportfolios.com/portfolios/LEAJournal-BarrettCarney.pdf

Bhattacharya, M., & Hartnett, M. (2007, October 10-13). E-portfolio assessment in higher education [Conference presentation]. 37th ASEE/IEEE Frontiers in Education Conference, Milwaukee, Wisconsin.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education, 103*(2), 20-50.

Bryant, L. H., & Chittum, J. (2013). ePortfolio effectiveness: A(n ill-fated) search for empirical support. *International Journal of ePortfolio, 3*(2), 189-198.

Dellinger, M. A., Koons, K., & McDonald, C. (2013). *ePortfolios and the study of civilizations and cultures* [White paper]. Inter/National Coalition for Electronic Portfolio Research, Cohort VI.

Myers, B. R. (2010). *The cleanest race: How North Koreans see themselves and why it matters*. Melville House.

Parkes, K. A., Dredger, K. S., & Hicks, D. (2013). ePortfolio as a measure of reflective practice. *International Journal of ePortfolio, 3*(2), 99-119.

Ramirez, J., & Sanborn, H. (2015). Creativity and reflection through multi-modal learning: Measuring creative expression in student ePortfolios. In K. Coleman & A. Flood (Eds.). *Capturing creativity through creative teaching.* Common Ground Publishing.

Sanborn, H. & Ramirez, J. (2017). Reflections on Asia: Borrowing lessons from art history in East Asia and China coursework. *ASIANetwork: A Journal for Asian Studies in the Liberal Art,s 24*(2), 70-88.

Schön, D. A. (1983). *The Reflective Practitioner*. Basic.

Chapter 11. Accountability and Actionable Data: A Comparison of Three Approaches to Program Assessment using ePortfolios

Rowanna Carpenter
Yves Labissiere
Portland State University

As increasing numbers of higher education institutions adopt and assess ePortfolios, programs and departments within those institutions must balance calls for accountability with the need to generate useful evidence. General education programs, in particular, need to provide external audiences with credible evidence that they advance students' skills in areas such as critical thinking and written communication. At the same time, faculty must be able to use data for program improvement. ***Assessment*** using ePortfolios makes it possible to meet both internal and external demands; however, programs need to plan carefully to do both well.

Calls for higher education accountability have proliferated in recent years. Assurances from educational institutions or accreditors claiming that universities and colleges successfully educate their students do not satisfy policy makers and the larger public. Books such as *Academically Adrift* (Arum & Roksa, 2011) claim universities fail to develop students' critical thinking skills. Efforts such as the Voluntary System of Accountability, internal to the higher education community, call on colleges and universities to present evidence of student learning publicly so potential students and other stakeholders can evaluate the outcomes of higher education. Some suggest standardized tests are the best way to provide such evidence, arguing that tests are easily administered, valid, reliable, and allow comparison across institutions (Arum & Roksa, 2011; Benjamin, et. al, 2012).

However, many assessment experts insist that standardized tests are disconnected from the work that students produce in the classroom and thus do not result in actionable data (Walvoord, 2010). In fact, one of the major critiques of standardized tests is that faculty are not able to use the resulting data to inform program improvement (Linn et al., 1991; McCollum, 2011). As Trudy Banta and her colleagues assert, "educators and policy makers in postsecondary education are interested in assessment processes that improve student learning and at the same time provide institutional data that may be used to demonstrate accountability" (Banta, Griffin et al., 2009). The Association of American Colleges and Universities' (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) initiative directly challenges the idea that standardized tests are

DOI: https://doi.org/10.37514/PRA-B.2020.1084.2.11

the only way to produce valid, reliable, and comparable information about student learning in higher education. As opposed to standardized tests, which often have no connection to students' course work, the VALUE ***rubrics***, developed by teams of national experts, are meant to assess authentic student work—work such as ePortfolios generated in the context of a course. Tracy Penny Light, Helen Chen, and John Ittleson (2012) argue that ePortfolios can "support student self-assessment but also inform and contribute to institutional improvement and educational effectiveness, involving all campus stakeholders ranging from senior leadership to individual students" (p. 98) (see Richardson et al., this collection).

This chapter focuses on reviewing ePortfolios for a program-level assessment that centers on whether the program as a whole is meeting its goals related to student learning, rather than investigating an individual student's or a particular faculty member's performance (Suskie, 2009; Walvoord, 2010). Examining ePortfolio assessment at the program level reveals the tensions between the demands for external accountability and the need to engage faculty to produce actionable data using limited resources (see Day, this collection). Based on assessment literature and our experience reviewing ePortfolios for program-level assessment, we propose the following key criteria for programs that are developing ePortfolio assessment processes.

- Reliability: Acceptable levels of reliability help ensure that any data produced can be viewed with confidence.
- Comparability: External audiences are often interested in comparing measures of student learning across institutions.
- Usability: In order to complete an assessment cycle, programs must be able to interpret findings in ways that help them identify areas for improvement.

As programs consider adopting rigorous ePortfolio assessment practices, factors beyond usability, reliability, and comparability should be considered. Programs must also seek:

- Efficiency: Portfolio review processes require resources in the form of funds and human time. It is important to acknowledge the resource and infrastructure requirements for any assessment process and to be sure that the investment of resources yields actionable information (Banta et al., 2009; Cooper & Terrell, 2013; Suskie, 2009).
- Rater Experience: It is important that any assessment process is seen as valuable, meaningful, and worthwhile to the faculty who participate.

These considerations (reliability, usability, comparability, efficiency, and rater experience) constitute the RUCER framework we propose for programs building ePortfolio assessment processes. In the next section, we explore each criterion in more depth. We then use the framework to compare three approaches to ePortfolio assessment. Finally, taking into account all of the criteria discussed above,

we advance some recommendations for programs considering implementing assessment with ePortfolios.

A Framework of Key Criteria

Reliability

Reliability in an ePortfolio assessment process refers to the extent to which, given the same piece of student work, different raters converge on the same score (see Sanborn & Ramirez, this collection). Acceptable levels of reliability help ensure that any data produced can be viewed with confidence. When assessing ePortfolios, rubrics help establish a scoring process that is consistent and unbiased (Suskie, 2009). In order to improve reliability, assessment approaches using rubrics usually incorporate a training session during which reviewers become familiar with the rubric, practice applying it to a particular piece of student work, and discuss any discrepancies in scoring so that they make scoring decisions consistently with each other. This type of training increases agreement among raters (Penny Light et al., 2012). While several measures of reliability for ePortfolio assessment processes exist, inter-rater agreement—how well two scores on the same piece of student work converge—is a common approach used by many universities (Finley, 2011).

