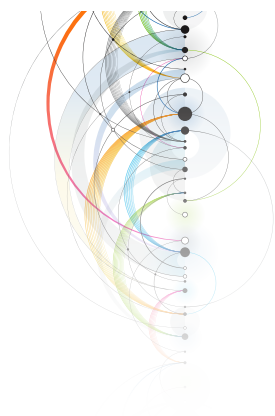


**Abstract**

The goals of a Liberal Studies education are designed to prepare citizens to live responsible, productive, and creative lives in a changing world. Ideally, a liberal education fosters well-grounded intellectuals with dispositions toward learning and an acceptance of responsibility regarding their ideas and actions. To measure the efficacy of a Liberal Studies education, a Midwestern regional university developed a systematic, rubric-guided assessment based on nationally recognized science principles and inquiry processes to evaluate student work in undergraduate science laboratory courses relative to a liberal education. The rubric presented a direct measure of student understandings of science inquiry processes. The assessment procedure used stratified random sampling at confidence levels of 95% to select student work, maintained anonymity of students and faculty, addressed concerns of university faculty, and completed a continuous improvement feedback loop by informing faculty of assessment results to assess and refine science-inquiry processes of course content. The procedure resulted in an assessment system for benchmarking science inquiry processes evident in student work and offered insights into the effect of undergraduate science laboratory courses on student knowledge and understanding of science inquiry. The current assessment was without additional burdening of faculty or supplementary testing of students.

**AUTHORS**

Judith Puncochar, Ph.D.  
Northern Michigan  
University

Mitchell Klett, Ph.D.  
Northern Michigan  
University

## A Model for Outcomes Assessment of Undergraduate Science Knowledge and Inquiry Processes

**L**iberal education is an approach to college learning that empowers individuals and prepares them to deal with complexity, diversity, and change inherent in a democracy (Carson, 1997). This approach emphasizes broad knowledge of science, culture, and society (Pingree, 2007). A liberal education is posited to help students develop an intellectual foundation to recognize real world issues and a sense of social responsibility to hone practical skills for solving problems in real-world settings (Schneider, 2008). The Association of American Colleges and Universities conducted a survey in 2013 and found 74 percent of employers would recommend a liberal education approach to college-bound students (Hart Research Associates, 2013). “What employers clearly want and need are liberally educated professionals” (Humphreys, 2013, para. 8). A commitment to advancing and improving liberal education must be measured and assessed to determine how well the liberal education approach meets the intended outcomes.

Assessment of science knowledge and learning is centuries old and initially used processes such as the Socratic Method. More recently, an upsurge of standardized testing has influenced assessment of science knowledge, but standardized tests do not offer a process by which to improve science inquiry processes and learning outcomes of natural sciences courses in higher education (Steedle, Kugelmass, & Nemeth, 2010). Standardized testing methods rarely assess student learning experiences, account for individual differences in learning needs, or assess the ability of students to think analytically, understand big picture concepts, or apply specific science details to the real world.

The American Association for the Advancement of Science (AAAS; 2013), the National Research Council (NRC; Shavelson & Towne, 2002), and National Science Teachers Association (NSTA; 2011) agree that scientific inquiry is a powerful way for students to understand science content. Assessing student understanding of science inquiry knowledge and processes has been a challenge to several assessment approaches. Students must learn

**CORRESPONDENCE**

*Email*  
jpuncoch@nmu.edu

how to ask science questions and use evidence to answer these questions. In the process of acquiring strategies of scientific inquiry, students learn to conduct an investigation, collect evidence from a variety of sources using evidence-based methodologies (Faust, 2000), develop an explanation from the data, and communicate and defend their conclusions. Scientific inquiry refers to these “activities through which students develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world” (National Academy of Sciences, 2013, p. 23).

Our research posits the use of a rubric based on guidelines from the AAAS, NRC, NSTA, and the National Numeracy Network (NNN) to assess undergraduates’ learning of science inquiry knowledge and processes in their science laboratory courses. We also propose an easy to implement, easy to replicate, and generalizable method of data collection of student work in undergraduate science laboratory classes.

Previous science program assessments may not have been based on science standards and their data collection methods may not have been easily transferable to other programs. For example, various methods to measure student knowledge of science inquiry and to assess natural science outcomes have included attitudinal surveys, interviews, journaling, performance assessments, portfolios, conceptual level tests, and rubrics (Ellis, Mathieu, & Brissenden, 2003). Traditional testing of students in science laboratory courses has shown little value in guiding student learning or in course or program improvement (Rennie, 1994). Evaluations of science laboratory instruction have lacked feedback on student learning outcomes (Seymour, Wiese, & Hunter, 2003). Alternative assessments of the influence of a science program on science literacy have included an internal evaluation conducted with teams of students in capstone courses to explore student perceptions of science learning (Augeri et al., 2011), an examination of the relationship between science knowledge and creating argumentation (Hakyolu & Ogan-Bekiroglu, 2011), and quasi-experimental comparisons of student achievement of inquiry-based science knowledge under conditions of the presence or absence of traditional tests and quizzes (Taylor, 2000) and student-centered versus teacher-centered instruction (Lord, Travis, Magill, & King, 2005).

