Learning about Learning in Computational Science and Science, Technology, Engineering and Mathematics (STEM) Education

A White Paper

Jamie Cromack, PhD
Microsoft Research

Wilhemina Savenye, PhD
Arizona State University

ABSTRACT

In conjunction with External Research & Program's (ER&P) continuing efforts to support development of curricula that make learning compelling, relevant and vibrant, ER&P also seeks, through strategic collaborations with leading academics, to identify pedagogical practices that result in improved learning and increased retention. ER&P is developing a support framework to assist faculty as they engage in assessment activities, including the *Microsoft Research Assessment Toolkit*. This white paper covers the nature of assessment practice, summarizes the fundamental types of assessment and offers examples of assessment practice that work well in computational science and other STEM education settings at the postsecondary level. Finally, a case study of effective assessment at Georgia Institute of Technology is presented.

INTRODUCTION

External Research & Program's (ER&P) central mission of stimulating innovation and research plays out in the hallways of postsecondary institutions worldwide with technology solutions to improve teaching and learning. ER&P, through internal efforts and funded projects, collaborates with faculty to investigate innovative uses of computing to enhance education. Historically, programs concentrated on technologies such as the Tablet PC, ConferenceXP, gaming and robotics; current emphasis has expanded to include interdisciplinary computational education. More information on ER&P's programs is available at http://research.microsoft.com/erp/.

The logical question derived from these curricular ventures asks how to measure the impact of these endeavors. External Research & Program is currently investigating the effect of these technological explorations in education. In conjunction with its continuing efforts to support development of curricula that make learning compelling, relevant and vibrant, ER&P also seeks to partner and collaborate with leading academics in the field of educational assessment. To underscore this commitment to learning, ER&P fosters assessment in field-based projects driven by academic Principal Investigators (PIs), and is developing a support framework to assist faculty as they engage in assessment activities. Part of that framework is the *Microsoft Research Assessment Toolkit*.

The Microsoft Research Assessment Toolkit was conceived to provide higher education faculty members with user-friendly, focused resources to assist in developing and conducting rigorous educational assessment. The Toolkit includes a variety of items to facilitate the planning and implementation process, including

1) An overview of assessment and assessment planning guide.
2) User self-assessment tool to help faculty determine their level of assessment experience. The components of the kit are coded to allow users to find tools and strategies appropriate for their needs.
3) Frequently Asked Questions (FAQ) that explain assessment issues with minimal use of jargon.

4) Case studies of effective assessment that provide a snapshot of faculty in various computational science or STEM disciplines.
5) A guide to selecting assessment tools.
6) Annotated resources primarily focused on computational science and STEM disciplines and organized by topical categories.
7) Web links to various assessment-related sites or organizations.

Development of the Toolkit will be an ongoing effort; Toolkit materials are available online at http://research.microsoft.com/erp/AssessmentToolkit.

This white paper addresses a basic starting point for educational assessment by promoting understanding of the nature of assessment practice, summarizing the fundamental types of assessment and offering examples of assessment practice that work well in computational science and other STEM education settings. This paper utilizes a conceptual framework that embraces the paradigm of good practice in assessment outlined by the American Association of Higher Education (AAHE), that, as a "systematic process of gathering, interpreting and using information about student learning, assessment is a powerful tool for educational improvement" (American Association of Higher Education[1], 1997, p. 1). Emphasis will be placed on assessment methodologies used to gather data with which faculty may make teaching and learning decisions in their courses. Examples will focus on what works in computational science and STEM disciplines; faculty from other fields will likely find the discussion applicable to their own domains, too. This paper will first address the academic backdrop for increased pressures to assess and then describe multiple approaches to conducting effective assessment. Finally, a case study of assessment practice in a media computation course at Georgia Institute of Technology provides a real-world example of how one professor implements educational assessment.

PROBLEM STATEMENT

Historically, a university education brought to mind musty lecture halls filled with earnest students frantically struggling to transcribe the sacred utterances of learned professors (Cohen[7], 1998; Lucas[16], 1994; Kimball[14], 2004). This teacher-centric pedagogy remained firmly entrenched in higher education practice for hundreds of years, with grades as the principal measure of student learning (Popham[18], 1993). Today's classrooms often take much different forms; modern-day teaching and learning takes place in classrooms, laboratories, in the field and online, with small and large groups of students, and with one or many teachers. Assessment and evaluation entail more than identifying the sometimes incremental changes in students' final grades – although this is a common misconception. The words "assessment" and "evaluation" evoke frequently conflicting meanings among educational decision makers; to make matters worse, the terms have no universally accepted definitions (Popham[18], 1993; Fitzpatrick, Sanders, & Worthen[10], 2004). The associated jargon of assessment and evaluation further confuse the issue. Is there a difference between a learning goal and a learning objective? How about performance criteria? Metrics? Outcomes?