Usability

Successful assessment processes are not only reliable, they must also result in meaningful data (Banta et al., 2009; Peterson & Einarson, 2001). If faculty cannot use the resulting data to inform program improvement, the process is not useful, regardless of reliability. Linda Suskie (2009) outlines four characteristics of useful assessments:

- They yield reasonably accurate and truthful information about what students have learned.
- They have a clear purpose so that assessment results are valued and don't end up sitting on a shelf.
- They engage faculty and staff.
- They focus on clear and important student learning goals. (p. 37)

As program faculty or staff are planning assessment, they should consider what type of data are produced, how those data relate to faculty work, and how the data can be used to inform action and improvement.

Comparability

Calls for accountability often include questions of whether evidence of student learning for a given program is comparable to evidence for other programs or in-

stitutions (Banta et al., 2009). Assessments of embedded, authentic assignments such as ePortfolios are often not appropriate for comparison because they vary so much from institution to institution (Suskie, 2009). Because such assessment approaches connect intimately to program practice and pedagogy and reflect the learning experiences of students in a particular program, we have to acknowledge the challenge to comparability that arises when adopting such localized practices. Nationally recognized rubrics, such as the VALUE rubrics, help mitigate these concerns because they were designed to provide for comparability without standardization (Rhodes, 2011).

Efficiency

It is important to acknowledge the resource and infrastructure requirements for any assessment process and to be sure that the investment of resources yields actionable information (Banta et al., 2009; Cooper & Terrell, 2013; Suskie, 2009). While some assessment or ePortfolio processes involve grading ePortfolios in the context of a course and aggregating those judgments to the program level, many program-level processes select a few samples of student work from key courses and review those in a process that takes place outside of regular classroom parameters. Taking the process outside of the classroom allows for a focus on a single goal or learning outcome, creates opportunities to involve faculty beyond those who teach the course in question, and addresses the problems that can arise when instructors across courses do not use the same assessment practices within their courses (Johnstone et al., 2001; Miller & Leskes, 2005; Suskie, 2009). An approach that takes assessment outside of the classroom uses resources beyond those required to deliver the course material and grade the student work from the course (Banta et al., 2009; Linn et al., 1991; Suskie, 2009). These resources include faculty time, any specific software requirements, and stipends for faculty participants (if applicable).

As a proxy for the resources needed for an assessment process, we use efficiency, including the time it takes to read ePortfolios, the number of ePortfolios reasonably possible to assess in a single day, and the time it takes to calibrate or train raters. Each of these measures relates to the overall resources—funds and human time—required to complete the task.

Rater Experience

It is important that any assessment process is seen as valuable, meaningful, and worthwhile to the faculty who participate. In addition to producing data that serve accountability purposes, an assessment process can provide the occasion for in-depth conversations about student learning and expectations for students within a program (Briggs, 2007; Hutchings, 2010; Suskie, 2009). Such conversations are only valuable, however, if faculty are invested in the process and the outcomes. Any changes supported by assessment results cannot be implemented

without faculty participation (Banta et al., 2009). With the use of embedded assessments, such as ePortfolios, faculty can make a direct link between assessment conversations and their classroom practices (McCollum, 2011; Suskie, 2009). Such conversations are also an opportunity to reinforce using assessment for improvement and not for individual faculty evaluation, a practice that promotes trust and continued engagement in the assessment process (Suskie, 2009; Walvoord, 2010). For faculty to learn from their experience reviewing portfolios, they need time to read the ePortfolios, have conversations with each other, and reflect on the results for their own classroom practices. An important consideration is the balance between the efficient production of assessment data and the time required for faculty to feel engaged and energized by the process.

Method

Given higher education institutions' need to respond to calls for accountability, including providing valid and reliable evidence of student learning, we want to inform the conversation about ePortfolio assessment by offering a comparison of three approaches to assessing ePortfolios using the framework we have presented above. We are interested in examining the reliability and usability of each process alongside a consideration of the efficiency of the process (how much does it cost to generate results?), the faculty experience in the process, and the comparability of the data across programs and institutions.

To examine our framework and learn about the potential contributions of alternative assessment strategies, we compared an approach using a ***holistic rubric***, an approach using ***adaptive comparative judgment***, and an approach using an ***analytic rubric*** (one of the VALUE rubrics), each described below (see Sanborn & Ramirez, this collection). We invited ten experienced reviewers to participate in our project. These faculty taught the freshman general education course from which we drew the ePortfolios or taught at other levels of the general education program. Each of them had participated in our established portfolio review process at least five times. All faculty were familiar with program learning goals and existing ePortfolio review processes.

Each assessment process took place on a single day. Because we value the conversation and collective experience of convening a group of faculty in one place, we did not examine asynchronous review processes. However, we believe our framework can help inform other assessment approaches as well. We conducted the review processes on three successive days. On each day, the faculty met in the morning to receive orientation to the day's task, spent the bulk of the day reviewing portfolios, and reconvened at the end of the day to provide feedback on the process. These review processes took place during the summer and each faculty reviewer was compensated for participation. Following the review, we analyzed reliability data, the ePortfolio score data, data on the length of time the reviews took, and our notes on the reviewers' experiences with each process.

Approach 1: Local Holistic Critical Thinking Rubric

During the first review process, the ten "expert" reviewers participated as part of an established ePortfolio review process with 25 other reviewers. The larger process involved scoring 265 student portfolios. All reviewers, including our experts, received orientation and training in the morning, during which they were introduced to a locally-developed six-point holistic critical thinking rubric (see Appendix A. Note: Since the writing of this chapter, the rubric has been revised substantially.). All reviewers read and scored one ePortfolio and convened for a conversation about the scores. Following a question and answer session during which reviewers worked toward consensus about how to score the first sample ePortfolio, the reviewers scored a second ePortfolio. After a second discussion, reviewers read and scored the rest of the ePortfolios.

Each ePortfolio was read by at least two people. If the two scores were the same or only one score apart, the average of the two scores was used as a final ePortfolio score. When the difference between the first two scores was two or more, a third reviewer read and scored the ePortfolio. To avoid having this process influence subsequent ratings, our expert reviewers read a sample of portfolios that were not included in the ACJ or VALUE processes.

