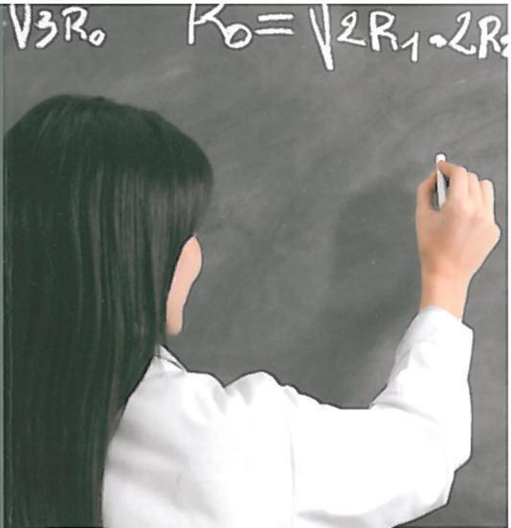


Assessment in the Disciplines
Volume 5

Assessment of Chemistry

Edited by
John Ryan, Ted Clark, and
Alexis Collier

Assessment of Chemistry

THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

© 2010 Association for Institutional Research
1435 E. Piedmont Drive, Suite 211
Tallahassee, FL 32308

All Rights Reserved

No portion of this book may
be reproduced by any process,
stored in a retrieval system, or
transmitted in any form, or by any
means, without the express written
permission of the publisher

Printed in the United States

ISBN 978-1-882393-19-0

Table of Contents

FOREWORD

This volume is the fifth in a series sponsored by the Association for Institutional Research (AIR) focused on assessment in the disciplines. The first year was dedicated to employing assessment in the teaching of business, the second year to the teaching of mathematics and related fields, the third year to the best practices for assessment of engineering, and the fourth year to assessment of writing. The next volume will focus on assessment of arts- and design-related fields of study.

Each of the volumes in this series has reflected both the culture of the profession and the personalities of the authors and editors, as might be expected, and this one is certainly no exception. One can detect in the following pages some of the struggles of the chemistry professoriate as it has grappled with, for example, the difficulties of teaching both its own majors and large numbers of nonmajors such as engineering and premedical and other biology-related students in lower division courses. At the same time, one can also see some of the pedagogical solutions that have been adopted and proven to be successful through creative use of disciplinary and interdisciplinary adaptation.

Of special note in this volume should be the fact that the editors are from three divisions of one university: John Ryan, the lead editor, is an institutional researcher working in the assessment arena; Ted Clark is a chemist; and Alexis Collier is a psychologist. Likewise many of the chapters, though written by chemists or chemistry educators, have contributions from other learning experts also. This richness of interdisciplinary interaction among the contributors helps make this volume stand out from others in the field. As a result, the lessons learned from it can be applied immediately.

Thanks to the Publications Committee of AIR for its continued support of this series and for all of the staff in the Executive Office who have provided assistance in producing it. Volumes such as this are a large team effort; much of the team goes unrecognized.

John A. Muffo
John A. Muffo and Associates, Inc.

EDITORS' PROFILES

John Ryan

and

Ted Clark

and

Alexis Collier

John Ryan is the Director of Assessment for the College of Education and Human Ecology at Ohio State University. His prior experience includes serving as an Assistant Provost and Associate Director of Institutional Research and Planning at Ohio State University as well as roles in student affairs and as a budget/policy analyst for the Ohio Legislative Budget Office. His publications and areas of interest focus on higher education policy and administration, student learning and persistence, and innovative approaches to improved data analysis and decision-support in colleges and universities. He also serves as a member of the Association for Institutional Research Publication Committee and as a member of the editorial board for the Enrollment Management Journal. He received his B.A. in political science with honors from Ohio State, his master's degree in political science from Ohio State, and his Ph.D. in higher education administration from the University of Nebraska-Lincoln.

Ted Clark is an Assistant Professor in the Chemistry Department at Ohio State University and Associate Director of the Ohio Consortium for Undergraduate Research-Research Experiences to Enhance Learning (OCUR-REEL) program. He earned his Ph.D. from the University of Michigan in the area of solid-state inorganic chemistry, completed a postdoctoral research position in the area of solid-state nuclear magnetic resonance spectroscopy at Ohio State, and has taught undergraduate general, analytical, and physical chemistry courses for more than a dozen years. Currently, he is involved in course and curricular development, environmental chemistry research, and the integration of technology in chemistry courses. An overarching theme of these interests is the inclusion of authentic in-class research experiences in undergraduate courses and an assessment of such experiences.

Alexis Collier is an Associate Provost and Associate Professor of Psychology at Ohio State University and has taught both graduate and undergraduate courses, including those designed to promote student development of independent research proposals and projects in integrative honors and laboratory courses. She coordinated the department's General Psychology 100 Instructional Program for 15 years and also received the OSU Department of Psychology Distinguished Teaching Award in 2000. Her research has been in the areas of learning, motivation, and memory with an integrating theme of age-dependent considerations in assessing cognitive and memorial processes. She has served as a grant review panel member for the National Institute of Mental Health and as consulting editor for the American Psychological Association's Journal Supplement Abstract Service. Dr. Collier served as a Provost Faculty Fellow in 1998–1999 and coordinated assessment across the Colleges of the Arts and Sciences during 2004–2007. She currently coordinates outcomes assessment activities university-wide as part of ongoing institutional initiatives to enhance the teaching and learning environment. She received her B.S. in psychology with distinction from Virginia Tech in 1973, and her Ph.D. in experimental psychology from the University of Washington in 1976, specializing in learning and motivation.

CHAPTER 1

ASSESSMENT AS A STRATEGY TO ENHANCE 21ST CENTURY CHEMISTRY EDUCATION

Ted Clark, Alexis Collier, and John Ryan
Ohio State University

In tandem with mathematics and other scientific fields of study, chemistry arguably represents one of the most challenging areas of study for students in higher education. In addition to its own inherent value as a discipline and a professional field, chemistry also serves as an important prerequisite to many "STEMM" (science, technology, engineering, mathematics, and medicine) programs and careers. The extent to which students have an interest in chemistry and are successful in developing the knowledge, skills, and affective dispositions and habits that lead to academic success in chemistry courses and programs, the greater both private and public returns on investments in chemistry education will be.