Web-based science assessment tools are available for science program evaluations (e.g., Assessment Tools in Informal Science, 2011). Several web-based computer models of active science processes offer additional approaches to help students understand science inquiry (Kastens & Rivet, 2008). Inquiry Science Environment (WISE) web-based modules provide visualizations of thermodynamics, electrostatics, and plate tectonics to guide students to connect scientific ideas when conducting inquiry investigations (Resnick & Zurawsky, 2007). The Student Assessment of Learning Gains (SALG) offers powerful individualized statistical analysis of science learning from a student’s perspective to help with immediate formative course evaluation (Seymour, Wiese, Hunter, & Daffinrud, 2000). However, student self-assessments are best used by instructors who seek to improve their courses or by students who can take responsibility for their own learning improvement. Such self-report methods are unlikely to be reliable or generalizable for science program improvement.

An evaluation of science inquiry processes ideally would include performance assessments of student understanding of tools and processes for addressing scientific relationships within the real world, which could be difficult to implement on a large scale (Buxton & Provenzo, 2011). Moreover, procedures for performing an evaluation of a postsecondary science program have political and educational importance. Results must be reliable, unbiased, meaningful, and based on the strength of evidence, but such program evaluations are few in number (Slavin, 2008).

For a process to be useful in measuring student knowledge and understanding of science inquiry, an assessment must focus on student learning, be useful for program and course improvement, employ replicable methods to assess student work, and have a process in place to act on the findings. The university implemented these criteria and followed steps in Wright’s (2003) assessment loop:

**Standardized testing methods rarely assess student learning experiences, account for individual differences in learning needs, or assess the ability of students to think analytically, understand big picture concepts, or apply specific science details to the real world.**

1. Setting goals or asking questions about student learning and development;
2. Gathering evidence that will show whether these goals are being met;
3. Interpreting the evidence to see what can be discovered about students' strengths and weaknesses;
4. Using those discoveries to change the learning environment so that student performance will improve.

The cycle was repeated to include improved interventions for student learning based on assessment data. Our assessment goal was to determine the extent of undergraduate science inquiry abilities and understandings as evidenced by student work in science laboratory courses.

**An evaluation of science inquiry processes ideally would include performance assessments of student understanding of tools and processes for addressing scientific relationships within the real world, which could be difficult to implement on a large scale.**

Laboratories have opportunities for students to design and conduct investigations. Students can collect evidence needed to answer a variety of questions, draw conclusions, and think critically and logically to create explanations based on evidence. In science laboratories, students have a setting to communicate and defend their results to peers and others. This study is limited to an examination of student knowledge and understanding of science inquiry processes within science laboratory courses taught during Spring Semester 2010.

The university's bulletin has a description of core competencies expected of students in science laboratory courses. Students in science laboratory courses are expected to be active in learning the processes and strategies of scientific inquiry. Students also are expected to demonstrate knowledge of science and abilities, design and conduct investigations, collect evidence from a variety of sources, develop an explanation from the data, and communicate and defend their inferences from data to conclusions. Student work in science laboratory courses should provide evidence not only of studied scientific knowledge, but also of the nature of scientific inquiry processes. Scientific, analytical, and logical processes should transcend particular course knowledge to provide students with greater talents and abilities to solve problems and reason rationally.

### Foundations of Assessment Process

The American Association for Higher Education and Accreditation (AAHEA) placed assessment as an ongoing process aimed at understanding and improving student learning (AAHEA, 2013). The goal of the current assessment process was to report results to faculty to implement appropriate curricular and instructional changes to support and improve student learning.

The Liberal Studies Committee (LSC) is a standing committee of the university's Academic Senate. The LSC has oversight and responsibilities to review, evaluate, and recommend changes or improvements of the Liberal Studies Program based on assessed effectiveness of undergraduates to develop knowledge, skills, and perspectives while progressing through their liberal studies education.

The LSC evaluated one of six different divisions each academic year for assessment purposes. In 2010, the LSC selected Division III Natural Science and Mathematics courses as the assessment focus. Mathematics courses were assessed separately. The current study reports only assessment of student work in Liberal Studies Natural Science courses with a laboratory component.

The LSC developed a plan using Wright's (2003) assessment procedure. First, the LSC identified instructors and science courses within Division III during Spring Semester 2010. Second, the Director of Institutional Research generated a randomly selected student sample from a list of science laboratory courses. Instructors were contacted and provided instructions regarding how to maintain student anonymity when submitting the requested sampling of student work. Meanwhile, the LSC created scoring criteria for assessing science abilities and understandings following guidelines of the AAAA, NNN, NRC, and NSTA, established reporting procedures, and identified an independent evaluation team of three faculty members to score student work with rubric. Each step of the assessment process is described more completely below.

The LSC determined courses in Astronomy, Biology, Chemistry, Environmental Science, Geography, Physics, and Psychology met criteria of Division III Foundations of Natural Sciences during spring 2010. All courses were designed to introduce students to quantitative reasoning and scientific understanding of current views of the natural world. Nearly all courses were introductory courses. Since no courses taught in spring 2010 were approved for advanced Liberal Studies credit, an assessment of the influence of advanced level courses on science inquiry knowledge and understanding was not conducted.