Reliability for this approach was measured through inter-rater agreement, calculated by determining the number of ePortfolios needing a third review and dividing by the total.

Approach 2: Adaptive Comparative Judgment (ACJ)

On the second day, our expert reviewers participated in an Adaptive Comparative Judgment (ACJ) process. For this process, we included a set of 100 portfolios that had not already been read by our reviewers. ACJ is an approach to rating ePortfolios that involves comparing two ePortfolios and selecting a "winner" between the two. In this case, raters were asked to select the ePortfolio that represented the better example of critical thinking, as defined by our holistic rubric, and declare it a winner. After making their selection, the judges made notes about why they made that decision. We were curious about ACJ as an assessment technique for several reasons. First, this approach promises superior reliability coefficients—well above .9 (Pollitt, 2012)—thus addressing one of the concerns with traditional rubric approaches to ePortfolio assessment. Second, this approach also involves a different way of reading student work. Rather than comparing the work with a set of criteria from a rubric, the ACJ approach asks reviewers only to compare two portfolios and make one choice. Each portfolio is then compared with several others over the course of the day and each portfolio is read by more judges than in a standard holistic rubric scoring approach. For details on this approach, see Pollitt (2012).

The ACJ software tracks agreement for each portfolio. As consensus forms on a particular portfolio, it is removed from the subsequent pair presentation. In

other words, if portfolio A is consistently judged better than other portfolios, it is removed from subsequent trials. The ACJ process relies on a modified Rasch model to calculate reliability (Pollitt, 2012).

Approach 3: VALUE Integrated Learning

We chose to use one of the AAC&U VALUE rubrics as a third assessment approach: the ***Integrative Learning*** rubric (see Appendix B). Having previously piloted several VALUE rubrics, we found that the Integrative Learning rubric correlated most closely with our local critical thinking rubric. The VALUE rubric also offered an opportunity to compare a holistic rubric approach with an analytic rubric approach.

The group of reviewers met in the morning to review the rubric. Because the rubric is analytic and includes five categories each associated with a four-point rating scale, the scoring conversation took longer than the comparable conversation for the holistic rubric. To have time to score the ePortfolios, we only scored one ***calibration*** ePortfolio and had a single conversation before asking the reviewers to proceed with reading the rest of the ePortfolios. Two reviewers read and scored each ePortfolio. We included the same sample of 100 portfolios that were reviewed during the ACJ process. Reviewers were instructed to skip any portfolio they had previously viewed, so we were getting fresh reads for each portfolio. We reconvened at the end of day for a conversation about the process.

Reliability for this approach was measured through inter-rater agreement for each of the five rubric criteria.

Findings

In this section, we compare our existing approach to ePortfolio assessment (Existing/Holistic), adaptive comparative judgment (ACJ), and an approach using a VALUE rubric (VALUE/Analytic) using the framework we have proposed and outline our findings for each criterion below. Table 11.1 summarizes the primary findings for each of our three assessment processes for the five criteria.

Existing/Holistic

Efficiency

Training raters in our existing holistic rubric approach takes approximately 90 minutes. This session includes an overview of the process and review and discussion of two calibration portfolios.

Because our "experts" were embedded as part of a larger ePortfolio review process, we could compare the time it took them to rate ePortfolios with the time clocked by other reviewers who participated in scoring ePortfolios that day. We found that our experienced reviewers took an average of 9.8 minutes for a first review and 6.5

minutes for a second review. Our less experienced reviewers took an average of 13 minutes for a first review and 8.7 minutes for a second review. In addition, our experienced reviewers assessed 64 portfolios, reading an average of 13 portfolios each.

We had not considered that our experienced reviewers would take so much less time to review portfolios than our less experienced reviewers. The result makes sense, but it also suggests that the rest of our findings need to be considered in light of this discovery. We are referring to processes as performed by faculty experienced in reading and scoring ePortfolios.

Reliability

Overall reliability, as measured by inter-rater agreement was 83%.

Usefulness of Data

The data produced through this process are an overall mean rubric score for the program and mean scores for faculty teams (generally consisting of three to five faculty members). We also produced distributions of portfolio scores at the program and team levels. These data are useful in that they are derived based on a rubric that was developed in house by our faculty and therefore align closely with the program definition of critical thinking. These data are also the basis of the assessment reports that faculty have received each year for more than a decade. Faculty are familiar with those reports and the underlying data and are accustomed to having conversations about the findings (see Appendix C).

A holistic rubric score gives an overall sense of student achievement, but does not pinpoint specific areas for development as an analytic rubric score would do. When faculty are presented with an aggregate team or overall score that has risen or declined, they must make some assumptions about the factors that may have played a role in that change. This shift in scores prompts discussion of pedagogical practices, assignments, and student responses to those as faculty work to explain and contextualize the results of the assessment process.

Comparability

Because we are using a locally developed rubric, we are not able to compare our critical thinking scores with any group beyond our institution. This emphasis on a local instrument can be problematic when faced with questions about accountability. However, because we have been using this rubric for an extended period, we do have historical data. We can compare scores over many years internally.

Raters' Experience

Our experienced reviewers enjoyed the opportunity to read student work and discuss that work with colleagues. Some raters expressed frustration with the process of scoring using a holistic rubric because an ePortfolio containing several

artifacts may exhibit evidence from more than one scoring level. Because of the volume of ePortfolios that need to be read in a day, reviewers can feel pressured to base their judgment on the first appropriate piece of evidence they find rather than spend time reading the entire ePortfolio. Generally though, reviewers enjoy the aspects of academic community, discussion, and deep reading of student work that are present in our current practice.

Adaptive Comparative Judgement (ACJ)

Efficiency

Training scorers for the ACJ approach took approximately 60 minutes. This included an overview of the process and practice sessions using a sample of portfolios during which reviewers practiced selecting the winner among pairs of ePortfolios based on the representation of critical thinking in the holistic rubric. Thus, speed was tacitly encouraged in the task. Reviewers knew that since the judgment of any one portfolio was a communal one, they could get away with a more "impressionistic" glance and get the job done. The reviewers were able to rank all 100 sample portfolios in one day. The average time spent making a comparison was 5.4 minutes and reviewers made an average of 41 comparisons each over the course of the day.

Reliability

Calculated using a modified Rasch model, overall reliability using ACJ was 93%. We predicted the reliability coefficient would reach 98%, had our reviewers been able to complete one more round of comparisons.