Beyond the provision of additional resources, innovations in chemistry teaching and learning are an essential component of a comprehensive strategy to enhance interest, skills, knowledge and performance in STEMM fields. Among these innovations, assessment represents one of the most important strategies that educators and postsecondary institutions can leverage to enhance the quality and quantity of chemistry knowledge and skills. The application of the scientific method, as well as action research and inquiry—skills that colleges and universities also seek to develop in their students—to the evaluation and assessment of instructional practices, experiences, and learning outcomes in chemistry provides a familiar approach for chemistry faculty, instructors, and programs. This approach to the assessment of student learning and success in chemistry can serve as a means to enhance student learning as well as program effectiveness and impact. In times of budget cuts and fiscal uncertainty, assessment and curricular innovations that address both learning and cost issues in delivering chemistry courses may possess an even stronger appeal for chemistry programs and their stakeholders.

Given the current economic and educational circumstances facing the United States, the contributions to this volume are apropos and timely. The authors, all faculty members and instructors in chemistry, provide a broad range of experiences, approaches, findings, and lessons learned—all from a disciplinary perspective. They also represent a diverse set of institutional contexts within which assessment must occur in order to improve student learning and academic achievement in chemistry. These contexts include public, private, two-year, four-year, research extensive, undergraduate-focused, and liberal arts-oriented colleges and universities.

Beyond the array of perspectives by institutional type, the chapters also represent programs and faculty at different stages of assessment development that reflect different rationales for conducting assessment, including a recognition of poor student performance, high student attrition, and concern with access and opportunity among under-represented student populations. The chapters capture faculty and programs at various stages in the evolution of a culture of assessment and present diverse perspectives on topics such as the role of common, standardized assessments in chemistry, the role of locally developed instruments, issues of assessment research design, and the need to assess both content and affective dimensions of learning and development. In addition, they represent a range of perspectives and strategies employed based on student population, including a view of practices in high school chemistry assessment, a key component of the chemistry pipeline linking K-12 and postsecondary sectors.

In the midst of the diverse array of ideas and perspectives offered in this volume, all chapters are united by two primary motives—to enhance their students' learning and success in chemistry and, in turn, to share helpful insights and lessons with other chemistry faculty across the country that will assist them in doing the same. Each contributor supports the primary goal of this volume to provide detailed presentations and analyses of real cases that can be adapted in a variety of contexts to enhance the impact of chemistry instructors and faculty on their students' learning and to increase our collective understanding of how students "learn chemistry" and how assessment can be employed as a strategy to make substantive improvements in instruction and outcomes.

In Chapter 2, Ted Clark discusses important innovations taking place in an introductory chemistry instruction via research experiences at a large, public research university and in the Ohio Consortium for Undergraduate Research: Research Experiences to Enhance Learning (OCUR-REEL) program. Clark's experience with the program and his approach to assessing the impact of it have produced some important insights regarding how students view their chemistry learning experience as well as the impact of embedding research experiences within an introductory course. Based on the assessment evidence, developing and offering a research immersion experience earlier in a student's postsecondary chemistry experience rather than later may provide a fairly intuitive, "organic" strategy for enhancing student interest in chemistry, shatter myths about scientific inquiry and work, and encourage more students to continue their journey in chemistry education.

Roehrig, Kern, Wood, and Nyachwaya focus their attention on the use of an "equation drawing task" as a formative assessment of the depth of students' understanding of ideas and concepts in chemistry in Chapter 3. However, the population of interest is not college students. Their chapter shares what they have learned based on experiences with high school students—a critical segment of the STEMM pipeline whose current experiences and learning in chemistry will play a large role in shaping their future success and interest in these fields.

In Chapter 4, Kinder and Johnson share their experiences in developing course-based assessments for an introductory chemistry course at a regional campus that is part of a larger research university. The study focuses on supporting and improving student learning via evidence-based course redesign and instructor development. Four assessment tools were specifically developed to collect data and were designed to be used inside the university's online course management system. The assessment tools, even as they continue to be enhanced and tested, have helped shed light on whether or not the course achieves established learning goals for its students. Their experiences and recommendations also provide an important perspective for those whose primary course audience comprises non-science majors.

Turning to the community college sector as another important part of the STEMM pipeline, in Chapter 5, Carver, Brothers, and Higgins describe the use of multiple assessments utilized during the STEM-ENGINES Undergraduate Research Collaborative, an NSF-funded Undergraduate Research Collaborative (URC) that focuses on the experiences of students at two-year community colleges. Students participated in research experiences during the academic year, with half of them continuing to do research at a four-year college or university during the summer. The authors describe the project and the use of assessment instruments that focus on student learning of chemistry content and process skills, emotional intelligence, and potential pursuit of STEM careers. In addition, the project evaluation design creates opportunities to gauge the impact of these experiences have on those who start their postsecondary education at a community college.

Hearne and her colleagues offer, in Chapter 6, an important perspective on chemistry assessment from the point of view of teaching at an institution that serves under-represented students. As part of the National Center of Academic Transformation's Course Redesign Initiative, the authors employed a "replacement model" to redesign a course where there was broad variation in the base knowledge of incoming students, a 55% student persistence rate into the second part of the freshman chemistry sequence, and a lack of coordination among the faculty members teaching the course, leading to course drift and inconsistent learning outcomes. Their redesign and assessment strategy provides some important ideas for others in the process of attempting comprehensive instructional reform with assessment as a key strategy to assess impact.