In March 2010, the Chair of the LSC met with the Director of Institutional Research to identify science faculty and instructors of Division III science laboratory courses taught during spring 2010 and create a list of randomly selected students for each identified science laboratory course. Each science instructor received a letter and an email. Natural Science laboratory instructors were also provided a list of Liberal Studies guidelines for Division III math and science courses (see Appendix A).

### Stratified Random Sampling

The LSC and the Office of Institutional Research compiled the population of students enrolled in all Division III Natural Science courses and determined a stratified random sampling of 350 students would provide a confidence level of 95%, which is the confidence level used by the LSC in previous assessments. The Office of Institutional Research's list of randomly selected students represented 8%–10% of students enrolled in each course. Since the list was generated prior to the drop date, some students had dropped the course before collection of student work occurred, which contributed to a return rate of less than 100% of requested student work.

Students were selected by stratified random sampling, which produces an allocated proportion of the total population. For example, if the population consisted of 60% women and 40% men, then three women and two men would reflect proportions of the sample. The LSC reviewed a random sample of about 9% of student work in the Division III Natural Science laboratory courses.

Individual science laboratory courses sometimes consisted of both lecture and lab or lab only. Faculty and instructors decided whether to submit student work from both lecture and lab or lab only because individual students could enroll in lecture and lab concurrently or separately. Many of the selected students in the sample were enrolled only in laboratory sections of a course. All work submitted for each randomly selected student counted as one set of student work or artifact. The Chair of the LSC collected student work and artifacts after finals week. All identifying features of students were removed from their work and artifacts. Student anonymity was maintained.

### Instrument Design

The LRC formed a subcommittee of three members to create a rubric to score criteria for assessing science abilities and understandings based on guidelines from the AAAS, NRC, NSTA, and NNN. Rubrics have long been used to assess student performance using criteria to focus an evaluation with a set of objective external scoring criteria and point-values associated with the criteria by level of performance (Schmoker, 2006). Data from rubrics are used for summative program assessment to compare worthiness of student performance and expected outcomes against external standards (Ebert-May, 2003). Rubrics provide faculty a readily accessible way to quantitatively assess student achievement based on the sum of a range of criteria determined by looking directly at student work (Dodgé & Pickette, 2001).

At the outset, subcommittee members read the AAAS, NRC, NSTA, and NNN guidelines. Each member arrived at the next meeting with an attempt to distill common core competencies into learning outcomes. The subcommittee discussion was facilitated by the Chair of the LSC, who was a professor of earth and space science. The subcommittee reached consensus on five separate learning outcomes based on AAAS, NRC, and NSTA guidelines (see Figure 1). Notably, rubric development did not start with the goals and objectives of the Liberal Studies Division III Natural Science courses. The rubric used science competencies and concepts based on nationally recognized science principles to assess student knowledge of science and scientific inquiry processes. The numerical scoring format was based on recommendations of the NNN

**Rubrics provide faculty a readily accessible way to quantitatively assess student achievement based on the sum of a range of criteria determined by looking directly at student work.**



for “Advancing Assessment of Scientific and Quantitative Reasoning,” which was a National Science Foundation funded project (DUE 0618599) to “further the development of collegiate scientific and quantitative reasoning assessment tools and procedures” (Sundre, Murphy, & Handley, 2009, para. 1).

The subcommittee developed a rubric using a five-point ordinal scale to reflect nuances within the Liberal Studies abilities and understandings of scientific concepts, recognition and use of scientific reasoning methods, understanding and discussion of general scientific articles, and use of mathematics in scientific reasoning and/or problem resolutions. A score of 0 meant the student work completely lacked evidence that the learning outcome was met (e.g., all evidence for the learning outcome was missing). A score of 1 indicated the student work was lacking sufficient evidence to meet the learning outcome (e.g., sporadic, patchy evidence and unfinished or imperfect responses). A score of 2 indicated the rater neither agreed nor disagreed that the outcomes were met and served as a neutral response for cases where a rater could not decide whether the student work did nor did not meet the learning outcome (e.g., perhaps a good start but lacking solid evidence). A score of 3 indicated the rater agreed the learning outcome was met by consistent, sufficient evidence provided by the student work (e.g., recognizes various forms of evidence and uses knowledge of natural phenomena and the physical world). A score of 4 indicated the rater strongly agreed that the student work provided quality evidence that exceeded expectations for the level of the course (e.g., synthesizes well-structured, articulated inquiry processes of natural phenomena and the physical world).