Rogers[19] (2002) emphasizes that meaning should take precedence over terminology, because many assessment terms are used differently or interchangeably. She warns against becoming mired in debate about specific words at the expense of focus on the underlying questions. Many education professionals distinguish between the two concepts by defining educational assessment as "data-gathering strategies, analysis and reporting processes that provide information that can be used to determine whether or not intended outcomes are being achieved" (Foundation Coalition[11], 2007, p. 1) and educational evaluation as a process of using the information gathered through assessment to "support decisions on maintaining, changing, or discarding instructional or programmatic practices" (Foundation Coalition[11], 2007, p. 1).

Many factors influence the nation's increased emphasis on assessment and evaluation. Accreditation requirements, calls for public accountability, increased education demands with decreased funding and changes in instructional design and delivery have necessitated gathering more information on the effectiveness of teaching and learning (Ewell[8], 1998). Accreditation boards such as ABET include in their decision-making process whether an institution has clearly articulated learning outcomes or objectives and if progress toward achievement is being made (Lattuca, Terenzini, & Volkwein[15], 2006; Felder & Brent[9], 2003). A majority of state governments in America require public institutions to "establish local, campus-based approaches to assessing and reporting student performance" (Ewell[8], 1998, p. 108). Decreased state funding, increased costs and burgeoning enrollments demand more efficiency on the part of institutions as they endeavor to provide educational options for students. As the landscape of learning research expands, the impetus for clinging to less-effective pedagogies such as lecture erodes. When teaching transforms, new ways of documenting student learning must also evolve (Huba & Freed[13], 2000; Ewell[8], 1998; Clough & Kauffman[6], 1999).

These external factors pressure faculty to adopt more rigorous assessment plans for their courses and curricula. Intrinsic factors also influence faculty to implement assessment measures; the desire to improve the educational experience for their students motivates many faculty to collect teaching and learning data. For faculty unsure of how to move beyond a few basic assessment practices, the quagmire of available resources online can seem insurmountable. The following information can be used as a starting point for the discovery process, and is intended to provide a contextual overview of assessment terminology and issues. This content also complements use of the Microsoft Research Assessment Toolkit.

ASSESSMENT SOLUTIONS

Type the word “assessment” into a search engine, and overwhelming millions of results flood the user with options. Even if faculty had time to wade through a fraction of the available web pages, how would it be possible to determine what information applies to a particular classroom setting or academic need? Specialization in a content area rarely includes formal preparation to teach that content or assess the ensuing learning; the Ph.D. historically has “never been a degree primarily aimed at preparing people to teach (Bogue & Aper[4], 2000, p. 159). Most faculty members develop teaching skills through time spent in the classroom and by following the example of preceding instructors (Hartman[12], 2001). If assessment did not play an integral part of previous teaching and no development activities focused on the assessment of learning have occurred, it is unrealistic to expect faculties to show expertise in the data gathering and analysis techniques that comprise educational assessment. So how can a faculty member receiving pressure to implement more assessment into the classroom move forward?

Getting Started

Good assessment equates with good teaching, and when faculties use assessment to guide instructional planning, students benefit (National Council of Teachers of Mathematics[17], 1995). Clearly articulated goals should also inform assessment project planning. Is the intent to improve learning (formative assessment) or provide data for grading (summative assessment)? Will students self-report the data (indirect assessment) or will they be required to demonstrate knowledge or skills (direct assessment)? Taking into account the context of the course, a reasonably balanced approach to assessment involves both formative and summative assessment and both indirect and direct assessment. Use of multiple assessment measures also provides validation for evidence of student learning (Rogers[20], 2006a). The items listed below do not comprise an exhaustive list of types of assessment; other effective and valuable means of assessment exist and are also in use by faculties in higher education institutions. The examples are included to help illustrate the assessment categories and to offer suggestions to faculty members on ways to implement the various techniques.

Summative Assessment of Learning

Summative assessment promotes accountability and quantifies levels of learning. Activities quantify students' progress against curricular objectives; this type of assessment is often comprehensive of a content unit, course or curricula and generally results in a grade.

Final grades – Final grades constitute the most commonly known summative assessment. It is important to note that grades commonly reflect performance in multiple content areas and should not be used as a measure of achievement of individual student learning outcomes. A single letter-grade eliminates the context of a student's achievement and provides little useful information for improvement (Atkin, Black, & Coffey[3], 2001).