Turning to the impact of teaching strategies, in Chapter 7, Loertscher details an NSF-funded project to explore the impact of active learning techniques in biochemistry courses taught at a private university. Over time, the use of assessment as a tool to enhance learning and improve instruction has moved from a less formal approach to one that is more systematic and rigorous. Starting with data collected from student examinations and student self-reports, the project now is moving to a pretest/posttest design to assess impact. This experience could serve as a model to inform efforts at other institutions to enhance student learning in biochemistry, particularly when attempting to make a transition from initial assessment activity to the practice of a scholarship of teaching and learning

In Chapter 8, extending the range of study on active learning, O'Sullivan and Copper provide a case study from their experiences teaching general chemistry at one of the nation's military academies. They examined the performance of more than 5,000 students over five years, comparing group work and active learning exercises to lecture-based approaches to instruction. Using a number of metrics, their findings suggest a positive effect on students who were part of an active learning environment in general chemistry. As the positive results became known to others in the department, more faculty members became interested in joining the project, offering important insights into how assessment strategies and innovations might be replicated and adopted throughout a department over time.

Pienta provides two case studies in Chapter 9 that represent different motivations for the desire to document or measure the knowledge, skills, and attitudes of chemistry students via changes at the course level and at the curriculum level in anticipation of reaccreditation and based on theory-driven plans and first principles, representing a phenomenological approach. The first example documents the process by which a traditional, large enrollment introductory-chemistry sequence underwent changes to address student dissatisfaction, unacceptable success levels, and demands from other programs that used these courses to fulfill their degree requirements. Demonstrating success required both qualitative and quantitative measures, the latter apropos to the discerning scrutiny of a faculty group made up of scientists. The outcome of the redesign was measureable, sustained, and transformative—student satisfaction and success increased as did the approval of constituencies who required the courses. The assessment plan for chemistry's undergraduate curriculum was motivated by an institutional reaccreditation, potential changes required for degree accreditation by chemistry's professional organization, and the turnover of a substantial number of faculty in the department. Each of those factors provided different timelines, motivation, and expectations. Originally skeptical of the need for a curricular assessment, the faculty eventually accepted its desirability. The success in the course redesign aided in the buy-in of the latter venture, and the faculty ultimately produced an exemplary model.

Kahle, Scantlebury, Woodruff, and Li discuss the evaluation of two large-scale projects to reform first- and second-year chemistry courses in Chapter 10. The projects, funded by the National Science Foundation under its Undergraduate Research Center program, focused on providing students with authentic research experiences in undergraduate chemistry courses with the goal of increasing both the number and the diversity of undergraduates electing to continue to study chemistry. Both projects chose to change the nature of their introductory courses through modules that included cutting-edge research and real-life applications of chemistry as well as multiple partners across a variety of institutions (research universities, two- and four-year colleges and universities, public and private institutions). The partnerships provided unique challenges (availability of equipment on some campuses, institutional support, timely reporting) as well as opportunities (replication of modules, cross-institutional research). Their findings suggest future directions for improving undergraduate chemistry courses as well as for evaluating large-scale, multisite projects, including implications for female students, students who planned on professional careers, and students of various ethnic/racial groups. The advantages of multisite projects are considered, and recommendations for more longitudinal designs are offered in chemistry outcomes assessment.

In the concluding chapter, Coppola shares his perspectives on his chemistry department's decision to eliminate the traditional two-semester general chemistry sequence and to allow students with some background in chemistry to take an organic chemistry course. From the start, the development of this course was based on sound pedagogical principles and contemporary instructional strategies. Over 20 years and roughly 50,000 students, the department has continued to evolve the course in both content and method and has carried out substantive research on student learning that has informed practice. Coppola traces the development of the course and describes in detail three cases of alignment of explicitly identified learning goals, pedagogical approaches to achieving those goals, and the methods used to assess our outcomes, including higher level learning goals. Beyond quasi-experimental designs that attempt to identify group differences based on instructional interventions and student performance, Coppola brings us back to the discipline itself and the critical role faculty members must play in developing criterion-based assessments of student learning and progress that go beyond comparative, norm-referenced criteria.

Putting chemistry education into a larger context, the United States finds itself at a historic crossroads in STEMM disciplines and professions. Future economic growth, the capacity for innovation and new discovery, and global competitiveness will depend in no small part on the development of more highly skilled researchers and professionals in STEMM fields. In addition, the level of scientific literacy and understanding across the larger population must be enhanced to ensure a well-educated, informed society and citizenry. According to the Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology (2007) in its report *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*:

> ... current trends in each of those criteria indicate that the United States may not fare as well in the future without government intervention. This nation must prepare with great urgency to preserve its strategic and economic security. Because other nations have, and probably will continue to have, the competitive advantage of a low wage structure, the United States must compete by optimizing its knowledge-based resources, particularly in science and technology, and by sustaining the most fertile environment for new and revitalized industries and the well-paying jobs they bring. We have already seen that capital, factories, and laboratories readily move wherever they are thought to have the greatest promise of return to investors. (p. 4)

In conclusion, we hope that the experiences and insights shared in this volume will prove to be a useful and valuable support as you seek to increase both the number of students who achieve academic success in chemistry as well as the level of knowledge and skills for the next generation of scholars, practitioners, and citizens. In this way, we hope to accelerate our collective progress in addressing the concerns and challenges identified by the scientific education community. The extent to which the challenges highlighted by national experts are addressed will depend in no small part on our ability to enhance student success in chemistry, both for those who will be directly employed in such fields as well as others whose understanding of chemistry will be essential to their individual quality of life and to the general well-being of our society.

REFERENCES

Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academy of Sciences.

CHAPTER 2

AN AMBITIOUS STATEWIDE TRANSFORMATION OF INTRODUCTORY CHEMICAL COURSES: ASSESSING THE OHIO CONSORTIUM FOR UNDERGRADUATE RESEARCH-RESEARCH EXPERIENCES TO ENHANCE LEARNING (OCUR-REEL) PROJECT[1]

Ted Clark
Ohio State University

Introduction

Many students have preconceptions and opinions of "scientific research" before entering college. A survey of students in an introductory science course at a large research university (general chemistry at the Ohio State University) found that student views of research before participating in a research project were occasionally positive ("the whole reason I came to OSU was to get involved in research") but frequently negative. This negativity could stem from earlier experiences ("I had a sour experience with research in high school, and I had a really bad attitude about it. I had crossed it off") or from a lack of experiences and an understandable uncertainty as to what research entails. This uncertainty manifests itself in considerable unease ("I had no idea...as to how one even went about research and was generally intimidated by it") and an unwillingness to participate in research.