Usefulness of Data

The data produced through this process are a rank-ordered list of portfolios. While this ranked distribution tells us how well each student does relative to the other students in the sample, there is no "objective" or independent assessment of quality such as in the other two methods. There is no way to know if the top ePortfolio in the sample represents a 6 on the rubric or a 3. There is no way to tell whether scores tend to cluster around a particular point or are widely distributed across a range of scores. A possible way to ameliorate this problem would be to use "anchor" portfolios that represent each point on our rubrics (these are ePortfolios that score a 1, 2, 3, 4, 5 or 6) and see where they fall on the distribution. That, we hope, would provide a sense of the quality of the distribution. However, without that additional scoring process to identify the ***anchor ePortfolios***, this assessment procedure produces a rank-order list without reference to specific levels of learning that might be represented. Along with the rank, the data also consist of reviewers' comments about why they made particular judgments. These comments are available for each portfolio. These qualitative data provide insight into the important aspects of critical thinking present in the samples that reviewers are using to make their judgments. The data can provide information about the tacit criteria that are in play during a

scoring session. Combined with the ranks, they provide descriptors of the top and bottom ePortfolios, an interesting addition that is not currently available from the other approaches examined here (see Appendix D).

Thinking about how we might present this data back to faculty also raised challenges. Given the ranking data, we could provide faculty teams a distribution of where their student portfolios fell across the ranks, but unless their students' scores were clustered toward the top or bottom, we are not sure how faculty would interpret these data. Without additional information such as anchor portfolios suggested above, faculty might not have enough information to make sense of the findings and identify areas for change or improvement.

Comparability

Since what we generate from this process is a relative distribution of scores of local samples of ePortfolios, our results cannot be generalized to another sample—local, non-local, internal, or external. What may be promising, however, may be the use of anchor ePortfolios shared across time within an institution (and perhaps even between institutions with common artifact types and rubrics). If this were possible, it might create an opportunity for comparability across institutions, while maintaining a superior inter-rater reliability.

Raters' Experience

The experience of our ACJ reviewers was mixed. As predicted, they found it easier to make the pairwise comparisons and select the winner among two portfolios. However, many expressed disappointment with not "having to get close" to the student's work. In other words, for many comparisons, the work to select the "winner" was somewhat shallow, easy, and therefore not as satisfying as reading to get a holistic sense of the student's work. The experience felt more impersonal to many raters, as if they were primarily "scorers" or rankers, a task that could be performed without connecting to the "meaning" in the work. Some reviewers also expressed feeling pressure to make a quick judgment. While some preferred the user friendliness of the task, others bemoaned the distancing (to the student's work) nature of the experience. This process also offered less opportunity for collegial conversation than the other two. In an ACJ process, all reviewers must engage in the comparison process at the same time to complete a round. In this way speed was incentivized and some reviewers reported feeling pressure to move quickly. In order to finish the scoring process, we needed to move quickly between rounds of comparisons. Reviewers did not have the self-pacing and breaks for conversation available in the other two processes.

VALUE Analytic

Efficiency

The calibration process took considerably longer for the VALUE rubric than for our existing holistic rubric calibration process. While the goal was to review two

test ePortfolios, the process of reviewing the VALUE rubric, which was unfamiliar to these reviewers, only allowed for one test portfolio and one calibration conversation within the 90 minutes we had allotted. We expect that if we were to adopt the rubric and use it consistently over several years, the time required for calibration would be reduced. We expect it would still be longer than the time required for a holistic rubric, but would be less than what we experienced in this study.

While reviewing ePortfolios using the VALUE Integrative Learning rubric, reviewers took an average of 12.3 minutes for a first review and 12.5 minutes for a second review. This is 2.5 and six minutes longer than our standard process, respectively. The reviewers were able to complete reviews of 46 ePortfolios (two reviews each) for an average of nine portfolios per reviewer.

Reliability

Inter-rater agreement varied across the rubric subcategories. Three categories reached or almost reached acceptable agreement levels: Connection to Discipline (78%); Transfer (80%); and Integrated Communication (83%). The two categories that related to how well students integrated their own experiences into their ePortfolios had lower levels of inter-rater agreement: Connection to Experience (71%); and ***Reflection*** and Self-Assessment (66%). Some of these lower levels of agreement would likely have increased with additional rounds of calibration.

Usefulness of Data

The VALUE rubric data are mean scores and score frequency distributions for the program and for each team on each of five rubric categories. The data are more nuanced than a holistic rubric score and can give some indication of specific areas in which the program is doing well and specific skills that may need more attention. Whereas the conversation following a holistic rubric scoring session may be fruitful, faculty may have a hard time pinpointing activities that could help improve the score if they feel that is appropriate because a global score does not offer the detail of a set of analytic rubric scores. The VALUE rubric data provide more specific information, which allows for more targeted conversations about specific skills students may need assistance in developing.

Comparability

The data produced through an assessment process using a VALUE rubric are not strictly comparable to any other university, even if they are using the same VALUE rubric because most universities adapt the rubrics to their uses and in fact are encouraged to do so (Finley, 2012). However, several states are participating in the Multi-State Collaborative to Advance Learning Outcome Assessment, which aims to use the VALUE rubrics to "produce valid data summarizing faculty judgments of students' own work, and also seeks to aggregate results in a way that allows for benchmarking across institutions and states" (State Higher Education Executive Officers Association, 2015). As a growing number of universities use

the rubrics for various projects, programs that use the rubrics join a national conversation about student learning and have a similar starting point for talking about critical thinking or integrative learning across campuses.

Raters' Experience

Generally, raters reported that they appreciated the additional time they got to spend with student ePortfolios in the VALUE rubric assessment process. Because they were rating students' work across five criteria, they needed to read more of the ePortfolio to determine their ratings.

Some raters liked the experience of using a rubric (VALUE or internal) less than the ACJ process. For these raters, the rubric seemed somewhat artificial, and they felt it could be difficult to distinguish among categories. Other raters liked that they were able to focus on specific elements of the portfolio to determine a rating and were not asked to compare other portfolios as part of the process.

