Assessing General Education Learning Outcomes

▸ **Linda Siefert,** director of assessment, University of North Carolina Wilmington

Much has been written about the current shift toward a knowledge economy and the resulting effects on our society and culture (Houghton and Sheehan 2000). Just as the practices of our business community are quickly evolving to keep pace with this shift, so is the way the education community thinks about assessing the learning of those who will function in this new knowledge economy. In the twentieth century, assessment relied on tests of *explicit* knowledge, or what we call content knowledge in education. Since content is now available quickly and inexpensively through electronic sources, simply knowing the correct answer no longer defines expertise. As educators, we need to prepare our students for success in life and in their careers by placing more emphasis on knowledge as transferable skills and abilities, such as the abilities to communicate thoughtfully and effectively, to think creatively and critically, and to access, evaluate, and use information to accomplish a purpose. To use the language of information and technology researchers, our focus is changing from assessing *codified knowledge* to assessing tacit knowledge (Stiles 2000). This requires more complex assessments that rely on authentic demonstrations and detailed and well-vetted rubrics. Whether these demonstrations are assessed individually from course-embedded assessments or in a purposeful offering like a portfolio, the processes are similar and the challenges remain the same.

FACULTY AS AN INTEGRAL PART OF THE PROCESS

The tacit knowledge and transferable skills that our faculty believe will prepare our students for the twenty-first-century workplace are reflected in the learning goals adopted by the University of North Carolina Wilmington (UNCW). Since curriculum and assessment are two sides of the same coin, our approach to assessing the UNCW Learning Goals places faculty at the center of the process. The faculty role begins within the general education courses. Each instructor of a course chosen for assessment selects a course assignment that is a regular part of the course and that addresses the specific learning goal. Since the assignments are part of the course content and course grades, students are motivated to perform at their best. This also means that little additional effort is required on the part of course faculty. Another benefit is that there is a natural alignment often missing in standardized assessments, and results can be linked directly to the curriculum, making it straightforward to identify areas of strength and areas for reinforcement of learning.

Faculty members also do all of the scoring in the UNCW assessment process. The student work products sampled are scored independently from the instructor grades by trained scorers from across disciplines using common rubrics. A disadvantage to this approach (when compared to standardized tests) is that results cannot be compared to those from other institutions. We are mitigating this by using the AAC&U VALUE Rubrics (Rhodes 2010) for four of our learning goals. Our hope is that institutions begin to share their findings from the VALUE Rubrics so that cross-institutional comparisons can be made.

Preparing scorers is key to obtaining reliable results. Metarubrics, such as the VALUE rubrics, are constructed so that they can be used to score a variety of student artifacts across preparation levels, across disciplines, and across universities. However, the generality of a metarubric makes it more difficult to use than a rubric that is created for one specific assignment. At UNCW, faculty scoring volunteers initially attend a two-hour workshop on one rubric. At the workshop, they review the rubric in detail, are introduced to the assumptions we've adopted for applying the rubrics, and practice scoring benchmark papers. Subsequent discussion begins the process of developing a common understanding of the rubric. On the day of scoring, scorers work in groups of two or three. Scoring papers from an assignment begins with the members of the group independently scoring one piece of student work and discussing their scores. This discussion continues the norming before scorers proceed through scoring the assigned papers.

VALUE rubrics detail four levels of achievement—benchmark, two levels of milestones, and capstones. One of the most important

assumptions we make when using the VALUE rubrics is that we are comparing each separate work product to the characteristics we want the work of UNCW graduates to demonstrate—level 4 of the rubrics. So even when we are using the rubrics to assess work products from 100- and 200-level courses, we are comparing the work to our expectations for graduating seniors, not to other students in the course or even students of the same level. We have not, as yet, determined our exact expectations for scores on work from these lower-division courses. That is why the results presented here give the percent of work scored at or above each of the milestones, levels 2 and 3.