Chemistry is intrinsically a research-based undertaking, and chemical education has long sought to provide students with beneficial laboratory experiences (Singer, Hilton, & Schweingruber, 2006). These laboratory experiences are currently undergoing exciting transformations. Educational researchers have frequently discussed the role of laboratory experiences (Hofstein & Lunetta, 2004), and some faculty and staff are beginning to alter their courses and chemistry curricula in innovative ways. One dramatic example of change is including authentic research experiences in introductory courses, with the Ohio Consortium for Undergraduate Research-Research Experiences to Enhance Learning (OCUR-REEL or REEL) program being an exemplar of such change. A distinguishing characteristic of REEL is its aim to provide research experiences for a large number of students enrolled in chemistry courses early in their academic careers. To the extent that features of undergraduate research experiences are successfully assimilated into these courses, the benefits associated with such experiences should occur (Osborn & Karukstis, 2009; Trosset, Lopatto, & Elgin, 2008). However, as the above quotes suggest, these students will have diverse views of, and interests in, research, which may shape their experiences.

[1] This chapter is based on the author's contribution "Does Chem-Research Make a Difference?" appearing in the National Association of Research in Science Teaching (NARST) 2010 conference proceedings.

This chapter describes the REEL program, discusses its current and future evaluation, and provides insights into how students view their participation in authentic in-class research experiences.

Motivation for Change

Undergraduate Research Collaboratives (URCs), formed in response to a call from the Chemistry Division of the National Science Foundation (NSF), are envisioned to be innovative models and partnerships that (a) expand the reach of undergraduate research to include first- and second-year college students; and (b) enhance the research capacity, infrastructure, and culture of participating institutions, thereby strengthening the nation's research enterprise (Undergraduate Research Collaboratives, 2005). The URC's aim is to improve undergraduate science education by providing active and engaging modes of learning, such as research opportunities, to a large number of students at diverse higher education institutions (National Science Foundation, 2003). There are currently five URCs, with more than 40 higher educational partners participating in multi-institutional collaborative efforts. Each of these URCs approaches the call for transformative change in a different way. This chapter describes the efforts of the REEL program to influence the pedagogy and culture of chemistry departments across Ohio.[2]

The REEL program is now entering its fifth year of NSF support. The goals of the program remain ambitious and have, in fact, increased as the program continues to evolve. The central aim of REEL is to introduce research experiences into first- and second-year chemistry courses to both increase Science, Technology, Engineering, and Mathematics (STEM) retention, and to generate new knowledge in the chemical sciences through multisite faculty-student collaborative research projects. The goal of increasing retention in STEM fields by having students participate in research is informed by the work of investigators who examine why students leave the STEM disciplines after taking introductory science courses like general chemistry (Daempfle, 2003–2004; Packard, 2004–2005; Seymour, 2001; Seymour, Hunter, Laursen, & Deantoni, 2004). The reasons why students leave STEM disciplines are varied, but interestingly they often *do not* include poor academic performance in first-year courses or negative views toward large classes. The quality of science instruction is, however, a strong determinant, especially if students believe that faculty view teaching as a burden and value research more strongly than teaching. This last point is especially relevant for the REEL program, for although negative student attitudes toward faculty research may result in students leaving STEM disciplines, when students are allowed to participate in that research, their views (and STEM retention) may

---

[2] The 15 higher education partner institutions in OCUR-REEL include the University of Akron, Bowling Green State University, Capital University, Central State University, University of Cincinnati, Cleveland State University, Columbus State Community College, University of Dayton, Kent State University, Miami (OH) University, Ohio University, Ohio State University, University of Toledo, Wright State University, and Youngstown State University. REEL partnerships also exist with high schools in Oxford, PA and Delaware, OH.

change positively. Clearly, to effectively increase STEM field retention by including research experiences in introductory courses, a large number of students must be included in the REEL program. For this reason, REEL classes are not intended for a subset of student (e.g., Honors students). Instead, the overarching program goal is to transform the experience for all students in these courses.

Program Overview

The OCUR-REEL institutions include community colleges, liberal arts colleges, and research universities across Ohio. In the first four years of the program, approximately 8,000 students have taken REEL chemistry courses, and more than 12,000 students will participate in the program's initial five years. The administration for REEL is situated at the Ohio State University (OSU), the largest participating institution in terms of number of students and research infrastructure.

The success of the REEL program is closely tied to the development, implementation, transfer, and evaluation of research modules. A research module is inspired by an actual research question and is preferably an extension of the research interests of REEL faculty or staff teaching the class. We have found a close involvement by module designers in the module's implementation to be a critical factor for increasing faculty and staff ownership of the project. Although this need for a close connection between module design and implementation may hinder a module's dissemination, the benefits from such a connection have been noteworthy, especially in terms of REEL-derived publications. A trade-off may well exist between ease of module dissemination and the likelihood a module leads to tangible gains in chemical knowledge, like research publications. REEL's approach, therefore, somewhat diverges from those URCs that favor the transfer of easily adopted research modules to many institutions (Weaver, Varma-Nelson, Wink, Morris, & Lytle, 2006; Weaver, Wink, Varma-Nelson, & Lytle, 2009).

The task of identifying and adapting research questions for classroom use is not a trivial matter, especially if the student participants are found in introductory courses. Finding a proper balance between logistical issues (e.g., time allocation in the laboratory, student safety, required chemicals), instrumentation needs, educational merit, and student engagement is a crucial and ongoing process. It has also become clear that the specific implementation of a research module is largely institution-dependent, as the balancing of these many needs depends on the resources and culture of a given chemistry department.