**Changing raw scores to percentages revealed that 27% of student work had evidence to exceed expectations (Strongly Agree) and 32% of student work had evidence to meet expectations (Agree), resulting in 59% of the student work meeting or exceeding expectations.**

Each of the five criteria in the rubric addressed specific scientific processes as defined by the AAAS, NRC, and NSTA. Understanding and use of scientific concepts referred to evidence of use of science knowledge as information in student work. Applying knowledge of science to everyday experiences referred to evidence of the ability to apply science outside of the laboratory to experiences in the natural world. Recognizing and use of scientific reasoning referred to evidence of scientific inquiry process and reasoning skills, which are distinct from the scientific procedures, observations, or concepts. Understanding and discussing general scientific articles required evidence of citations, references, or referrals to science articles, research, or researchers in student work. Use of mathematics in scientific reasoning and/or problem resolutions required evidence of credible use of scientific and mathematical information in scientific developments and public policy issues. Construction of the rubric used “sound assessment methods and practices” (Sundre et al., 2009, para. 2). After creating the rubric, the subcommittee selected three faculty members to form an assessment team to score student work in science laboratory classes.

### Assessment Team

The assessment team was selected using the following criteria: (a) at least one member must teach courses in the Liberal Studies Division III Natural Sciences, (b) at least one member must not teach in Division III, and (c) a third member who may or may not teach in the Division III. Faculty members from Psychology, Biology, and Chemistry formed the assessment team. Team members consisted of voluntary faculty volunteers from departments that offered undergraduate science courses. Members were chosen based on their experience in teaching math and science courses and on their expertise in science knowledge, assessment, and evaluation. Each faculty volunteer received a stipend to work on the assessment team. Assessment team members attended a training session to practice scoring samples of laboratory science work not included in this study. Reviewers completed their review of student work from 350 students in five to eight hours. The LSC chair acted as the coordinator of the assessment team.

### Results

As with earlier collections of student work in other Liberal Studies Divisions, faculty who submitted student samples did so in a timely fashion. Science laboratory faculty and instructors were 79% in compliance with submitting student work, which represents the highest percentage of compliance within the six Liberal Studies divisions. The LSC commended efforts of the Office of Institutional Research staff and of faculty who submitted student work for assessment. All departments represented within Division III turned in student work for the assessment process.

The three raters scored student work from entry-level science laboratory courses using the rubric. Numbers with the symbol # in Table 1 refer to the following rubric criteria:

1. Understanding and use of scientific concepts
2. Application of knowledge of science to everyday experience
3. Recognition and use of scientific reasoning methods
4. Understanding and discussion of general scientific articles
5. Use of mathematics in scientific reasoning and/or problem resolutions

Table 1

*Frequency of Raters' Scores Using Five-point Likert Scaling to Assess Science Abilities and Understandings*

Rating	Rater 1					Rater 2					Rater 3				
	Science Abilities and Understandings					Science Abilities and Understandings					Science Abilities and Understandings				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
SD	2	0	129	309	111	21	24	21	66	95	0	0	1	54	24
D	18	36	47	12	116	13	8	8	5	5	17	24	140	163	201
ND NA	238	141	70	2	21	20	16	16	2	14	21	171	102	12	28
A	124	201	116	69	90	79	27	90	30	14	252	93	51	71	22
SA	14	17	34	4	54	158	216	156	58	153	14	16	10	4	25
Total	396	395	396	396	392	291	291	291	161	281	304	304	304	304	300

*Note.* Abbreviations are as follows: SD = Strongly Disagree, D = Disagree, ND NA = Neither Disagree Nor Agree, A = Agree, and SA = Strongly Agree. The symbol # followed by a number refers to the order of rubric criteria for assessing abilities and understandings of science inquiry processes.

Cohen's Kappa statistic was used to assess the degree to which two or more raters who examine the same ordinal data agree when assigning data to rubric categories. Kappa is a "chance-corrected proportional agreement" (Stawicki, 2010, para. 2) with possible values ranging from +1 (perfect agreement) to 0 (no agreement above that expected by chance) to -1 (complete disagreement). Kappa values were statistically significantly different from 0, suggesting that ratings between coders were largely similar. Table 2 includes the 15 Kappa ratings comparing raters with one another across the five rubric categories. Nine of the 15 Kappa ratings were in the substantial range (0.61–0.80), three were in the moderate (0.41–0.60) range, and three were in the fair range (0.21–0.40; see Landis & Koch, 1977). Cross tabulation reaffirmed that coders largely agreed.

Higher inter-rater agreement occurred in ratings associated with understanding and use of scientific concepts, recognition and use of scientific reasoning methods, understanding and discussion of general scientific articles, and use of mathematics in scientific reasoning and/or problem resolutions (see Table 2). Lower inter-rater agreements occurred in ratings associated with understanding multiple problem-solving perspectives.

## Ratings of Student Work

Summing ratings by similar levels of the rubric (e.g., Strongly Agree) showed higher ratings on student work with evidence of an understanding of current views of natural phenomena, specifically through "Understanding and Use of Scientific Concepts" and "Application of Knowledge of Science to Everyday Experience." Lower ratings occurred on student work with evidence of an "Understanding and Discussion of General Scientific Articles" and "Use of Mathematics in Reasoning and Problem Solving." Coders used all five points of the rubric (see Figure 1).