Table 11.1. Summary of findings

Holistic	ACJ	VALUE
Reliability		
Acceptable	Highest	3/5 Acceptable 2/5 Not acceptable
Usability of Data		
One score requires assumptions and interpretation Have historical data Faculty developed the rubric—they understand the score	Ranking of ePortfolios highest to lowest is most difficult to interpret	5 scores result in more specific information about particular skills
Comparability		
Internal history No external comparability	No external comparability	This is the most comparable with other institutions Many universities are using the rubrics
Efficiency		
Middle in terms of training and rating times 64 ePortfolios scored	Quickest judgment process Most ePortfolios scored (100)	Longest training process Fewest ePortfolios scored (46)
Reviewer Experience		
One score can be difficult to determine Like the interaction with colleagues	Some liked the comparison process Least interactive process Less connection to student work	Mixed Some liked the anchor points Liked the time necessary to read and score

Discussion

This research study offered an opportunity to take a close look at three ePortfolio assessment practices using our proposed RUCER five-criteria framework. As a result, we better understand each process and the framework of criteria that we suggest institutions need to balance as they consider implementing ePortfolio assessment practices. Our framework, which adds a consideration for the quality of evidence, faculty experience, and the resources required to run each process to more common considerations of reliability and comparability, gives institutions a realistic view of what each of these approaches can offer. The framework and its criteria represent the complexity of considerations institutions should entertain when making decisions about any ePortfolio assessment process, synchronous or asynchronous, using many types of scoring schemes. We weigh each process below to provide an example of the way our framework can play out as a decision-making tool.

If the only considerations on the table were reliability and efficiency, and the institution has decided against a standardized test, then the ACJ approach is promising. During that review process, reviewers completed comparisons of all 100 ePortfolios with the highest reliability. The VALUE rubric and holistic rubric processes offer similar measures of reliability, and the holistic rubric was the second most efficient. Not surprisingly, the process in which reviewers had to make the largest number of judgments, the VALUE process, resulted in the fewest number of portfolios being read, indicating that it would cost the most to produce the data.

If comparability beyond the institution is important, the only process offering that possibility is one using a VALUE, or other nationally developed, rubric. Although the data may not be strictly comparable across programs or institutions, an approach using a VALUE rubric makes it possible to have cross-campus and cross-institutional conversations about our students' strengths and weaknesses.

However, if an institution or department wishes to implement a robust and meaningful assessment cycle, it must look beyond reliability, efficiency, and comparability to considerations of data usefulness and rater experience, considerations both squarely anchored in the experience of faculty. Assessment planners must consider, even before any data are collected, what type of data will be produced by a given process, how those data might be presented to faculty for consideration, and whether those data will be considered meaningful by the faculty in question.

Although the ACJ process was most efficient and reliable, the rank-ordered data are not, in and of themselves, meaningful. We had difficulty envisioning how faculty could interpret the data in ways that led to meaningful shifts in practice and improvement in student learning. The holistic rubric, on the other hand, is based on our program's definition of critical thinking, so it is closely connected to our program and the data are meaningful to our faculty who have worked with them for several years. The VALUE rubric process produces the most detailed data related to student work, pinpointing specific skills that emerge as strengths for our students and others that may need work.

Any of these three sets of data can be improved by aligning them with other program data about students and their learning. For example, we often supplement our holistic rubric score with an inventory of the types of student work included in ePortfolios or student responses to aligned items from course evaluations. The ACJ process would be enhanced by having a few of the ePortfolios reviewed using a rubric and including those as anchor portfolios.

Conversations related to assessment and faculty support are converging (e.g., Carpenter & Fitzmaurice, 2019; Stanny, 2018) with increased recognition that assessment processes serve as rich opportunities for faculty support, not just data gathering activities, supporting our inclusion of reviewer experience as a criterion for consideration. ePortfolio review processes give faculty a glimpse into each other's courses through the work students produce. Gathering faculty on an ePortfolio review day offers many opportunities for conversation about the meaning and manifestation of program learning outcomes, the strengths and challenges observed in the student artifacts, and collegial conversation about assignment and curricular design. These are outcomes that are not captured in measures of efficiency, but certainly matter if we want to be accountable for improving student learning outcomes.

As part of this research project, we had conversations with experienced ePortfolio reviewers. They provided important insight into the experience with each ePortfolio assessment process, information programs should consider before adopting new practices. Although they did not articulate it as such, these faculty members helped us see the importance of the human aspects of an assessment process. Reviewers are not data production tools. As reviewers, they agreed to be part of a social process that is focused on students' experiences and learning. When they were not able to interact with each other as much (as in the ACJ process), they were less satisfied with their personal experience of the day. Any assessment process must take these human needs into account (Briggs, 2007). Because of these candid conversations with our reviewers, we have a much clearer sense of the ways in which people seem to prefer to read ePortfolios and interact with other raters, the rubric or task, and the ePortfolio.

These conversations with faculty help illuminate the workings of the assessment process itself, contribute to a deeper understanding of the process, and allow assessment practitioners to improve the quality of future assessment endeavors. The conversations and interactions can also contribute to deeper faculty engagement with the learning outcomes and more authentic buy-in for any improvement effort that results.

The framework proposed in this project proved to be a useful set of criteria for evaluating ePortfolio assessment processes. The criteria, taken together, highlight the need to center the human aspects of the process (review and meaning making) while weighing important considerations of reliability, efficiency, and comparability. Practitioners who are tasked with developing an ePortfolio review process can use these criteria to weigh the relative importance of all factors. We argue that adding a focus on the usability of the results, the experience of the re-

viewers, and resources helps ensure that the process will produce assessment that both supports program improvement and provides for accountability.

References

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* The University of Chicago Press.

Banta, T. W., Griffin, M., Flateby, T. L., & Kahn, S. (2009). *Three promising alternatives for assessing college students' knowledge and skills*. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice.* Jossey-Bass.

Benjamin, R., Miller, M. A., Rhodes, T. L., Banta, T. W., Pike, G. R., & Davies, G. (2012). *The seven red herrings about standardized assessments in higher education*. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Briggs, D. C. (2007). Assessing what students know, how they know it, or both? *Measurement: Interdisciplinary Research and Perspectives, 5*(1), 62-65.

Carpenter, R., & Fitzmaurice, C. (2019). Assessment and faculty support: Fostering collegial community to strengthen professional practice. *The Journal of General Education, 67*(1-2), 90-108.

Cooper, T., & Terrell, T. (2013). *What are institutions spending on assessment? Is it worth the cost?*. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Finley, A. (2011). How reliable are the value rubrics? *Peer Review, 13*(1), 31-33.