Research modules are designed to be authentic in the sense that core attributes of scientific reasoning are included, and students contribute to a project in which the scientific research is potentially publishable and not predefined. This means that students have the opportunity to generate research questions, select variables to investigate, coordinate results with other studies, and communicate their results, *inter alia* (Chinn & Malhotra, 2001). By having student investigations fit into the context of a broader research question, it is possible to encourage student ownership of a narrow research question while also contributing to a meaningful research project (Stefanou, Perencevich, DiCintio, & Turner, 2004).

The REEL module entitled "Nontoxic Inorganic Pigment Design" illustrates common attributes of a research module. Since many traditional pigments (e.g., red, orange, and yellow ones) contain toxic heavy metals such as cadmium, mercury, and lead, a need exists to identify alternate pigments that are environmentally friendly. General chemistry students are challenged to synthesize and characterize such nontoxic pigments. For a given class of solid-state compounds, students have an extremely large number of compositional possibilities they may investigate. By exploring a suggested composition parameter space, students synthesize compounds and examine relationships between structure, composition, synthesis method, and color. By combining results from a large number of students, an impressive number of samples may be prepared and analyzed. Characterization instrumentation for this module includes X-ray diffraction, transmission UV-Vis spectroscopy, and reflectance UV-Vis spectroscopy instrumentation. Synthesis methods include precipitation reactions and traditional high temperature solid-state methods. These activities replace about three or four weeks of traditional laboratory experiments. This module has resulted in one research publication (Dolgos, Paraskos, Stoltzfus, Yarnell, & Woodward, 2009), and comparable modules are in use at several partner institutions including Capital University, Youngstown State University, the University of Akron, as well as OSU.

Research modules examining aspects of environmental chemistry are also popular at multiple REEL institutions, including Bowling Green State University, Columbus State Community College, Capital University, and OSU. The chemical analysis of natural water samples from a variety of ecosystems, including wetlands, rivers or man-made ponds is possible on a large scale with ion-selective electrodes and UV-Vis spectrophotometry for analytes that affect water quality including nitrate, phosphate, chloride, and fluoride ions. The analysis of heavy metals in soil samples by X-ray fluorescence spectroscopy and flame atomic absorption spectrophotometry has also been implemented on a large scale. Environmental chemistry modules such as these are well suited for quantitative analysis and general chemistry courses and afford a merging of student research questions and an overarching investigation.

Students in REEL courses typically work closely with classmates in small groups. These small groups receive guidance from "experts" that may include the instructor, a teaching assistant, and perhaps a peer-mentor, all of whom are familiar with the module's overarching research theme and are capable of providing guidance for module-specific needs in terms of instrumentation, software applications, data analysis, etc. Such close guidance is a crucial component often lacking when instructing students in authentic scientific practices (Hodson, 1996). Peer-mentors are especially important at OSU, given the large number of students who participate in REEL each year (approximately 1,000 and increasing). Peer-mentors are students who complete a REEL research experience and then enroll for module-specific training. Other URCs have also successfully incorporated peer-led learning in their research programs (Weaver et al., 2006; Weaver et al., 2009).

Finally, opportunities exist for students to share their results with a larger audience. Most REEL courses conclude with research poster sessions or research presentations at which student groups communicate their findings. Other, more innovative ways to communicate scientific information, like web sites or digital video narratives (Clark & Clark, 2008) have also been explored. Statewide conferences and symposia hosted by REEL or the American Chemical Society are also important venues with students meeting researchers from other institutions.

REEL vis-à-vis Other Research Experiences

As Richard Blanton states in a review of undergraduate research experiences, "Undergraduate research has traditionally been an informal cottage industry of individual arrangements between students and professors, plus a network of federally funded summer research opportunities" (Blanton, 2008, p. 233). He goes on to observe that, while this cottage industry still thrives, undergraduate research has grown into a movement that is becoming institutionalized. One theme of this movement is its expansion to include a large number of students. As discussed by Boyd and Wesemann (2009), the variety of programs providing research experiences to undergraduates is remarkable. The approach of a faculty member mentoring an upperclassman and perhaps guiding a senior thesis still occurs. The broadening of participation in research, however, is becoming an increasingly common goal in higher education. Approaches for broadening participation take many forms, such as providing a small number of nonscience majors the opportunity to collaborate with scientists (Beane & Urquhart, 2009), recruiting students early in their academic careers to conduct research in fields and disciplines of interest (e.g., the Undergraduate Research Opportunities Program [Gregerman, 2009; Locks & Gregerman, 2008]), or the URC model that situates research experiences in discipline-specific introductory courses. Before considering the evaluation of the URC model as implemented in the REEL program, it is valuable to describe salient features of REEL's research philosophy in terms of generating research questions and the recruiting and mentoring of students.

Research questions. Given REEL's ambition to generate new knowledge in the chemical sciences, it should be noted that REEL's research philosophy inverts the traditional science research paradigm.   Research in chemistry departments is usually characterized by having a small number of highly trained individuals dedicate a large amount of time to investigate a particular research question. This is true for research groups employing graduate student researchers and postdoctoral researchers, and this is also the model adopted by NSF's Research Experiences for Undergraduates (REU) program. The REEL approach, in contrast, utilizes a large number (perhaps 300 at a time) of novice students (with general chemistry experience) who invest only a limited amount of time (approximately three to five weeks). In addition, as the introductory quotes to this chapter illustrate, many of these students will not be enthusiastic with starting a required research experience in chemistry. Given these constraints, it is clear that many research questions are ill suited to REEL's approach. There are, however, investigations that benefit from an approach with a massive scale, with combinatorial chemistry being a prime example.

Student recruitment. Students enrolling in general chemistry courses at REEL institutions in Ohio are usually STEM majors, but only a small minority are actually chemistry majors. Across REEL, more than one-half of general chemistry students intend to attend a professional school upon graduation (e.g., medical, veterinarian, pharmaceutical; Kahle, Li, & McFaddin, 2007). These students will be required to take several chemistry courses, but most *do not* enter general chemistry anticipating (or looking forward) to doing chemical research. REEL's goal and practice of providing research opportunities to all general chemistry students is markedly different from other programs that recruit students and pair them with faculty mentors from diverse fields and disciplines in which students have an expressed interest.