Capstone projects – Capstone projects generally represent the culmination of a student's experiences in a degree program, have some relevance to their degree and career interests and are often applied in nature. Assessment can focus on higher-order thinking and problem-solving skills, ability to work in groups or design experiments, communication skills, writing skills, project management and planning or any related skill that demonstrates attainment of the desired learning outcomes for the curricula. The assessment process may also involve a committee of faculty members who determine the student's eligibility to graduate.

Practicum or Internship – The purpose of the practicum or internship is to provide students with supervised experience in an actual work setting similar to one in which the student might end up after graduating. Assessment generally involves analysis of work performance by the student's supervisor, grading of a formal report by the faculty member, and a self-analysis written by the student.

Formative Assessment of Learning

Formative assessments focus on improvement within the scope of the current course. Classroom Assessment Techniques (CATs) are formative activities designed to ascertain what and how well students are learning (Angelo & Cross[2], 1993). These techniques focus on learning, are directed by the instructor, benefit both the instructor and students, emphasize improvement rather than grades, are context-specific to the classroom in which they occur, take place on an on-going basis and are rooted in good teaching practice.

The following Classroom Assessment Techniques (taken from Angelo & Cross's estimable text of the same name) have proven effective for computational science and STEM classrooms (for more examples, please refer to the CATs book).

Background Knowledge Probe – This activity goes beyond the common practice of asking students what courses they have already taken in the field. Using a survey, the instructor elicits information that can be used to focus instruction on appropriate content and level of difficulty. The questions address information the students will need to know to succeed on course assignments and activities, include both easy and difficult questions, and avoid general knowledge areas. A more advanced application of the technique could entail implementing the survey pre- and post-course.

Pro & Con Grid – This technique helps faculty determine how well students can imagine more than one side to an issue by having them develop lists of pros/cons, costs/benefits, advantages/disadvantages to an issue. The instructor identifies a decision/judgment/issue that has relevance in the context of the course and writes out a question that will elicit thoughtful responses to the topic. A more advanced application of the technique could involve role playing as students choose a character (project manager, programmer, technical artist, etc.) and answer the question.

Student-generated Test Questions – By having students write test questions and compose answers, faculty discover what students identify as key content, what they consider reasonable test questions and how well they can answer the questions they create. Instructors pre-determine the types of questions (essay, multiple choice, short-answer, etc.) and the topics to be addressed. The questions can then be compiled for a study guide or, as an added incentive, chosen (if suitable) to appear on the actual test.

Minute Paper – This popular CAT helps ascertain what students felt was the most important information they learned during a particular class meeting and if they have any lingering questions about the content. Answers to these questions help faculty focus instruction, make mid-course corrections or identify areas that need more emphasis.

Muddiest Point – This very simple technique identifies areas of confusion from a lecture, discussion, homework or other activity. When students write out the answer to the question, "What was the muddiest point in _______?" they not only must reflect on the content material but also articulate their thoughts. This CAT works well when large amounts of information has been presented.

Direct Assessment of Learning

When planning direct assessments, the first step is to define desired outcomes—what do you want students to know, be able to do or demonstrate, and at what level of proficiency (Brualdi[5], 1998)? Next, consider the activity— what is the timeline, what resources are available, and what data are needed to make a judgment on the student's performance? The third step is to determine what constitutes success for the student, and finally, to create grading rubrics. Rubrics, simply stated, are the "scoring rules" that allow more objective grading by identifying criteria of performance at different levels of mastery (Huba & Freed[13], 2000).

Direct assessment of student learning "provides for the direct examination or observation of student knowledge or skills against measurable learning outcomes" (Rogers[20], 2006a, p. 3). Faculty will find many of the learning activities that result in direct assessment familiar.

Grades – It is impossible to make assumptions about a student's knowledge or abilities from a single grade, because "many factors contribute to an assigned grade" (Rogers[22], 2006c, p. 3). A grade on an exam or in a course generally represents aggregate learning of multiple concepts or skills and is reflective of the instructor's priorities and expertise. Additionally, grades are often curved or include contributing factors not indicative of learning (Rogers[22], 2006c).

Portfolios – A student portfolio provides evidence of student achievement through a grouping of selected works that can be evaluated for specific competencies. Rogers[21] (2006b) lists several concerns about using portfolios to assess student learning. Faculty should determine whether the assessment will be focused on individual students or student cohorts and establish ahead of time what measurable performance criteria will be employed. The process of assessing the portfolios should be spelled out, too. If more than one person will be analyzing the materials, issues of consistency in rating and incentives for participation need to be addressed.