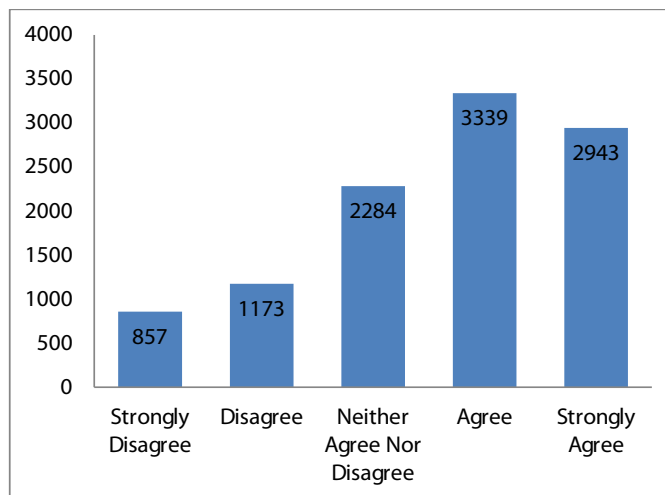
**Table 2**  
*Kappa Calculations between Three Raters on their Assessments of Students' Abilities and Understandings of Science Inquiry Processes*

<i>n</i> = 145	R1 R2	R1 R3	R2 R3
1. Understanding and Use of Scientific Concepts	0.66	0.33	0.64
2. Application of Knowledge of Science to Everyday Experience	0.54	0.23	0.38
3. Recognition and Use of Scientific Reasoning Methods	0.65	0.57	0.65
4. Understanding and Discussion of General Scientific Articles	0.66	0.53	0.65
5. Use of Mathematics in Scientific Reasoning and/or Problem Resolutions	0.77	0.66	0.78

*Note.* Abbreviations R1, R2, and R3 refer to Rater 1, Rater 2, and Rater 3 respectively. The symbol *n* designates number of scores in a randomly selected, limited portion of the total sample of 10,596 scores.

Changing raw scores to percentages revealed that 27% of student work had evidence to exceed expectations (Strongly Agree) and 32% of student work had evidence to meet expectations (Agree), resulting in 59% of the student work meeting or exceeding expectations. Forty-one percent of student work did not provide evidence to meet expectations (i.e., 19% of student work lacked evidence for the criteria and 22% of student work had equivocal evidence).

Likert Scaling	Number of Scores	Percent of Total Scores
Strongly Disagree	857	8%
Disagree	1173	11%
Neither Agree Nor Disagree	2284	22%
Agree	3339	32%
Strongly Agree	2943	27%
	10596	100%



*Figure 1.* Ratings of student work submitted by instructory science laboratory courses as evidence of student knowledge and understanding of science inquiry in Liberal Studies. Results are displayed from left to right according to rubric-guided Likert scaling by number of scores, percentage of total scores, and bar graphs with a display of raw scores.

### Discussion

Liberal education is an approach posited to prepare students to deal with complexity, diversity, and change (Carson, 1997). An assessment of outcomes of a liberal education establishes a baseline to measure practical skills for solving problems in real-world settings (Schneider, 2008) and science competence of all students taking Liberal Studies science laboratory classes, including students typically underserved by the undergraduate learning experience (Seymour, 2002). Creating a baseline of undergraduate knowledge and inquiry processes helps to determine how well the liberal education approach meets its intended outcomes.

Information garnered from the assessment of students' understanding and use of scientific concepts, recognition and use of scientific reasoning methods, understanding and discussion of general scientific articles, and use of mathematics in scientific reasoning and/or problem resolutions taught us at least three important lessons to enhance future practices

in liberal education outcomes assessment. These lessons include maintaining excellent communication practices, developing a valid and reliable rubric for the assessment, and using internal experts to conduct the assessment.

First, maintaining transparency in communications about the process was imperative to gain faculty cooperation with the assessment process. Initially, we used email and phone calls to communicate with faculty about collection procedures of student work and development of the assessment rubric. A faculty-led discussion on the assessment process at an Academic Senate meeting was helpful in garnering faculty support.

Second, science laboratory courses are well suited for performance-based assessments. Students and faculty are familiar with inquiry-based assessments and external science standards allowed the development of a robust rubric based on valid criteria for the assessment process. A concise rubric scoring scale helped to avoid scoring bias and unreliability. The validity and reliability of the process provided a vigorous, easily defensible assessment process.

Third, a committee comprised of faculty from all colleges developed the evaluation rubric, and an assessment team of faculty from diverse science backgrounds conducted the assessment process, both of which added credibility and included an explicit process to avoid scoring bias.

With assessment results in hand, we looked for ways to “close the loop” on how these results are being used to improve student outcomes (see the National Institute for Learning Outcomes Assessment, 2013). Using assessment evidence at department, program, and course levels to make actual improvements in student learning and inform curriculum decisions is challenging (Bailie, Marion, & Whitfield, 2010; Banta & Blaich, 2011). At first, assessment results went directly to department heads to share with faculty of science laboratory courses and the LRC assessment report was posted on the university’s assessment website. No formal reporting mechanism was initially in place to follow whether or how faculty and instructors used assessment information to improve their science laboratory courses or student learning of science inquiry processes. In 2012, a process was initiated as an Academic Quality Improvement Program initiative to have faculty from all disciplines work together in small groups to develop learning outcomes for their syllabi (see Hammock & Richardson, 2011, for a similar process). Science laboratory faculty who attended the workshops developed inquiry-based learning outcomes to provide student data for a continuous improvement feedback loop to assess and refine science-inquiry processes of course content.