Mentoring students. The manner in which science students are mentored is frequently cited as an important factor influencing the quality of the student's research experience (Taraban & Blanton, 2008). Alternatives to faculty-student mentoring relationships, like REEL's peer-mentor program, may be one way to overcome the structural barriers that discourage mentoring on a grand scale at large universities (Packard, 2004–2005), and other URC sites have pursued similar strategies. However, it is important to realize that a peer-mentor relationship differs sharply from a long-term faculty-student mentoring relationship that often accompanies undergraduate research experiences. This limitation may affect REEL's ability to influence student views in those areas that could benefit from close collaboration with a faculty member, for example, career decisions (Lopatto, 2007).

Program Evaluation

Discussion of REEL's evaluation begins by noting those stakeholders most invested in the program, since their interests will inform any assessment effort. Prominent stakeholders for REEL include NSF, the faculty and staff involved in REEL, the administrators at REEL institutions in addition to science educators interested in issues like laboratory reform, student engagement, and curricular change.

Recalling that URCs were formed in response to a call from the Chemistry Division of NSF, one measure of success will be research publications derived from REEL research. The extent to which REEL expands the reach of undergraduate research to increase the number of first- and second-year college students is also valued, as is the effect these research experiences have on student retention in STEM disciplines. REEL faculty and staff share these concerns, but must also seek formative assessment to guide program development. Administrators at partner institutions, interested in both the sustainability of REEL and its extension to other fields of study, seek tangible student gains following participation in REEL. Finally, REEL's evaluation will have value to science educators once placed in the context of other pedagogical or curricular innovations. Any such comparisons will require the use of common assessment instruments, be they focused on content learning, changes in student attitudes, insights into the nature of science, or other areas.

As shown in Figure 1, the evaluation of REEL is being accomplished in several stages. Stage 1 was completed in the initial year of the program, and stage 2 has been an ongoing process in years 1 through 5. These assessments have been conducted in consultation with external evaluators from the Evaluation & Assessment Center for Mathematics and Science Education at Miami (OH) University. Stage 3 depends on longitudinal data and is currently in development.

Stage 1: Initial level of inquiry in non-REEL labs.	**Stage 2:** Student perceptions of teaching and learning in REEL and non-REEL courses.	**Stage 3:** Retention in STEM	**Stages 4 and 5:** Comparison with other research experiences. Instrument development and student case studies

Figure 1. Stages of evaluation for the REEL Program.

In stage 1, laboratory experiments from pre-REEL general and organic chemistry courses were initially rated using a modified inquiry rubric (Lederman, 2004). This entailed examination of syllabi, lab manuals, and instructor manuals for general chemistry and organic chemistry courses (Kahle & Marks, 2007). Following establishment of inter-rater reliability (IRR) for the rubric (IRR ³ 0.80 for undergraduate chemistry laboratory experiments), pre-REEL courses were assigned to a level. As described in Lederman (2004) and summarized in Kahle and Marks (2007), the defining characteristics of these levels include:

Level 0—Problem area, methods and "correct" interpretations are given or are immediately obvious from either statements or questions in the students' laboratory manual or textbook. Includes activities in which students simply observe or "experience" some unfamiliar phenomena or learn to master a particular laboratory technique.

Level 1—Laboratory manual proposes problems and describes ways and means by which the student can discover relationships he/she does not already know from lab manuals and texts.

Level 2—Problems are provided, but methods and solutions are left open.

Level 3—Problems, as well as solutions and methods, are left open.

Without exception, pre-REEL laboratory experiments were characterized as having defined problems for students, detailed procedures, and defined methods for solutions/analyses (Level 0 on the rubric). Thus, prior to implementation of the REEL program, experiments were not designed to facilitate inquiry and instead focused on students performing experiments to verify results.

An expository laboratory instruction style, like the one used in pre-REEL laboratory experiments, is the most popular and most heavily criticized style of laboratory instruction (Domin, 1999). Advantages of this style are largely logistical, as activities may be performed simultaneously by a large number of students with minimal involvement from an instructor. It is also possible to design expository experiments so that they are completed in a two- to three-hour time span with costs kept to a minimum. Although the limitations of this instructional style have long been recognized (Schwab, 1962) and calls have been made to move forward by increasing the levels of student inquiry (e.g., the influential Boyer Commission Report [Kenny, 1998]), prior to REEL, an expository instruction was deeply entrenched in chemistry departments across Ohio.

REEL's stage 2 assessment began in year 1 of the program. This effort has focused on evaluating the teaching and learning practices that accompany the implementation of REEL research modules. As discussed previously, research modules are the central means by which a research-intensive program is being introduced into the first- and second-year chemistry curriculum. Quantitative methods are primarily used to assess the impact of these research modules on student views of teaching and learning in REEL courses by administering a 35-item Likert-type questionnaire to all students at a course's conclusion. This survey is an adaptation of a valid and reliable instrument developed previously and consists of three subscales that ask students to describe teaching and learning practices in their chemistry course. The questionnaire also includes eight Likert-type items that ascertain student understanding of the nature of science (NOS). The NOS is a component of scientific literacy that may be affected by participation in authentic research (Lederman, 1992; Matthews, 1998; Schwartz, Lederman, & Crawford, 2004). The subscales for the questionnaire are "What Instructors Do," "What Students Do," and "My Views About Science." Responses for the subscales concerning teaching and learning behaviors are on a 5-point Likert-type scale with responses ranging from *Almost Never* (1) to *Very Often* (5), while responses to the "My Views about Science" subscale range from *Strongly Disagree* (1) to *Strongly Agree* (5). Additional items, specific to REEL's goals, were added to explore students' future career plans and experience in independent research projects. Reliability coefficients (Cronbach alpha values) are calculated for each subscale each year. Cronbach alpha values are usually ~0.80 to 0.90 for each subscale. Data are combined for a given course across all institutions, so OSU's large number of students is an important factor to note. Within each course, analyses then investigate student responses in REEL versus non-REEL course with variables such as gender or student-expressed career plans examined. Rasch psychometric techniques have also been utilized with this dataset.