Embedded concepts – This technique allows faculty to determine whether students are attaining course or curricular objectives. By embedding assessment of a particular concept throughout the instruction process, it becomes possible to document change in learning by a student or group of students.

Indirect Assessment of Learning

Indirect assessment of learning involves a process of self-reflection by students. In contrast with direct assessment, which deals with the observable evidence of student learning, indirect assessment relies on students reporting on their own attitudes, reactions or behaviors. This type of assessment is not considered as strong because "assumptions must be made about what exactly the self-report means" (Rogers[20], 2006a, p. 3). However, because direct measures cannot assess areas such as attitudes or values, indirect assessment comprises a valuable part of the overall assessment picture. Rogers[20] argues for the use of both types of assessment measures to create a "meaningful assessment program" (p. 3).

Surveys, questionnaires – Well-designed surveys can help students reflect on their learning and offer insight into student knowledge and skills, in addition to satisfaction data. Results from student surveys can be used to triangulate with direct assessment data to provide a rich, comprehensive picture of student learning. Surveys can be implemented before, during or after the course to elicit feedback that can be used for improvement of current or future courses. Quantitative questions should avoid biased-sounding prompts; open-ended, qualitative survey questions should be specific enough to provide useful information.

Electronic Mail Feedback – This classroom assessment technique provides a way for instructors to receive feedback from students who may be reluctant to respond to face-to-face questions (Angelo & Cross[2], 1993). The approach is useful for diagnosing problems or soliciting reactions to stressful situations. Questions faculty may ask include, "What is one specific, small change I could make that would help you learn more effectively in this class?" or "If you were the teacher of this class, what would you do to make the ____ assignments more useful?" (p. 328).

Course-related self-confidence surveys – This example of indirect assessment helps pinpoint student confidence in a domain-specific skill or ability (Angelo & Cross[2], 1993). Faculty can use this information to structure assignments that will help build confidence; this assessment is effective in courses where students are learning new skills and works well when employed before and after the new material is introduced. Faculty should identify the key elements of learning and ask students some variation of, “How confident do you feel about _____?” Response options can include *Very*, *Somewhat, Not Very*, and *Not at All*.

MEDIA COMPUTATION AT GEORGIA INSTITUTE OF TECHNOLOGY

A Case Study in Effective Assessment

Mark Guzdial, a professor in the College of Computing at Georgia Institute of Technology, is engaged in a comprehensive program of design, development and assessment of a media computation course. He has developed, and allows other faculty to share and adapt, a series of survey instruments and interview guides to determine the impact of the course development work. He measures, in particular, student attitudes, engagement, attrition and retention. The results of this assessment work have been published widely. The general goal of Guzdial’s media computation project is to build an engaging introductory CS sequence, as reflected in higher retention and a decreased rate of course withdrawals.

The course covers content defined by the ACM and IEEE computing standards (2003). Guzdial generally focuses on improving the engagement of students, and measures their attitudes toward the course and course processes at several stages throughout, in order to determine impact. He is also interested in student retention and determining longer-term attitudes about the class and about computer science activities. Guzdial notes that his assessment techniques and tools are being adapted by other faculty in CS departments in various types of institutions.

Guzdial has been using a comprehensive general survey approach, with some interviews, to assess the effects of the CS1315 media computation course. He suggests faculty consider using surveys at various points throughout the course, including Initial Participant Surveys, Mid-Term Surveys and Final Participant Surveys. Guzdial also implemented a Longitudinal Survey that he used as a follow-up measure a year after he first offered the CS1315 course. Finally, he developed an Interview Guide he also uses as part of the research. Overall, Guzdial has found students’ attitudes to be favorable, and has been able to utilize survey results both formatively, to improve the course as it proceeds, and summatively, to improve the course over several semesters.

More information about Guzdial's assessment approach and tools are available on his "Assessment Tools" web page at http://coweb.cc.gatech.edu/mediaComp-teach/16. The tools may be adapted for use by other faculty. For plans and ongoing work on the media computation classes see http://coweb.cc.gatech.edu/mediaComp-plan.

CONCLUSION

All teaching faculty conduct educational assessment of some type. Good assessment provides faculty with the information required to make informed instructional and curricular decisions, improve teaching and assign grades. Because most faculty members do not receive instruction on assessment practices alongside disciplinary content, user-friendly, focused tools and resources are needed to assist with educational assessment planning and implementation. External Research & Programs supports faculty in their efforts by providing the Microsoft Research Assessment Toolkit, a simple and comprehensive collection of materials to help faculty develop and carry out rigorous, balanced plans for assessing student learning, available as part of ER&P's larger support framework.

