Cohort V Final Report – University of Cincinnati

1. Research question

At the outset of our participation in Cohort V (that is, in summer 2008), we defined our main research question as follows: Does the comparison of the CLA to ePortfolios – in terms of effective and persuasive measurement of learning outcomes – warrant the adoption of the Collegiate Learning Assessment (CLA) as a university-wide assessment instrument? We wanted to compare data from the CLA with that from evaluated reflective essays collected in ePortfolios, this comparison then to provide comparative information on how baccalaureate competencies at the University of Cincinnati are effectively adding value as measured over time and across disciplines.

2. Main findings

Findings from our investigations into published research:

The Voluntary System of Accountability (VSA) in 2009 completed a "Test Validity Study" between three standardized tests – the CLA, the MAPP (the Measure of Academic Proficiency and Progress, now known as the ETS Proficiency Profile), and the CAAP (the Collegiate Assessment of Academic Proficiency) – that is, the three tests permitted for measuring student learning outcomes under the VSA protocol (Klein, Liu, & Sconing, 2009). While the study presents reassuring evidence that the VSA-approved tests provide internally consistent institutional rankings based on aggregated student achievement in the areas of critical thinking, effective writing, and problem solving, the study provides only very weak evidence of correlation of test scores for instruments measuring similar constructs. In particular the CLA shows only weak evidence of convergent or divergent validity between the CLA and the MAPP and the CAAP. The VSA study concludes that the CLA "Make an Argument" and "Critique an Argument" sub-test are "relatively unreliable at the student level," and the study's authors hypothesize that such "brief constructed-response tests" have "lower reliability" (Klein, Liu, & Sconing, 2009, p. 25). They conclude:

> The correlations between scores on different tests, especially at the student level, are affected by the reliability of those scores. When the individual student is the unit of analysis, multiple-choice tests are known to yield more reliable scores per hour of testing time than constructed-response measures. This is one explanation for the fact that correlations tended to be lower when one or both tests employed a constructed response format (it may also be that they measure different

constructs). This suggests that, when scores are used to make decisions about individual students, such as for course placement, special attention should be given to their reliability. Similarly, drawing conclusions about a student's relative strengths across skill areas should be limited to instances where the differences are statistically significant. (p. 34)

Similarly, in a recent Lumina-funded study comparing cross-sectional and longitudinal CLA results, Klein (n.d.) raises concerns about the reliability of the CLA results. His findings indicate that "reliability of the effect sizes (are) modest, suggesting that the effect sizes (are) not as precise an estimate of gain as we would like. For this reason, schools should base assessments of their students' progress on more than one or two CLA administrations worth of data, and they should try to reduce measurement error by increasing sample sizes."

Such studies have thus raised concerns about using the CLA test results for individual student-level and program assessment and suggest that great care must be taken in assigning meaning to individual student CLA scores or to outcomes that are not drawn from large data samples covering several years.

Findings from our own research and experience:

Our assessment of student artifacts was conducted using the Association of American Colleges and Universities VALUE rubrics measuring written communication and critical thinking. These rubrics appear to be quite useful in providing a framework for assessing student-constructed responses for individual student evaluation, at least within the context of a single institution. When used carefully, the rubrics can yield high inter-rater reliability among faculty from various units in the assessment of the traits associated with the competencies that faculty experts have agreed represent competent writing and critical thinking. VALUE rubrics further provide student-level assessment of clearly identified and separable factors, thus providing useful insight for individual student diagnostic assessment. Moreover, the rubrics can also help guide continuous program improvement in that students' average or individual scores on a particular trait can be observed over time and provide useful data before and after the introduction of program modifications or changes in pedagogical methods.

Our study further found that VALUE-scored student artifacts designed by faculty to capture critical thinking and effective writing do not correlate to any meaningful degree with the CLA test designed to measure the same or similar skills. This lack of alignment with faculty evaluations of the same skill sets for the same student in the same time frame suggests the CLA may not reflect faculty valuation and thus cannot replace faculty-driven

assessment of and attention to student achievement in providing individual diagnostic assessment of performance in the areas of writing or critical thinking. This result is particularly troubling because faculty improvements in the student skill sets as measured by the VALUE rubrics may not result in any commensurate increase in the student CLA scores.

Our study reinforces the widely held notion that an institution needs to employ a complex and varied assessment strategy to understand the achievement of student learning outcomes, to demonstrate the value of the educational experience offered by that institution, to make sure that students are well prepared for lifelong learning and ongoing intellectual inquiry, and to provide a framework for program feedback and improvement. More specifically, however, our study suggests major limitations in the CLA's ability to contribute significantly to such a strategy, at least from the standpoint of faculty involvement in assessment at the University of Cincinnati.