How does a REEL course differ from a non-REEL course? Although a full discussion of questionnaire data is beyond the scope of this chapter, a summary of the results (Table 1) provides insights into how inclusion of in-class research experiences and accompanying changes in classroom pedagogy affect student views.

Analysis of questionnaire data provides an overview of student views of teaching and learning in REEL classes compared to non-REEL courses. As expected, the REEL courses are perceived as including many activities consistent with an increase in student inquiry (e.g., more likely to design activities to test their own ideas, consult with classmates as sources of learning) and away from a traditional expository format emphasizing learning by studying a textbook, taking notes in lecture, and memorizing scientific facts. An expanded discussion of this analysis may be found in Chapter 10 of this volume.

Table 1

Views of General Chemistry Students in REEL and Non-REEL Courses (statistically significant responses at the $p < 0.05$ level shown)

"In REEL General Chemistry, students..."	"In non-REEL General Chemistry, students..."
Argue or debate with one another about the interpretation of data.	Learn science by studying the course textbook.
Repeat experiments.	Take notes and listen to lectures.
Consider alternate explanations to accepted theories.	Memorize scientific facts.
Design activities to test their own ideas.	Learn scientific facts by using charts and diagrams.
Consult one another as sources for learning.	
Talk with one another to promote learning.	
Use educational technology in class.	
Develop scientific literacy skills.	

In addition to questionnaire data, many REEL instructors and module designers desire formative student feedback for specific research modules. This usually is generated by including open-ended questions for the students at the conclusion of a module. Student responses may be submitted anonymously and are not linked to student descriptors, like gender or performance in the class. The construction of open-ended questions, of course, will influence the nature of the responses. It has been observed, for example, that students' observations of "what they liked" about a laboratory experience are less useful than their estimates of "what they gained" (Seymour, Wiese, Hunter, & Daffinrud, 2000). Although our experiences with question construction are similar to those noted by Seymour et al., the focus here has not been on student-stated learning gains, but rather on identifying those aspects of a REEL course students viewed as important and worth retaining, and those aspects requiring change by asking questions like, "What aspects of the REEL project should be removed next year." Framing the question in this manner allows students to reflect on their laboratory experiences in a broad sense and provides module designers with valuable module-specific information. It has also been useful to ask students whether they would "advise future students to enroll in a REEL or a non-REEL course." This also elicits very insightful responses as students reflect on their REEL experiences and situate them in the context of other academic and nonacademic commitments. Also, given the goals of REEL, students are asked to comment on their views of research, to describe how these views may have changed, and whether they are interested in pursuing additional research opportunities.

The formative feedback provided for research modules may be discussed in terms of students' perceptions of laboratory learning environments. The subscales included in the Science Laboratory Environment Inventory (Fraser, Giddings, & McRobbie, 1995, that is, student cohesiveness, open-endedness, integration, rule clarity, and material environments, provide a useful framework for examining student responses. An examination of ~150 student responses for a particular research module ("Investigation of Heavy Metals in Urban Soil Samples") provides representative responses. For example, when asked, "What aspects of the REEL project should definitely be retained," these students strongly favored the open-endedness of the assignment. The excitement of actually forming research questions was quite clear, especially when compared with traditional laboratory instruction. As a student noted,

> I liked the idea that we were able to choose what we wanted to research. It was good to be able to test something you wanted to research instead of reading from a book and doing what it tells you.

Many students directly contrasted their REEL laboratory experiences with the non-REEL labs they had completed earlier in the year:

> I enjoyed how it was not just "copy this procedure out of the book, do it word for word, go home, write a lab report, and do it again." We got to make our own procedures and I think that definitely should be retained for next year.

The strong support for open-ended tasks is clear for these students, with 56% including this trait when describing characteristics of REEL that should be retained and only 14% criticizing the lack of defined procedures and direction. The relationship between open-endedness and student attitudes is not always favorable and may vary with student population (Fraser et al., 1995) but was viewed positively here.

Student views of group work were also quite positive, with 38% identifying this as an aspect that should be retained, and 10% viewing it negatively. Group work, as described here, is consistent with the laboratory dimension of student cohesiveness. The value of group work, for many students, was the opportunity to learn from their peers. This is illustrated in the following student responses:

> By working in groups I had the opportunity to actually discuss results and learn about how others view data differently,

> The aspects of the program that I enjoyed were working together with a group and forming a hypothesis. It was very interesting to work with a group who come from many different experiences to see what everyone can bring to the table. Working with groups is very important,

> I loved having a group because we could split the work and bounce ideas off each other. We are all intelligent students because we have made it to 123 [the final general chemistry course] so why not share the knowledge?

Especially welcome in these comments is an indication that students view their peers as legitimate sources of learning. This is not entirely unexpected since questionnaire data indicated that REEL students are more prone to "Argue or debate with one another about the interpretation of data" than non-REEL students. However, student remarks illuminating their views on this subject are insightful. In terms of student epistemological development, the recognition of classmates as sources of learning is noteworthy. Traditional expository laboratory instruction provides little opportunity for student epistemological growth. Indeed, the rule-bound nature of expository instruction may do more to retard such growth (Finster, 1991), and REEL courses have the potential to dramatically expand "what counts" as sources of knowledge.

Another dimension of group work that should not be neglected is the social one. Traditional chemistry labs are, for the most part, solitary experiences with students working independently. Many students welcomed the opportunity to break out of this laboratory format. In the words of one such student:

> I feel I learned so much about research, chemistry, and made so many new friends in the process. Usually chemistry lab is just a scary place where no one talks to each other for the whole quarter, but REEL lab is definitely not like that.

