Assessment of science knowledge and learning is centuries old and includes processes such as the Socratic Method (Doran, Lawrenz, & Helgeson, 1994). Within the United States, assessment of science achievement in Kindergarten-12 education has expanded over several decades through state-mandated science achievement tests and outside agencies such as the National Assessment of Educational Progress (NAEP) and the National Center for Education Statistics (NCES), which hosts the closely watched Trends in International Mathematics and Science Study (TIMSS). The call for assessment of students’ science knowledge at the postsecondary level is clear (Wright, 2004), yet colleges and universities face considerable complexities when undertaking science program evaluation, communicating student learning outcomes, and recommending program improvements (Scharmann, James, & Stalheim-Smith, 2004).

**Background**

Various approaches for assessing science programs have used external influences, including a series of national discussions on education about “how best to impart a certain body of knowledge and know-how” (Schulman, 2005). Internal evaluations have been conducted with
teams of students in a capstone course to explore students’ perceptions of science learning (Augeri et al., 2009) in relation to the university’s Liberal Arts core competencies (University of Iowa, 2009). Ideally, the evaluation of science literacy not only would include an understanding of tools and processes for addressing scientific relationships of natural objects and phenomena, but also would involve performance assessments, which could be difficult to implement on a large scale (Buxton & Provenzo, 2011).

One study of performance assessment revealed that, before inquiry-based instruction, students in sixth grade only used prior experiences and perceptions in their explanations on written exams, but after science inquiry experiences, students also utilized the scientific knowledge and concepts learned during the science inquiry intervention and their answers reflected deeper complexity and understanding (Dumbrajs, Helin, Kärkkäinen, & Keinonen, 2011). Classic experimentation, however, is not the only way to help students understand scientific processes. An examination of the relationship between science knowledge and creating argumentation determined that college students’ science knowledge did not predict their contribution to an argumentation (Hakyolu & Ogan-Bekiroglu, 2011). Indeed, no statistically significant difference existed between the science knowledge achievement of undergraduate elementary science methods students under conditions of the presence or absence of traditional classroom tests and quizzes (Taylor, 2000). Multiple approaches to science inquiry exist with modern analogs of active science processes using physical and computer models (Kastens & Rivet, 2008). Web-based Inquiry Science Environment (WISE) modules, for example, use visualizations of thermodynamics, electrostatics, and plate tectonics to guide students to connect scientific ideas when conducting inquiry investigations (Resnick & Zurawsky, 2007).

The American Association for the Advancement of Science and the National Science Teachers Association agree that scientific inquiry is a powerful way of understanding science content. Students must learn how to ask questions and use evidence to answer science questions. In the process of learning the strategies of scientific inquiry, students learn to conduct an investigation and collect evidence from a variety of sources, develop an explanation from the data, and communicate and defend their conclusions. Ultimately, scientific inquiry refers to the activities as helping develop knowledge and understanding of scientific ideas, as well as an understanding of how science is done in the natural world.

The goals for a liberal education include intellectual development and skills, broad knowledge, social responsibility, integrative learning and the demonstrated ability to use one’s knowledge in real-world contexts (Schneider, 2008). College faculty members have a responsibility to evaluate the learning outcomes of the liberal education that institutions seek to impart to students. Procedures for performing an evaluation of the Natural Science division of a Liberal Studies 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).

We drew from a definition of assessment as a process that incorporates five steps (Wright, 2003):

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; and
5) *Repeating* the cycle with interventions and/or new questions about learning.

We used Wright’s (2003) assessment process to evaluate whether science courses in a Liberal Studies Program were able to enhance students’ science abilities and understandings in five areas:

1. Students should improve their ability to understand and use scientific concepts.
2. Students should be able to apply their knowledge of science to everyday experiences.
3. Students should improve their ability to use scientific reasoning.
4. Students should be better able to understand general scientific articles.
5. Students should have an improved ability to think quantitatively and apply mathematics to scientific problem solving.