The next phase of science laboratory course assessments is slated for 2016. Links to national science standards, the rubric, and a report about the assessment process are on the university assessment website. Discussions are underway to explore the benefits of creating a “connections” type of science laboratory course with a focus on applying/integrating science inquiry processes. Presently, the Liberal Studies Natural Science Division III has two course levels (i.e., 100–200 [Emerging aka “lower division”] and 300–400 [Innovating aka “upper division”]).

## Summary

Evidence of assessment and evaluation are critical to a university’s accreditation processes. We recommend selecting a non-intrusive, statistically defensible, stratified random sampling of student artifacts for the assessment and evaluation process. The method of data collection worked well and met the usual goal of sampling, which is to produce a representative sample. Occasionally, faculty would inquire whether they could submit the “best examples of student work,” rather than submitting the work of randomly selected students. The LSC insisted on conforming to accepted statistical practices on the collection of student artifacts from stratified random samples.

After the assessment, raters gave their feedback on the assessment and evaluation process. They suggested more training on initial ratings of student work samples to hone their skills to automaticity with the scoring rubric.

We advise giving clarifying information to faculty and instructors on how to select examples of student work and artifacts to submit. For example, laboratory reports, papers, essays, and even short answer problem-based items were excellent artifacts for assessing

**Creating a baseline of undergraduate knowledge and inquiry processes helps to determine how well the liberal education approach meets its intended outcomes.**



**Laboratory reports, papers, essays, and even short answer problem-based items were excellent artifacts for assessing science understanding and use of scientific concepts, recognition and use of scientific reasoning methods, understanding and discussion of general scientific articles, and use of mathematics in scientific reasoning and/or problem resolutions.**

science understanding and use of scientific concepts, recognition and use of scientific reasoning methods, understanding and discussion of general scientific articles, and use of mathematics in scientific reasoning and/or problem resolutions. Submitting student grades was of no value to raters for assessing science abilities or understandings and resulted in a rating of zero.

Student work for this assessment was gathered from science laboratory courses taught in Spring Semester 2010 in entry-level science courses. No student work came from advanced courses. Consequently, finding only 28% of student work exceeded expectations is not surprising on an assessment of science knowledge in introductory courses. Comparing our findings to a baseline of TIMSS 2007 results offered insights into trends in student knowledge of science and science processes. When compared to the international median, about 38% of U.S. eighth-graders performed at a high benchmark (28%) or above the advanced benchmark in science (10%; TIMSS 2007). In comparison, 59% of the study's undergraduates performed at expectations in entry-level undergraduate Division III Natural Science courses. For lower performing students, TIMSS 2007 results had 29% of U.S. eighth-graders performing at or below the low benchmark in science. Our raters determined 19% of undergraduates performed below expectations and 22% were approaching expectations.

Our research offers a feasible, systematic, outcomes assessment approach to evaluation of undergraduate science programs. We have honored Wright's (2003) outline of the assessment process and met criteria outlined by Slavin (2008) for a reliable, rigorous, unbiased, and meaningful assessment based on the strength of evidence.

Next steps include using the assessment results of student knowledge and understanding of science inquiry processes to improve teaching and learning in Division III Natural Science courses and to invite other postsecondary institutions to use the rubric to assess student knowledge and understanding of science inquiry processes. The assessment process provides a meaningful measurement and documentation of undergraduates' science learning and offers an opportunity for faculty and instructors to bridge the gap between undergraduate science teaching and student learning of science theory and practice.

Goals for a liberal education include intellectual development and attainment of intellectual skills, broad knowledge, social responsibility, integrative learning, and demonstrated ability to use one's knowledge in real-world contexts (Schneider, 2008). To assess whether the goals of a liberal education have been achieved, college faculty members have a responsibility to evaluate science inquiry learning outcomes of a general education that academic institutions seek to impart to students. The fundamental worth of our assessment method is use of a generalizable stratified random sampling assessment method of student work and an easy to implement and replicate rubric based on nationally recognized science standards and inquiry processes, which strives to rise above the studied scientific knowledge to assess student understanding of the nature of scientific inquiry processes. Such understanding of scientific inquiry processes should transcend particular course knowledge to provide students with greater talents and abilities to solve problems, reason logically, and live rationally.