In short, it is troubling that there is no systematic relationship between the scores generated for the same students performing tasks designed to showcase similar skills (writing and critical thinking) within a similar time frame. At worst, it calls into question the validity of the instruments, and at best, it highlights the differences between the constructs and intentions of the instruments.

For faculty at the University of Cincinnati, the lack of correlation between the CLA results and the results from the rubric-scored assignments will only further diminish their regard for such tests as the CLA since it seems to be measuring learning outcomes quite different from what the faculty themselves value and seek to develop within their own classrooms and disciplines. Given these results and UC's history of using rubrics to assess student learning outcomes in relation to general education, it appears likely that UC faculty will have greater confidence in data that comes from rubric-scored student artifacts created within the higher-stakes context of course assignments. At least within the setting of this experiment, we know that VALUE rubrics can be applied in ways that yield high agreement on student performance across faculty from different disciplines. What is needed now is a multi-institution investigation replicating the work of this study across a representative sample of institutions. Only then will we have evidence that VALUE- scored student artifacts can be scored with adequate inter-rater reliability to produce the "comparable scores" required for VSA reporting.

Looking back at our original research question, however – to examine "comparison of the CLA to ePortfolios" – we came to realize that ePortfolios had much less to do with our research than we had originally intended. In scoring individual student artifacts using the VALUE rubrics, we were examining individual pieces of writing that indeed existed

within ePortfolios but that could just as easily have been handed to the reviewers as a printed stack of essays. More meaningful assessment needs to emerge out of the interaction between pieces of evidence within an ePortfolio rather than from individual and relatively isolated pieces of evidence.The complex relationships between these various elements – the VALUE rubrics, individual student artifacts, ePortfolios, and assessment strategies – are described in Darren Cambridge's study (2010, pp. 99-107) in ways that help clarify the various pieces we are still working to assemble into a more ambitious institutional assessment plan.

3. **Inquiry processes,** including the methods of inquiry that you used and, if applicable, the value of embedding research in your ongoing electronic portfolio practice

During the fall term of 2008-09 at the University of Cincinnati, 111 Honors students were enrolled, for academic credit, in an introductory honors seminar requiring both ePortfolios and the CLA. The student learning goals of this seminar included progress on the university's baccalaureate competencies of effective communication, social responsibility, critical thinking, and knowledge integration, as well as personal reflection on individual learning styles, goals, and accomplishments.

For the ePortfolio portion of the study, students were asked to submit two class assignments to be evaluated using scoring rubrics designed in the spirit of the AAC&U's VALUE rubrics, in this case for written communication and critical thinking. At the beginning of the seminar (and during the same week in which the CLA was administered), students submitted a reflective essay in response to an assignment to consider what was important to them from their high school experiences, why they had chosen to attend the University of Cincinnati, and what their expectations had been of their first academic year at UC. This assignment was then scored using a variation of the VALUE rubric for written communication.

Seven weeks after completing the CLA, and again through their ePortfolios, students submitted a second assignment, this one designed to exhibit their critical-thinking skills. Students now had to describe and propose an experiential-learning project in which they identified possible learning opportunities, including objectives, prior assumptions, expectations, and plans for reflection and dissemination. With the focus now on critical thinking, these proposals were scored – again by two faculty members – only for those 55 students who had completed the CLA performance task and not for the full cohort of 111.

Although the number of rating teams had increased for this part of the study – five teams instead of only three – the correlations were again reliably high: the average scoring group correlation was 0.669, and the total rater correlation was 0.765.

4. **Value of participation in the Coalition,** including specific background, changes in research direction, choice of methodologies, learning from other cohorts or members of Cohort V

The main value of our participation came about through the cohort structure that served often to remind us to ground our practice in research and test it against research-based theory. That led us at one point, for instance, to ask David Shulenburger for feedback on a draft version of what has become a manuscript for an article. Along with George Mehaffy and Christine Keller, Shulenburger had published an autumn 2008 article that acknowledged the controversy already swirling around the VSA and its attention to core educational outcomes. Shulenburger thus had the expertise to provide us with a valuable perspective on our work, and his feedback did indeed result in significant changes in the draft.

Our ability to ground our recommendations upon research will also lend credibility to our faculty-development work at UC as we seek to increase the involvement of our faculty in the assessment process more generally. The title of Pat Hutchings' (2010) paper for the National Institute for Learning Outcomes Assessment – "Opening Doors to Faculty Involvement in Assessment" – points towards one recurring theme underlying her six recommendations. All of Hutchings' recommended strategies are designed to help advance "what happens when faculty are significant participants in the assessment process – not just token members of a committee cobbled together for an accreditation visit or an after-the-fact audience for assessment results they had no part in shaping but central voices and shapers of activity" (p.7).