Hutchings, P. (2010). *Opening doors to faculty involvement in assessment*. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Johnstone, S. M., Ewell, P., & Paulson, K. (2001). *Student learning as academic currency.* American Council on Education.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

McCollum, D. L. (2011). The deficits of standardized tests: Countering the culture of easy numbers. *Assessment Update, 23*(2), 3-5.

Miller, R., & Leskes, A. (2005). *Levels of assessment: From the student to the institution.* Association of American Colleges and Universities.

Miller, R., & Morgaine, W. (2009). The benefits of e-portfolios for students and faculty in their own words. *Peer Review, 11*(1), 8-12.

Penny Light, T., Chen, H. L., & Ittleson, J. C. (2012). *Documenting learning with eportfolios: A guide for college instructors*. Jossey-Bass.

Peterson, M. W., & Einarson, M. K. (2001). What are colleges doing about student assessment? Does it make a difference? *Journal of Higher Education, 72*, 629-669.

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281-300.

Rhodes, T. (2009). The VALUE project overview. *Peer Review, 11*(1), 3-7.

Rhodes, T. (2011). Emerging evidence on using rubrics. *Peer Review, 13*(4), 4-5.
Stanny, C. (Ed.) (2018). Special issue: Assessment in action: Evidence-based discussions about teaching, learning, and curriculum. *New Directions for Teaching & Learning, 2018*(155), 1-116.
State Higher Education Executive Officers Association (2015) MSC: A Multi-state collaborative to advance learning outcomes assessment (Pamphlet). https://www.aacu.org/sites/default/files/files/VALUE/MSC_DY_FAQ_100815.pdf
Suskie, L. (2009). *Assessing student learning: A common sense guide*. Jossey-Bass.
Walvoord, B. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. Jossey-Bass.

Appendix A. University Studies' Holistic Critical Thinking Rubric[3]

4	➤ Identifies and develops a compelling question or problem that meaningfully recognizes context. ☗ Demonstrates a systematic approach to exploring a topic, problem, or issue through research, documented experimentation, and/or other methodologies. ◆ Analysis integrates a diverse range of relevant considerations and/or points of view. ● Conclusions and related outcomes reflect student's informed evaluation and ability to assess and weigh evidence and perspectives. ▣ Creates a novel or unique idea, question, format, or product, and incorporates new directions or approaches in the final product. ✪ Evaluates own strengths, challenges, and/or assumptions and identifies important areas for further exploration, learning, or understanding.
3	➤ Identifies and develops a question or problem that acknowledges context. ☗ Demonstrates awareness of methodology, though the approach is not always thorough or fully developed. ◆ Analysis represents a range of relevant considerations and/or points of view. ● Conclusions and related outcomes follow from the evidence and reflect student's evaluation and ability to assess and weigh evidence and perspectives. ▣ Experiments with creating a novel or unique idea, question, format, or product, and considers new directions or approaches to the final product. ✪ Identifies own strengths, challenges, and/or assumptions and some areas for further exploration, learning, or understanding.

3. The University Studies Critical Thinking Rubric was updated in 2016. This is the updated version of our rubric, but it reflects the structure of our previous holistic rubric. For a copy of our previous rubric, please contact either of the authors.

2	➤ Identifies a question or problem with limited understanding of context ☗ Demonstrates some awareness of methodology, but the approach is neither thorough nor in-depth. ◆ Analysis represents a limited range of considerations and/or points of view. ● Conclusions and related outcomes reflect student's attempt at evaluation and ability to assess and weigh evidence and perspectives. ▣ Reformulates a collection of available ideas, and may acknowledge alternate, divergent, or contradictory perspectives or ideas. ✪ Mentions own strengths and/or challenges, with little recognition of own assumptions or the possibility of further exploration, learning, or understanding.
1	➤ Does not clearly identify a question or problem. Shows little understanding of context. ☗ Demonstrates little awareness or understanding of methodology. ◆ Analysis represents no range of considerations and/or points of view. ● Conclusions are not connected to evidence. ▣ Primarily summarizes or repeats available information. ✪ Minimal acknowledgment of own strengths, challenges and/or assumptions.
0	➤ Demonstrates no attempt to identify a question or problem and shows no understanding of context. ☗ Demonstrates no awareness of methodology. ◆ Demonstrates no analysis. ● The student reaches no conclusions, and evidence is either missing or inaccurate. ▣ Does not identify relevant information. ✪ No acknowledgment of own strengths, challenges and/or assumptions.

Appendix B. AAC&U Integrated Learning VALUE Rubric

Integrative Learning VALUE Rubric

for more information, please contact value@aacu.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 by shared nationally through a common dialog and understanding of student success.

Definition

Integrative learning is an understanding and a disposition that a student builds across the curriculum and co-curriculum, from making simple connections among ideas and experiences to synthesizing and transferring learning to new, complex situations within and beyond the campus.

Framing Language

Fostering students' abilities to integrate learning—across courses, over time, and between campus and community life—is one of the most important goals and challenges for higher education. Initially, students connect previous learning to new classroom learning. Later, significant knowledge within individual disciplines serves as the foundation, but integrative learning goes beyond academic boundaries. Indeed, integrative experiences often occur as learners address real-world problems, unscripted and sufficiently broad, to require multiple areas of knowledge and multiple modes of inquiry, offering multiple solutions and benefiting from multiple perspectives. Integrative learning also involves internal changes in the learner. These internal changes, which indicate growth as a confident, lifelong learner, include the ability to adapt one's intellectual skills, to contribute in a wide variety of situations, and to understand and develop individual purpose, values, and ethics. Developing students' capacities for integrative learning is central to personal success, social responsibility, and civic engagement in today's global society. Students face a rapidly changing and increasingly connected world where integrative learning becomes not just a benefit . . . but a necessity.

Because integrative learning is about making connections, this learning may not be as evident in traditional academic artifacts such as research papers and academic projects unless the student, for example, is prompted to draw implications for practice. These connections often surface, however, in reflective work, self-assessment, or creative endeavors of all kinds. Integrative assignments foster learning between courses or by connecting courses to experientially-based work. Work samples or collections of work that include such artifacts give evidence of integrative learning. Faculty are encouraged to look for evidence that the student connects the learning gained in classroom study to learning gained in real life situations that are related to other learning experiences, extra-curricular activities, or work. Through integrative learning, students pull together their entire experience inside and outside of the formal classroom; thus, artificial barriers between formal study and informal or tacit learning become permeable. Integrative learning, whatever the context or source, builds upon connecting both theory and practice toward a deepened understanding.