WHAT WE'VE LEARNED ABOUT STUDENTS' ABILITIES

UNCW has results from four of the VALUE Rubrics (written communication, inquiry, critical thinking, and information literacy) for our general education courses (mainly 100- and 200-level courses). We have discovered things about our students that are contrary to anecdotal information. For example, within the learning goal of written communication, students are not in general weaker in the control of syntax and mechanics than they are in other dimensions of writing, although this is an area often discussed by faculty as problematic. Instead, our results show that students struggle most with using sources to support ideas. Results also help point out relative strengths and weaknesses across the learning goals. Findings to date illustrate relative strengths in information literacy, followed by written communication and inquiry. Critical

thinking scores have shown the most need for improvement, and have also provided the most useful information for curriculum change. Details are provided for information literacy and critical thinking in tables 1 and 2.

INFORMATION LITERACY FINDINGS

We have assessed student work products on information literacy in one academic year, sampling seventy-eight work products from four sections of our culminating English composition class, one of two main general education courses that emphasize information literacy skills.

For this UNCW Learning Goal, the scores are fairly consistent across all dimensions of the rubric with respect to the percent of work products scored at or above a level 2. Relative strengths and weaknesses show up more clearly for the work scored at or above a level 3. At this milestone, we see, for example, that almost 60 percent of students in this course are able to use at least three of the information-use strategies that provide evidence of ethical and legal use of information (IL5). However, only 40 percent of the work products evaluated the information to the level of identifying the students' and others' assumptions when presenting a position (IL3). Almost half (49 percent scored at a level 2) of the students identified some assumptions, although they may have demonstrated more awareness of others' assumptions than their own. With 81 percent of the students in the sample in their first and second years, the findings indicate that students have a sound base from which to continue to practice their information literacy skills within their majors.

CRITICAL THINKING FINDINGS

We have two years of results for critical thinking, from 302 student work products (187 in year 1 and 115 in year 2) sampled from fourteen sections of history, music, psychology, and sociology introductory courses.

Although not at all surprising, results from student work scored on critical thinking are the lowest across all learning goals. Within the dimensions of critical thinking, student performances were scored highest on explaining the issues (CT1), with over a third of the students able to clearly describe and clarify the issue to be considered (scores of 3 and 4), and another third of the students able to describe the issue, although with some information omitted (scores of 2). Students had the most difficulty identifying context and assumptions when presenting a position (CT3), and tying conclusions to a range of information, including opposing viewpoints and identifying consequences and implications (CT5).

First- and second-year students accounted for 77.8 percent of the work products scored in this sample. The results indicate that much practice is needed for these students in subsequent years. It also suggests that general education courses, in addition to courses in the majors, will likely need to stress critical thinking more in order for future graduates to attain capstone, or level 4, scores in critical thinking before graduation.

It is important to mention that we made a small change to the critical thinking rubric between years 1 and 2. According to feedback we received from faculty scorers after the first round of using the VALUE critical thinking rubric, the second dimension, Evidence, was difficult to apply. This dimension contains two statements, one addressing the level of interpretation and development of analysis, and the other focused on questioning the viewpoints of experts. Based on this feedback, we piloted a change to the rubric in which the two statements were applied independently. When we did this, the scores on the first

TABLE 1. INFORMATION LITERACY RESULTS

DIMENSION	PERCENT OF WORK PRODUCTS SCORED 2 OR HIGHER	PERCENT OF WORK PRODUCTS SCORED 3 OR HIGHER
IL1 Determine Information Needed	87.2%	46.2%
IL2 Access Needed Information	89.6%	46.8%
IL3 Evaluate Information and Sources	88.5%	39.7%
IL4 Use Information Effectively	85.9%	43.6%
IL5 Access and Use Information Ethically	93.6%	59.0%

part, interpreting the evidence and developing an analysis (CT2.1), are the highest of all dimensions, and the scores on the second part, questioning the viewpoints of the experts (CT 2.2), are the lowest of all dimensions. The information found from dissecting the dimension is quite important, as it suggests that students need to be instructed on the importance of including an author's viewpoint in critical analysis.