## References

- American Association for Higher Education and Accreditation. (2013). *AAHE/AAHEA*. Retrieved from <http://www.aahea.org/aahea/>
- American Association for the Advancement of Science. (2013). *AAAS homepage*. Retrieved from <http://www.aaas.org/>
- Assessment Tools in Informal Science. (2011). *ATIS homepage*. Retrieved from <http://www.pearweb.org/atis/>
- Augeri, M., Brents, M., Christiansen, E., Etzenhouser, B., Fox, S., Giese, K. ... van Andel, M. (2011). *Assessment of student learning outcomes at an institutional level*. Retrieved from [www.uiowa.edu/~outcomes/documents/FocusGroups-FINAL.pdf](http://www.uiowa.edu/~outcomes/documents/FocusGroups-FINAL.pdf)
- Bailie F., Marion B., & Whitfield, D. (2010). How rubrics that measure outcomes can complete the assessment loop. *Journal of Computing Sciences in Colleges*, 25(6), 15-28.
- Banta, T. W., & Blauch, C. (2011). Closing the assessment loop. *Change: The Magazine for Higher Learning*, 43(1), 22-27.
- Buxton, C., & Provenzo, E. F. (2011). "Natural Philosophy" as a foundation for science education in an age of high-stakes accountability. *School Science and Mathematics*, 111(2), 47-55. doi:10.1111/j.1949-8594.2010.00060.x
- Carson, R. N. (1997). Science and the ideals of liberal education. *Science & Education*, 6(3), 225-238. doi:10.1080/00091383.2011.538642
- Dodge, B., & Pickette, N. (2001). *Rubrics for web lessons*. Retrieved from <http://webquest.sdsu.edu/rubrics/weblessons.htm>
- Ebert-May, D. (2003). *Classroom assessment techniques: Scoring rubrics*. Retrieved from <http://www.flaguide.org/>
- Ellis, A., Mathieu, B., & Brissenden, G. (2003). *Field-tested Learning Assessment Guide (FLAG)*. Retrieved from <http://www.flaguide.org/>
- Faust, D. (2000). The concept of evidence. *International Journal of Intelligent Systems*, 15, 477-493.
- Hakyolu, H., & Ogan-Bekiroglu, F. (2011). Assessment of students' science knowledge levels and their involvement with argumentation. *International Journal for Cross-Disciplinary Subjects in Education*, 2, 264-270.
- Hammock, G., & Richardson, D. (2011). *Closing the loop: Linking assessment with course design*. SoTL Commons: A Conference for the scholarship of teaching & learning. Abstract retrieved from <http://eaglescholar.georgiasouthern.edu:8080/jspui/handle/10518/3833>
- Hart Research Associates. (2013, April 10). It takes more than a major: Employer priorities for college learning and student success. Retrieved from [http://www.aacu.org/leap/documents/2013\\_EmployerSurvey.pdf](http://www.aacu.org/leap/documents/2013_EmployerSurvey.pdf)
- Humphreys, D. (2013). Success after college: What students, parents, and educators need to know and do. *Liberal Education*, 99(2).
- Kastens, K. A., & Rivet, A. (2008). Multiple modes of inquiry in earth science: Helping students understand the scientific process beyond laboratory experimentation. *The Science Teacher*, 75(1), 26-31.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. doi:10.2307/2529310
- Lord, T., Travis, H., Magill, B., & King, L. (2005). Comparing student-centered and teacher-centered instruction in college biology labs. Retrieved from <http://stemtec.org/pathways/Proceedings/Papers/Lord-p.doc>
- National Institute for Learning Outcomes Assessment. (2013). *NILOA homepage*. Retrieved from [www.learningoutcomeassessment.org/](http://www.learningoutcomeassessment.org/)
- National Academy of Sciences. (2013). *National science education standards*. Washington, DC: National Academy Press. Retrieved from [http://www.nap.edu/openbook.php?record\\_id=4962](http://www.nap.edu/openbook.php?record_id=4962)

- National Science Teachers Association. (2011). *Positions: Official NSTA positions on a range of issues*. Retrieved from <http://www.nsta.org/about/positions.aspx?lid=abt>
- Pingree, S. E. (2007). Bringing theory to practice & liberal education. *Liberal Education*. Retrieved from [http://www.aacu.org/liberaleducation/le-wi07/Le-wi07\\_feature4.cfm](http://www.aacu.org/liberaleducation/le-wi07/Le-wi07_feature4.cfm)
- Rennie, L. J. (1994). Measuring affective outcomes from a visit to a Science Education Centre. *Research in Science Education*, 24(1), 261-269. doi:10.1007/BF02356352
- Resnick, L. B., & Zurawsky, C. (2007). Science education that makes sense. *American Educational Research Association (AERA): Research Points*, 5(1), 1-4.
- Schmoker, M. (2006). *Results NOW: How we can achieve unprecedented improvement in teaching and learning*. Washington, DC: ASCD.
- Schneider, C. G. (2008). Liberal education takes a new turn. *The NEA 2008 Almanac of Higher Education, 2008*, 29-40. Retrieved from [www.nea.org/assets/img/PubAlmanac/ALM\\_08\\_03.pdf](http://www.nea.org/assets/img/PubAlmanac/ALM_08_03.pdf)
- Seymour, E. (2002). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education*, 86(1), 79-105. doi:10.1002/sce.1044
- Seymour, E., Wiese, D., & Hunter, A. (2003). *Classroom assessment techniques: Student assessment of learning gains*. Retrieved from <http://www.flaguide.org/>
- Seymour, E., Wiese, D., Hunter, A., & Daffinrud, S. M. (2000). *Creating a better mousetrap: Online student assessment of their learning gains*. Paper presentation at the National Meeting of the American Chemical Society, San Francisco, CA. Retrieved from <http://www.salgsite.org/docs/SALGPaperPresentationAtACS.pdf>
- Shavelson, R. I., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academies. Retrieved from <http://www.nap.edu>
- Slavin, R. E. (2008). Perspectives on evidence-based research in education--What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5-14. doi:10.3102/0013189X08314117
- Stawicki, S. P. (2010). Kappa – Agreement – Statistic. *Scientific Training and Research*. Retrieved from [http://www.medschoolwiki.org/msw/main/index.php/Kappa\\_-\\_Agreement\\_-\\_Statistic](http://www.medschoolwiki.org/msw/main/index.php/Kappa_-_Agreement_-_Statistic)
- Steedle, J., Kugelmass, H., & Nemeth, A. (2010). What do they measure? Comparing three learning outcomes assessments. *Change: The Magazine of Higher Learning*, 42(4), 33-37. doi:10.1080/00091383.2010.490491
- Sundre, D. L., Murphy, C., & Handley, M. (2009). *Advancing assessment of scientific and quantitative reasoning*. The National Numeracy Network. Retrieved from <http://serc.carleton.edu/nnn/numeracyprojects/examples/32007.html>
- Taylor, A. (2000). The effect of traditional classroom assessment on science learning and understanding of the processes of science. *Journal of Elementary Science Education*, 12(1), 19-32. doi:10.1007/BF03176895
- Trends in International Mathematics and Science Study (TIMSS). (2007). Retrieved from <http://timss.bc.edu/timss2007/index.html>
- Wright, B. D. (2003). *More art than science: The postsecondary assessment movement today*. Retrieved from <http://www.apsanet.org/media/WordFiles/MoreArtThanScience.doc>