Within the context of understanding the natural world, laboratory experiences provide students opportunities to interact directly with the material world using the tools, data collection techniques, models, and theories of science. Laboratories allow students to design and conduct investigations to collect the evidence needed to answer a variety of questions, draw conclusions, and think critically and logically in creating explanations based on their evidence. The process allows students an opportunity to communicate and defend their results to their peers and others. The majority of courses assessed in this study had a lab component. The university, however, has a separate requirement for a laboratory science course that is apart from the Liberal Studies Division III Natural Science requirement. This study is limited to courses that quality as Liberal Studies Division III courses taught during Spring Semester 2010.

**Liberal Studies Committee**

The Liberal Studies Committee (LSC) is a standing committee of the Academic Senate and is assigned responsibility for the Liberal Studies component of the undergraduate curriculum. The Committee has oversight of the Liberal Studies Program with major responsibilities to review, evaluate, and recommend changes or improvements of the Liberal Studies Program. The American Association for Higher Education and Accreditation (AAHEA) notes assessment is an on-going process aimed at understanding and improving student learning. The Committee is committed to evaluating the Liberal Studies Program’s effectiveness in developing students' knowledge, skills, and perspectives in their liberal arts education (Klett, 2011). The primary purpose of assessment is to improve student learning by making appropriate curricular and instructional changes. The university’s Academic Quality Improvement Program report for the university had the following quote from the Higher Learning Commission (Prosen, 2010):

“While on campus, the Team found ample evidence that the University is currently operating in a data-rich environment and making good use of that information, including using data to drive decision making. There appears to be good progress being made in the area of developing academic outcome measures for each academic program as well as for
the Liberal Studies core. An ongoing Action Project ‘Documenting and Benchmarking the Outcomes Assessment Process’ will assist in the maturation of a developing system of assessment and improvement of student learning outcomes.’’

The Liberal Studies Program consists of six foundations, also known as divisions. The Liberal Studies Committee has focused on a different division each academic year for assessment purposes. Early in the academic year, the committee selected Division III Natural Science and Mathematics courses as the focus for 2009-2010. As currently configured, Division III also contains mathematics courses. The mathematics courses were assessed separately. In the current study, we report only the assessment of student work in Natural Science courses.

We applied Wright’s (2003) assessment procedure to the Division III Natural Science evaluation. First, we identified instructors and science courses in Division III during Spring Semester 2010. Second, a randomly selected student sample for each course was generated. Third, instructors were contacted and provided with instructions for submitting the requested sampling of student work. Fourth, an assessment rubric was developed, and finally, an Assessment Team was selected. Each of these steps is described in more detail below.

**Procedure**

**Sample Selection**

The Liberal Studies Committee determined courses in Astronomy, Biology, Chemistry, Environmental Science, Geography, Physics, and Psychology all met the criteria of Division III Foundations of Natural Sciences and were taught during Spring Semester 2010. All courses were designed to introduce students to focus on scientific and quantitative reasoning and understanding of the natural world. Nearly all courses were introductory courses. No courses taught in Spring Semester 2010 were approved for advanced Liberal Studies credit, so additional assessments of advanced levels science knowledge and understanding were not conducted.

**Collection of Students’ Science Work**

In March 2010, the Director of Institutional Research Office met with the Chair of the Liberal Studies Committee to identify science faculty and instructors who taught Division III courses during Spring Semester 2010 and to provide a list of selected students for each identified science course. Each science instructor received a letter and an email with the following information:

> 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.

Division III Natural Science instructors also were provided a list of Liberal Studies guidelines for Division III science courses:

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

**Stratified Random Sampling of Undergraduates**

The population size of students enrolled in all Division III Natural Science courses was compiled. Collaboration between the Liberal Studies Committee and the Office of Institutional Research determined a stratified random sampling of around 350 students would provide a confidence level of 95%, which is the confidence level used by the Committee in previous assessments. The Office of Institutional Research generated a list of randomly selected students representing 8%-10% of students enrolled in each course. Because this list was generated prior to the drop date, the list did contain some students who had dropped the course before the 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. As depicted in Figure 1, stratified random sampling uses 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 Liberal Studies Committee reviewed a random sample of about 9% of students’ work in Division III Natural Science courses.
Figure 1. Stratified random sampling consists of an allocated proportion of students enrolled in each course.