5. **Application to practice** and/or **dissemination** of your research, present and future

Our work during our participation with Cohort V has resulted in conference presentations at the Assessment Institute in Indianapolis, IN (2008 and 2009) the ISSOTL Annual Meeting in Bloomington, IN (2009), the Student Success Assessment Summit at the Ohio State University in Columbus (3 presentations in 2009), the Higher Learning Commission Annual Meeting, Chicago, IL (2009), and the National Association for Schools of Art & Design Conference (2010). We have also produced an article-length study of our work that is currently under consideration with an academic journal.

In relation to our ongoing work with faculty at UC: As noted above, the results of our study emphasize the importance of keeping assessment of student learning firmly anchored within the curriculum and also very much under the ownership of the faculty.

Moreover, the VALUE rubrics can also help guide continuous program improvement in that students' average or individual scores on a particular trait can be observed over time and provide useful data before and after the introduction of program modifications or changes in pedagogical methods.

As we approach the conversion from quarters to semesters at UC beginning in autumn 2012, such a framework for assessment – and what we have learned about assessment through our attention to ePortfolios and to the CLA – can prove particularly valuable. As it looks at ways to document student learning outcomes, semester conversion for UC has taken some of its shape from a variety of external expectations about accountability. If we are able to improve our assessment of student learning through our attention to ePortfolios, might we simultaneously strengthen our efforts in scholarly teaching as well as the scholarship of teaching & learning (SoTL)?

One more insight regarding our approach to assessing ePortfolios: As we evaluated the student artifacts using the VALUE rubrics, we found too many discrepancies between ratings through the first pass. Two strategies emerged for addressing that: make some adjustments and revisions to the rubrics so that they fit more closely with the actual assignments, and spend more time discussing the rubrics among the raters, using more examples for norming purposes. In other words, don't assume that the AAC&U rubrics, in and of themselves, will fully address issues of assessment with ePortfolio artifacts (or, presumably, with ePortfolios as digital, networked structures). They need close examination and also in-depth discussion so that raters fully understand the basis for the rubric as well as its relationship with that which is to be assessed.

6. **Next research questions** following the work done during the Cohort

Our preliminary experience with the Honors students who are developing ePortfolios suggests that one important element is "portfolio as presentation." That is, the need to present one's work more formally and in a more encompassing format creates a powerful "crisis moment" that plays a key role in learning, by focusing and crystallizing the various organizing structures in that learning with reference to strategic check points in the student's development.

Further research now will continue to explore this emerging insight through such overlapping/interlocking questions as these:
- At the most general level, how can we develop assessment strategies that will apply not just to individual student artifacts within ePortfolios but to ePortfolios as multi-faceted and complex structures themselves?

- To what extent can we measure student learning in ePortfolios with reference to this "crisis moment" concept?
- To what extent will students choose to use the crystallizing functions of such a "crisis moment," and to what extent does this need to be required?
- To what extent can such a "crisis moment" be captured and understood through the reflection components of an ePortfolio?
- To what extent can such a "crisis moment" affect students' reflection, or disposition to reflect?
- What role does the potential for or the experience of failure play in learning, in those cases where the ePortfolio presentation seemed to result in a failed format?

In exploring these questions, we have interview information to draw upon that was gathered in August 2010 from students in Honors who have been using iWebfolio ePortfolios.

In addition, we have now administered another nationally standardized test (the ETS Proficiency Profile), and its results from October 2010 will allow us further to expand our comparative study of ePortfolio artifacts with the results from nationally standardized tests.

One ongoing account of our work, with a number of documents relevant to our study, is Rich Robles' "UC E-Portfolio Project" at
http://www.iwebfolio.com/portfolios/customReviewerView?y=1235630268572

Endnotes:

Cambridge, D. (2010). *Eportfolios for lifelong learning and assessment.* San Francisco: Jossey Bass.

Hutchings, P. (2010, April). *Opening doors to faculty involvement in assessment* (Occasional Paper No.4). Urbana, IL: National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomeassessment.org/documents/PatHutchings_000.pdf

Klein, S., Liu, O.L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., Nemeth, A., Robbins, S., & Steedle, J. (2009). *Test Validity Study Report*. Retrieved from the Fund for the Improvement of Postsecondary Education website: http://www.voluntarysystem.org/docs/reports/TVSReport_Final.pdf

Klein, S. (n.d.). *The Lumina Longitudinal Study: Summary of Procedures and Findings*. Retrieved from http://www.collegiatelearningassessment.org/files/CLA_Lumina_Longitudinal_Study_Summary_Findings.pdf

Shulenburger, D., Mehaffy, G, & Keller, C. (2008). The voluntary system of accountability: Responding to a new era. *Liberal Education, 94* (4), 48-54.