## Appendix A

### Email to Science Faculty and Instructors of Science Laboratory Courses

As a professor who teaches a course listed as Division III, you have been selected to be part of the outcomes assessment evaluation. The Liberal Studies Committee will be evaluating your students' work as a part of a programmatic evaluation of liberal studies program. Please provide a sample of your students' work, making sure the sample best demonstrates the liberal studies skills and abilities that students have achieved in your course. Additionally, we need an explanation of how you have assessed your students' work. Examples of students' work could include written papers or essays, projects, tests or final exams. The Liberal Studies Committee decided on this option as possibly the least intrusive method of collection of student work samples. This effort was modeled after successful collection of student work samples from the Division I Humanities, 2006, Upper level Divisions II and IV, 2008 evaluations, and Division V Formal Communications, 2009.

1. How does this course enhance the students' ability recognize and understand the scientific processes?
2. Ability to evaluate various forms of evidence and knowledge
3. Ability to engage in analytical reasoning and
4. How does this course enhance the students' ability to understand and use scientific concepts?
5. How does this course enhance the students' ability to understand and discuss general scientific articles?
6. How does this course enhance the students' ability to apply their knowledge of science to everyday experience?
7. Are the division goals and objectives included as part of the course syllabus?
8. Ability to engage in argumentation and quantitative analysis
9. Ability to engage in scientific inquiry and processes
10. Ability to see across disciplinary boundaries
11. Understanding natural phenomena and the physical world
12. Understanding multiple problem-solving perspectives

## Appendix B

### Natural Science Rubric

Division III Natural Science description is “These courses primarily focus on scientific and quantitative reasoning and understanding the natural world.”

To the Reviewer: Indicate your level of agreement regarding the demonstration of the following components per the learning outcome artifacts reviewed as related to Division III.

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
<i>The learning outcome artifacts of this course (class?) demonstrate <b>Understanding and Use of Scientific Concepts</b></i>	0	1	2	3	4
<i>The learning outcome artifacts of this course (class?) demonstrate <b>Application of Knowledge of Science to Everyday Experience.</b></i>	0	1	2	3	4
<i>The learning outcome artifacts of this course (class?) demonstrate <b>Recognition and Use of Scientific Reasoning Methods.</b></i>	0	1	2	3	4
<i>The learning outcome artifacts of this course (class?) demonstrate <b>Understanding and Discussion of General Scientific Articles.</b></i>	0	1	2	3	4
<i>The learning outcome artifacts of this course (class?) demonstrate <b>Use of Mathematics in Scientific Reasoning and/or Problem Resolutions</b></i>	0	1	2	3	4

<ul style="list-style-type: none"> <li>• Ability to write and communicate clearly and effectively</li> <li>• Ability to evaluate various forms of evidence and knowledge</li> <li>• Ability to engage in analytical reasoning and argumentation</li> <li>• Ability to engage in quantitative analysis</li> <li>• Ability to engage in scientific inquiry and processes</li> <li>• Ability to see across disciplinary boundaries</li> </ul>	<ul style="list-style-type: none"> <li>• Understanding cultural diversity within the United States</li> <li>• Understanding the world as a diverse and interrelated community</li> <li>• Understanding the relationship of the individual to society and its culture and institutions</li> <li>• Understanding the role of the fine and performing arts and the humanities in shaping and expressing a culture’s values and ideals</li> <li>• Understanding natural phenomena and the physical world</li> <li>• Understanding multiple problem-solving perspectives</li> </ul>
--	--