Bibliography

1. American Association of Higher Education (AAHE). (1997). *Principles of Good Practice for Assessing Student Learning*. Retrieved June 1, 2007, from DePaul University Academic Affairs Office of Faculty Development & Research: http://condor.depaul.edu/~acafflpc/aahe.htm

2. Angelo, T. A., & Cross, K. P. (1993). *Classroom Assessment Techniques, 2nd Edition.* San Francisco: Jossey-Bass.

3. Atkin, J. M., Black, P., & Coffey, J. (. (2001). *The Relationship Between Formative and Summative Assessment--In the Classroom and Beyond*. Retrieved June 24, 2007, from Classroom Assessment and the National Science Education Standards: http://books.nap.edu/html/classroom_assessment/index.html

4. Bogue, E. G., & Aper, J. (2000). *Exploring the heritage of American higher education: The evolution of philosophy and policy.* Phoenix, AZ: Oryx Press.

5. Brualdi, A. (1998). *Implementing performance assessment in the classroom.* Retrieved June 25, 2007, from Practical Assessment, Research & Evaluation, 6(2): http://pareonline.net/getvn.asp?v=6&n=2

6. Clough, M. P., & Kauffman, K. J. (1999, October). Improving Engineering Education: A Research-Based Framework for Teaching. *Journal of Engineering Education* , pp. 527-534.

7. Cohen, A. M. (1998). *The shaping of American higher education: Emergence and growth of the contemporary system.* San Fransciso: Jossey-Bass.

8. Ewell, P. T. (1998, April). National Trends in Assessming Student Learning. *Journal of Engineering Education* , pp. 107-113.

9. Felder, R. M., & Brent, R. (2003, January). Designing and teaching Courses to Stisfy the BET Engineering Criteria. *Journal of Engineering Education* , pp. 7-25.

10. Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines, 3rd ed.* Boston: Pearson.

11. Foundation Coalition. (2007, January 24). *Assessment and Evaluation*. Retrieved January 24, 2007, from http://www.foundationcoalition.org/home/keycomponents/assessment_eval/AE_sec.html

12. Hartman, H. J. (2001). Teaching metacognitively. In R. M. Joshi (Series Ed.), & H. J. Hartman (Vol. Ed.), *Neuropsychology and cognition, Vol. 19: Metacognition in learning and instruction* (pp. 149-172). Dordrecht, The Netherlands: Kluwer Academic Publishers.

13. Huba, M. E., & Freed, J. E. (2000). *Learner-Centered Assessment on College Campuses.* Needham Heights, MA: Allyn and Bacon.

14. Kimball, B. A. (2004, Summer). Christopher Langdell: The case of an "Abomination" in teaching practice. *Thought & Action* , pp. 23-28.

15. Lattuca, L. R., Terenzini, P. T., & Volkwein, J. F. (2006). *Engineering Change: A study of the Impact of EC2000.* Baltimore: ABET.

16. Lucas, C. J. (1994). *American higher education.* New York: St. Martin's Press.

17. National Council of Teachers of Mathematics. (1995). *Executive Summary: Principles and Standards for School Mathematics.* Retrieved June 6, 2007, from Principles and Standards for School Mathematics: http://www.nctm.org/uploadedFiles/Math_Standards/12752_exec_pssm.pdf

18. Popham, W. J. (1993). *Educational evaluation.* Needham Heights: Allyn & Bacon.

19. Rogers, G. M. (2002). *The Language of Assessment: Humpty Dumpty Had a Great Fall.* Retrieved June 1, 2007, from ABET: http://www.abet.org/Linked%20Documents-UPDATE/Assessment/Assessment%20Tips3.pdf

20. Rogers, G. M. (2006a, August). *Direct and Indirect Assessments: What are they good for?* Retrieved June 16, 2007, from ABET Community Matters: Assessment 101- Assessment Tips with Gloria Rogers, Ph.D.: http://www.abet.org/Linked%20Documents-UPDATE/Newsletters/06-08-CM.pdf

21. Rogers, G. M. (2006b, October). *Got Porfolios?* Retrieved June 16, 2007, from Assessment 101: Assessment Tips with Gloria Rogers, PhD: http://www.abet.org/Linked%20Documents-UPDATE/Newsletters/07-04-CM.pdf

22. Rogers, G. M. (2006c, December). *Using Course or test Grades for Program Assessment.* Retrieved June 16, 2007, from ABET Community Matters: Assessment 101 - Assessment Tip with Gloria Rogers, PhD: http://www.abet.org/Linked%20Documents-UPDATE/Newsletters/06-12-CM.pdf