Assignments to foster such connections and understanding could include, for example, composition papers that focus on topics from biology, economics, or history; mathematics assignments that apply mathematical tools to important issues

and require written analysis to explain the implications and limitations of the mathematical treatment, or art history presentations that demonstrate aesthetic connections between selected paintings and novels. In this regard, some majors (e.g., interdisciplinary majors or problem-based field studies) seem to inherently evoke characteristics of integrative learning and result in work samples or collections of work that significantly demonstrate this outcome. However, fields of study that require accumulation of extensive and high-consensus content knowledge (such as accounting, engineering, or chemistry) also involve the kinds of complex and integrative constructions (e.g., ethical dilemmas and social consciousness) that seem to be highlighted so extensively in self reflection in arts and humanities, but they may be embedded in individual performances and less evident. The key in the development of such work samples or collections of work will be in designing structures that include artifacts and reflective writing or feedback that support students' examination of their learning and give evidence that, as graduates, they will extend their integrative abilities into the challenges of personal, professional, and civic life.

Glossary

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

- Academic knowledge: Disciplinary learning; learning from academic study, texts, etc.
- Content: The information conveyed in the work samples or collections of work.
- Contexts: Actual or simulated situations in which a student demonstrates learning outcomes. New and challenging contexts encourage students to stretch beyond their current frames of reference.
- Co-curriculum: A parallel component of the academic curriculum that is in addition to formal classroom (student government, community service, residence hall activities, student organizations, etc.).
- Experience: Learning that takes place in a setting outside of the formal classroom, such as workplace, service learning site, internship site or another.
- Form: The external frameworks in which information and evidence are presented, ranging from choices for particular work sample or collection of works (such as a research paper, PowerPoint, video recording, etc.) to choices in make-up of the ePortfolio.
- Performance: A dynamic and sustained act that brings together knowing and doing (creating a painting, solving an experimental design problem, developing a public relations strategy for a business, etc.); performance makes learning observable.
- Reflection: A meta-cognitive act of examining a performance in order to

explore its significance and consequences.
- Self Assessment: Describing, interpreting, and judging a performance based on stated or implied expectations followed by planning for further learning.

Integrative Learning VALUE Rubric

for more information, please contact value@aacu.org

Definition

Integrative learning is an understanding and a disposition that a student builds across the curriculum and co-curriculum, from making simple connections among ideas and experiences to synthesizing and transferring learning to new, complex situations within and beyond the campus.

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 4	**Milestones 3**	**Milestone 2**	**Benchmark 1**
Connections to Experience *Connects relevant experience and academic knowledge*	Meaningfully synthesizes connections among experiences outside of the formal classroom (including life experiences and academic experiences such as internships and travel abroad) to deepen understanding of fields of study and to broaden own points of view.	Effectively selects and develops examples of life experiences, drawn from a variety of contexts (e.g., family life, artistic participation, civic involvement, work experience), to illuminate concepts/theories/ frameworks of fields of study.	Compares life experiences and academic knowledge to infer differences, as well as similarities, and acknowledge perspectives other than own.	Identifies connections between life experiences and those academic texts and ideas perceived as similar and related to own interests.
Connections to Discipline *Sees (makes) connections across disciplines, perspectives*	Independently creates wholes out of multiple parts (synthesizes) or draws conclusions by combining examples, facts, or theories from more than one field of study or perspective.	Independently connects examples, facts, or theories from more than one field of study or perspective.	When prompted, connects examples, facts, or theories from more than one field of study or perspective.	When prompted, presents examples, facts, or theories from more than one field of study or perspective.

	Capstone 4	Milestones 3	Milestone 2	Benchmark 1
Transfer *Adapts and applies skills, abilities, theories, or methodologies gained in one situation to new situations*	Adapts and applies, independently, skills, abilities, theories, or methodologies gained in one situation to new situations **to solve difficult problems or explore complex issues in original ways.**	Adapts and applies skills, abilities, theories, or methodologies gained in one situation to new situations **to solve problems or explore issues.**	Uses skills, abilities, theories, or methodologies gained in one situation in a new situation **to contribute to understanding of problems or issues.**	Uses, in a basic way, skills, abilities, theories, or methodologies gained in one situation **in a new situation.**
Transfer *Adapts and applies skills, abilities, theories, or methodologies gained in one situation to new situations*	Adapts and applies, independently, skills, abilities, theories, or methodologies gained in one situation to new situations **to solve difficult problems or explore complex issues in original ways.**	Adapts and applies skills, abilities, theories, or methodologies gained in one situation to new situations **to solve problems or explore issues.**	Uses skills, abilities, theories, or methodologies gained in one situation in a new situation **to contribute to understanding of problems or issues.**	Uses, in a basic way, skills, abilities, theories, or methodologies gained in one situation **in a new situation.**
Integrated Communication	Fulfills the assignment(s) by choosing a format, language, or graph (or other visual representation) **in ways that enhance meaning**, making clear the interdependence of language and meaning, thought, and expression.	Fulfills the assignment(s) by choosing a format, language, or graph (or other visual representation) **to explicitly connect content and form**, demonstrating awareness of purpose and audience.	Fulfills the assignment(s) by choosing a format, language, or graph (or other visual representation) that **connects in a basic way** what is being communicated (content) with how it is said (form).	Fulfills the assignment(s) (i.e., to produce an essay, a poster, a video, a PowerPoint presentation, etc.) **in an appropriate form**.