Only one set of a student’s work was collected on each randomly selected student. The Liberal Studies Committee left the discretion to individual course instructors whether submitted student work from lecture, lab, or both lecture and lab would be considered, since individual students could enroll in lecture and lab concurrently or separately. Many selected students were only in laboratory sections of a course. The Chair of the Liberal Studies Program collected student work after final exam week. Identifying features were removed.

**Rubric Development**

Rubrics are authentic assessment tools used to measure students' work (Dodge & Pickette, 2001), acting as scoring guides for evaluation of a student's performance based on the sum of a range of criteria, rather than on a single numerical score, and providing a summative assessment. The Liberal Studies Committee formed a sub-committee to develop an assessment rubric for the evaluation of undergraduates’ science understanding and abilities based on student work in Liberal Studies Division III Natural Science courses. The subcommittee researched and followed guidelines from the American Association for the Advancement of Science (AAAS), the National Research Council (NRC), and National Science Teachers Association (NSTA). Based on the AAAS, NRC, and NSTA guidelines, the subcommittee distilled the science competencies into five learning outcomes:

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

These five learning outcomes were compared with the university’s bulletin description of the core competencies expected of students in science laboratory courses. Specifically, students who enroll in science laboratory courses were expected to be *active* in learning the process and strategies of scientific inquiry. Students were expected to demonstrate understandings and
abilities of how to design and conduct investigations to collect evidence from a variety of sources, develop an explanation from the data, and communicate and defend their conclusions. Consequently, student work in science laboratory courses should show evidence of students’ knowledge of science and scientific inquiry processes.

The subcommittee’s final assessment rubric was based on The National Numeracy Network’s recommendations for “Advancing Assessment of Scientific and Quantitative Reasoning”, which was an NSF 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 rubric met Sundre, Murphy, and Handley’s (2009) call for “sound assessment methods and practices” (para. 2). In the rubric, a score of zero (0) meant the rater strongly disagreed that the data provided met the learning outcomes. A score of one (1) indicated the rater disagreed the data had evidence that learning outcomes were met. A score of two (2) meant the rater agreed the data were approaching outcome expectations. A score of three (3) indicated the rater agreed the data met outcome expectations, and a score of four (4) indicated the rater strongly agreed the data had evidence that the quality of the students’ work exceeded outcome expectations.

**Assessment Team**

An Assessment Team was selected according to the criteria set by the Liberal Studies Committee. The criteria included three requests for the Assessment Team: (1) at least one member must teach courses in the Liberal Studies Division III Natural Sciences, (2) at least one member must not teach in Division III, and (3) a third member who may or may not teach in the Division III. Faculty members from Psychology, Biology, and Chemistry formed the Assessment Team.

Members of the Assessment Team were faculty volunteers from departments with undergraduate science courses and were chosen based on experience in teaching mathematics and sciences and for their expertise in science, assessment, and evaluation. Each faculty volunteer received $400 for work on the Assessment Team. The Assessment Team reviewed all submitted materials in about five hours. The Chair of the Liberal Studies Committee acted as the coordinator of the Assessment Team.

**Results**

For the fourth year in a row, the Liberal Studies Committee commended efforts of the Office of Institutional Research staff and of faculty who submitted their students’ work. As with earlier collections of student work in other Liberal Studies Divisions, faculty who submitted student samples did so in a timely fashion.

**Compliance of Faculty and Departments**

The Division III Natural Science faculty members were 79% in compliance with submitting student work, which represents the highest percentage of compliance within the six Liberal Studies Divisions.
All departments within Division III turned in student work for evaluation. Biology, Chemistry, and Psychology had the highest enrollments consisting of 24% of the population.