RELIABILITY OF THE SCORES

While the reliability of the scores used in our analysis is vital, a complete discussion of this topic is beyond the scope of this article. As already mentioned, reliability starts with norming the scorers, helping them come to a common understanding of the rubric. Measuring interrater reliability (IRR) tells us how well our norming process is working. We compute reliability statistics that are based on a subsample of work products that are scored by multiple scorers, usually between 20 and 30 percent of the total number of work products. The statistics we report include agreement measures—percent agreement and Krippendorff's alpha—and a consistency measure—Spearman's Rho. Benchmarks were determined based on the work of Krippendorff (2004) and Nunnally (1978).

We have met our benchmarks on some dimensions of the rubrics, but we have much more work to do. While our process was designed to help scorers navigate the difficul-

ties and standardize their application of the rubric, we are using our inter-rater reliability results to enhance our scorer norming procedures. Among other things, work needs to be done to set standards for aligning the dimensions of the metarubric to the requirements of an assignment.

CHALLENGES OVERCOME AND STILL TO BE ADDRESSED

Based on the year of work by a committee convened on general education assessment, we were able to anticipate many challenges as we began our work. The committee recommended a structure for accomplishing the assessment activities and for recommending changes based on those results. Among other structures, funds were allocated for faculty stipends for participating in the norming and scoring activities, and the Learning Assessment Council was charged with making recommendations to the faculty senate committee in charge of the general education curriculum.

Other challenges are arising as we implement general education assessment full scale. Our general education curriculum has grown from 214 to 247 courses with the introduction of our new University Studies curriculum, and this number will continue to increase as we implement the additional components of this new four-year curriculum. Appropriate sampling from the curriculum is key to being able to generalize

our findings to our student body, and planning that sampling over time is an integral part of our work. We are also working on additional avenues for disseminating the findings more directly to the faculty responsible for courses in the general education curriculum. To this end, we are providing a series of workshops through the Center for Teaching Excellence in which instructional and assessment experts provide the findings and discuss best practices in teaching and assessing each learning goal.

Looking ahead, we have begun assessing senior-level courses in the majors using the same set of VALUE rubrics. With these data, we will have information on our students' ability levels in their early years at the university and near graduation. This process will help determine where student skills are growing the most, and which skills may need additional emphasis over time.

We have built an assessment process around the skills that will matter most in the twenty-first century, one in which faculty participation and feedback is central. We will continue to look closely at our work, refining both the process and the rubrics based on evidence, in order to make sure our assessment results most closely reflect what our students know and can do. And we look forward to comparing our results to those of other universities. ▪

REFERENCES

Houghton, J., and P. Sheehan. 2000. *A Primer on the Knowledge Economy.* Victoria, Australia: Victoria University, Centre for Strategic Economic Studies Working Paper No. 18. Retrieved from http://www.cfses.com/documents/Wp18.pdf.

Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology.* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Nunnally, J. C. 1978. *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Rhodes, T. L. ed. 2010. *Assessing Outcomes and Improving Achievement Tips and Tools for Using Rubrics.* Washington, DC: Association of American Colleges and Universities.

Stiles, M. J. 2000. "Developing Tacit and Codified Knowledge and Subject Culture within a Virtual Learning Environment." *International Journal of Electrical Engineering Education* 37, 13–25.

TABLE 2. CRITICAL THINKING RESULTS

DIMENSION	PERCENT OF WORK PRODUCTS SCORED 2 OR HIGHER	PERCENT OF WORK PRODUCTS SCORED 3 OR HIGHER
CT1 Explanation of Issues	68.3%	35.5%
CT2 Evidence Year 1	65.0%	28.2%
CT2 Evidence Year 2*		
Interpreting and Analysis	72.8%	38.6%
Questioning viewpoint	40.9%	13.6%
CT3 Influence of Context and Assumptions	48.8%	21.2%
CT4 Student's Position	54.5%	24.0%
CT5 Conclusions and Related Outcomes	47.7%	17.0%

*In Year 2, CT2 was scored as two separate statements. See discussion below.