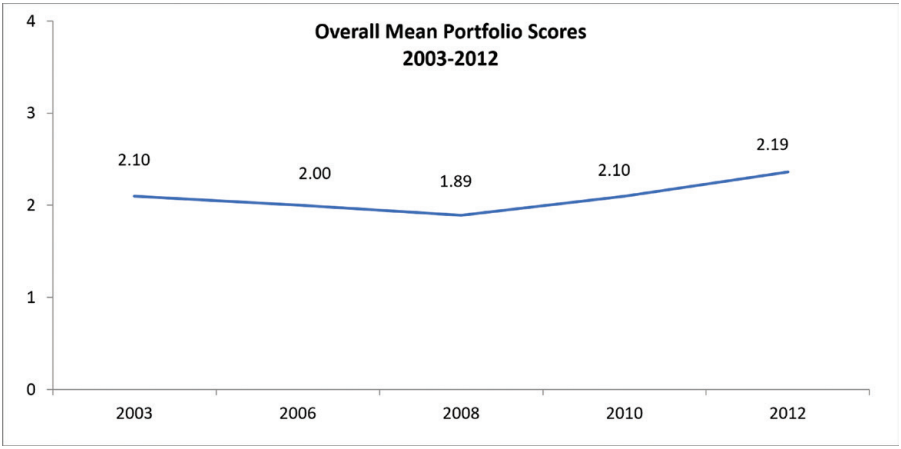
Continued

	Capstone 4	Milestones 3	Milestone 2	Benchmark 1
Reflection and Self-Assessment *Demonstrates a developing sense of self as a learner, building on prior experiences to respond to new and challenging contexts (may be evident in self-assessment, reflective, or creative work)*	Envisions a future self (and possibly makes plans that build on past experiences that have occurred across multiple and diverse contexts).	Evaluates changes in own learning over time, recognizing complex contextual factors (e.g., works with ambiguity and risk, deals with frustration, considers ethical frameworks).	Articulates strengths and challenges (within specific performances or events) to increase effectiveness in different contexts (through increased self-awareness).	Describes own performances with general descriptors of success and failure.

Integrative Learning VALUE Rubric
for more information, please contact value@aacu.org

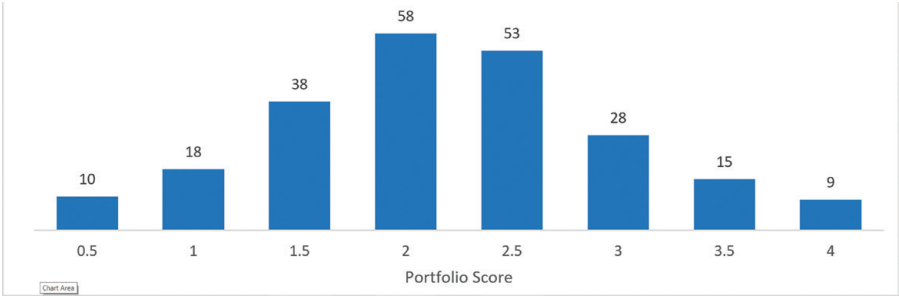
Appendix C. Example Assessment Report

Inquiry and Critical Thinking Assessment



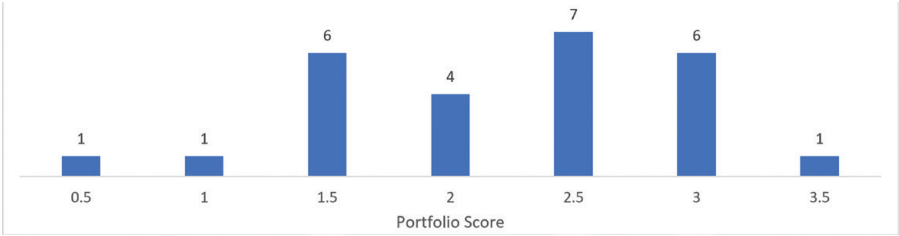
Overall

Number of student work samples: 229
Mean Score: 2.19 on a 0 to 4 scale.
Number of papers per score.



Team 1.

Number of student work samples: 26
Mean Score: 2.21 on a 0 to 4 scale.
Number of papers per score



Inquiry and Critical Thinking Discussion. Inq & CT Rubric Data

- Look at the holistic rubric. Where would you expect your students to score? How does the distribution of rubric scores for your theme compare with those expectations?
- Given these scores and your experience with your students, do you identify areas in need of improvement? What aspects of critical thinking do your students do well? struggle with?
- What actions will you take as a team or as individual faculty to further enhance inquiry and critical thinking in your courses?
- Is there support the UNST program can offer you to assist in addressing those areas?

Favorite Assignment

- Share a description of an assignment related to this goal with each other.
- Look at the inquiry and critical thinking rubric. Which of the criteria is most relevant to the assignment?

- As you look across assignments in your group, does your theme emphasize a particular kind of approach to inquiry and critical thinking?
- Are there aspects of inquiry and critical thinking that you could enhance through assignment redesign or course adjustments?
- Do the rubrics provide ideas for modifying the assignment?

Appendix D. Adaptive Comparative Judgment Ranking and Comments

Port. No.	Rank	ACJ Score	Theme	Comments
68	1	17.4	1	Strong research project in Portfolio A and it's even listed under Critical thinking!
				A is more reflective; uses analysis and synthesis in his/her thinking. Also, included various pieces of evidence.
				Beautiful voice; applies concepts well to self and other texts/experiences. I am biased in favor of voice, which may have influenced by very slight preference for B
				The author of portfolio A used critical thinking in all aspects of the assignments presented. There was just a lot more detail about that process than in B
86	2	16.5	3	B makes connections to other classes; also, comparative and analytical approaches are highlighted.
				Both portfolios were good. Portfolio A had some great short writing exercises, which were a less formal assignment, which helped me make my decision.
57	3	14.7	1	Portfolio B has a fully realized research paper with an original thesis
				Though A demonstrated the process of inquiry and learning, B did so more proficiently and with more gusto.
				A had more depth and development of the students' own ideas.
78	4	13.3	6	Both portfolios were quite thoughtful but I chose B because it offered more samples each one accompanied by a reflective part. As total, it was more nuanced
				Really just a bit more sophisticated than B, but a tough distinction for me, as I am impressed by both.

Continuing through the rankings to the lowest ranked portfolios

Port. No.	Rank	ACJ Score	Theme	Comments
74	97	-8.1	5	Merely supporting a conclusion.
				Very little was shown of any substance in this portfolio. This was a difficult decision because neither portfolio was very robust or displayed assignments calling for risk and critical thinking.
				Material in A was mostly summary.
8	98	-8.8	3	Really doesn't move beyond presentation of discovered information. Very basic
				This portfolio did not really include any examples of critical thinking. There was a lot of description but little evidence of engaging.
				A was informative, but lacked depth or critical inquiry.
7	99	-9.5	7	Lacking work samples to assess, so could only go by the final reflection, which lacked depth compared to the analysis paper in the other portfolio.
				This was tough, because neither portfolio had much evidence in terms of students' work. I picked B since there was a little more evidence and the students referred to other sources and connected them to their ideas.