Inter-rater Reliability

Inter-rater reliability between members of the assessment team was characterized using Kappa statistics. The overall agreement between Raters 1 and 2 and Raters 2 and 3 showed the greatest agreement. Raters 1 and 3 showed the lowest agreement, but agreement was still in the fair to moderate range. The highest agreements were between Raters 1 and 2 and Raters 2 and 3 in 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 which ranked substantial. The lowest agreements, understanding multiple problem-solving perspectives, were between Raters 1 and 2 and Raters 1 and 3. Nevertheless, substantial agreement existed between Raters 2 and 3 on the same question, which achieved an overall inter-rater reliability of 80%.

Outcomes of Undergraduate Science Understanding and Abilities

The highest ratings were on student work that demonstrated an understanding of natural phenomena, specifically through “Understanding and Use of Scientific Concepts” and “Application of Knowledge of Science to Everyday Experience”. The lowest ratings were “Understanding and Discussion of General Scientific Articles” and “Use of Mathematics in Reasoning and Problem Solving”. Understanding multiple problem solving perspectives received mixed reviews.

The raters determined 27% of students’ work provided evidence that met expectations for competencies in science understanding and abilities. As shown in Figure 2, the majority of students demonstrated science understanding and abilities that neared expectations. None of the student work examined exceeded outcome expectations.

Figure 2. Percentages of randomly selected students’ levels of learning science based on their work in Liberal Studies Division III Natural Science introductory-level courses.
Recommendations and Conclusions

Since 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 evaluation. The method of data collection worked well. Occasionally, faculty would inquire whether they could submit the “best examples of student work”, rather than submitting the work of randomly selected students. The Liberal Studies Committee insisted on artifacts of students from stratified random samples.

Although streamlining the scoring process made the evaluation process faster, raters were surprised at the level of simplification and the limitations of the scoring rubrics. Raters noted more initial training would have been helpful. To help raters evaluate student artifacts efficiently, we recommend giving faculty and instructors some additional information on how to submit examples of students’ work. For example, laboratory reports, papers, essays, and short answer problem-based items were excellent artifacts for evaluating science processes. Submitting student grades was of no value to raters for evaluating science abilities or science understandings.

Departments received feedback on the Liberal Studies Division III results. The highest ratings of student work were on understanding and using scientific concepts and applying science knowledge to everyday experience. The lowest ratings were on using and discussing general scientific articles and using mathematics in reasoning and problem solving. Using multiple problem solving perspectives received mixed reviews. Faculty and instructors were encouraged to use the results for improvements in their courses and in student performance. We plan to repeat the assessment of Division III Natural Science courses in 2014.

Outcomes for Knowledge of Natural Science

All of the students’ work came from courses taught in Spring Semester 2010. No students’ work was from an advanced science course. Consequently, finding 0% of the ratings did not exceed expectations is not surprising for introductory science courses.

Comparing our results with TIMSS 2007 results offers some insight into the pattern of trends in students’ knowledge of science and science processes. When compared to the international median (TIMSS, 2007), about 38% of U.S. eighth-graders performed at a high benchmark (28%) or above the advanced benchmark in science (10%). In comparison, 27% of undergraduate students performed at expectations.

The TIMSS 2007 results had 29% of U.S. eighth-graders performing at or below the low benchmark in science. Our raters determined 17% of the samples performed below expectations and 56% of the samples were approaching expectations. In addition, 17% of the students’ grades by instructors did not meet expectations (i.e., grades of D and F).
Assessment of Undergraduate Science Outcomes

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 (2007) for a reliable, rigorous, unbiased, and meaningful assessment based on the strength of evidence. The next steps include assessing use of the results for improvements in teaching for students’ science learning in Division III Natural Science courses. The assessment process provided measurement and documentation of the level of meaningful science learning of undergraduates and offers an opportunity for faculty and instructors to bridge the gap between undergraduate science teaching and students’ learning of science theory and practice.

References