What about students' perceptions of research? Students take up this question when considering whether to recommend REEL. Overall, for this module, 65% of students recommended future students enroll in a REEL course, 27% offered a conditional recommendation, and only 8% recommended avoiding a REEL course.[3] The value of research was prominent in many recommendations:

> I would definitely recommend to take the course that includes REEL, especially if you are interested in research but don't have a clue of what it actually entails. I think it is a good experience for everyone to get a taste of research-based science.

Finally, most students (75%) indicated their views of scientific research changed by participating in REEL. Comments frequently demonstrated an understandable ignorance on the part the student prior to participation in REEL:

> I have greater appreciation as to how much time and how many people it takes to achieve a small piece of information.

> (I have) better insights. Before this I had no idea what to expect in research. It is a lot tougher than it seems or I thought it would be.

> I used to think that research was just testing a hypothesis and hoping for the best, but I now realize a lot more effort comes into it. You have to plan what you will be doing and then you must think about what will happen.

These comments speak directly to the authenticity of the research in this module. This experience is designed to stress genuine chemical research, not simple inquiry-based laboratory experiences that include epistemologically deficient portrayals of science (Chinn & Malhotra, 2001).

---

[3] REEL courses are not electives. At smaller REEL institutions, all general or organic chemistry courses in a given term may be REEL offerings. At larger institutions, both REEL and non-REEL sections may be offered in the same term but students are not usually given advance notice as to which sections include research modules.

In addition to describing the difficulty of scientific research, students also remarked how research could be more fun or interesting than they had imagined, and that research was now something they would consider pursuing. As a student stated, "I had always thought research was done by the smartest people in the world and I never thought I would fit into that category." On balance, having identified research as an activity requiring more effort and patience than first imagined, do students want to pursue additional research experiences?

Of the 86 students responding to this question, only 7 students indicated a decreased interest in pursuing research. Far more students indicated an increased or continued level of interest (n = 50), or an unchanged (usually high) level of interest (n = 29). Negative comments generally focused on the level of effort research required and its open-endedness:

> No, I do not want to do research. It is a very intense process and it's very open-ended without a lot of guidelines.

The positive student views regarding REEL and future research often described how research was a new experience that now seemed more accessible:

> Personally, research scared me because I knew nothing about it. Now I know it is not that scary and can be brought down to the level that I understand. I believe it has given me confidence I need to make the next step. I even had an interview today for a lab position for the next coming year!

> Before REEL I had no opinion of research except one word: intimidating. Now I feel completely different. I would be really excited to find a research project that interests me. Without REEL I would still be very hesitant to sign up for any other projects.

Overall, based on these student responses, the "Investigation of Heavy Metals in Urban Soil Samples" research module has undergone little change. The REEL laboratory environment for this module favors open-endedness and encourages student cohesiveness, both of which are welcomed by the students. In addition, students view this experience as a legitimate introduction to chemical research, with many describing an interest in pursuing additional research experiences.

Beyond describing student views of REEL courses, several pragmatic metrics are being examined to evaluate REEL's impact. One such metric is student involvement in research. The goal here is to identify students who have participated in REEL courses and track their subsequent undergraduate (and perhaps postgraduate) research experiences; such data are currently being compiled. It is also worth noting, however, that REEL is strongly influencing the culture of undergraduate research at REEL institutions beyond simply the REEL classes. For example, the notion that undergraduates are capable of contributing to research projects early in their academic careers has taken hold in many chemistry departments among both students and faculty. This has resulted in an increase in the number of students participating in chemical research, even for students and faculty who have not participated in REEL. For example, the number of undergraduates participating in research in the chemistry department at OSU is at an all-time high, having grown by 300% following REEL's inception. This transformation of the department's culture of research is exciting but makes a definitive analysis of REEL's contribution challenging.

Recalling that REEL modules are not designed solely to improve laboratory learning environments but also to contribute to the generation of new scientific knowledge, another useful metric is the number and quality of REEL-derived research publications. Faculty in the REEL program are increasingly using student results in research publications (Binkley, Ziepfel, & Himmeldirk, 2009; D'Cunha, Collins, Richards, Vincent, & Swavey, 2006; Demaray, Thuener, Dawson, & Sucheck, 2008; Dolgos et al., 2009; Lunsford, Choi, Stinson, Yeary, & Dionysiou, 2007; Matthews et al., 2008; Presswala et al. 2008), and other articles have been published describing REEL laboratory experiments and experiences (Clark, 2009; Lunsford, Speelman, et al., 2008; Lunsford, Stinson, & Widera, 2006; Lunsford, Yeary, & Nguyen, 2008; Lunsford, Yeary, Stinson, Choi, & Dionysiou, 2007; Novak et al., 2007). The likelihood of REEL-derived findings being published is quite dependent on the nature of the research question (as discussed above) and the degree to which the faculty member takes responsibility for the research project.

The Future of REEL's Evaluation

The REEL program is best described as a pseudo-experiment with numerous innovations taking place simultaneously throughout the chemistry curriculum at diverse higher education institutions across Ohio. Although the evaluation of the program to date suggests exciting transformations, it has been understandably limited and directed toward meeting the needs of the immediate stakeholders (i.e., the funding agency and the faculty at partner institutions). The evolution of REEL and its evaluation (stages 4 & 5 in Figure 1) requires a shifting of focus away from indirect measures (such as self-reports) describing the benefits of incorporating research into undergraduate classes. It will be important to assess the actual benefits of including research, not just the potential for such gains (Prince, Felder, & Brent, 2007). Beyond analyzing student retention data for REEL institutions (an investigation currently underway), the use of common instruments will be required to consider the pros and cons of the REEL program vis-à-vis other URCs, other programs promoting undergraduate research, and undergraduate experiences in general. This is the direction of REEL's future assessment, and a few preliminary thoughts regarding REEL's continued evaluation conclude this chapter.

The Survey of Undergraduate Research Experiences (SURE) and the Classroom Undergraduate Research Experience (CURE) survey are both instruments that David Lopatto has used to assess the benefits of undergraduate research experiences. A distinct advantage of these instruments is their implementation on a large scale (e.g., more than 40 institutions provided data concerning undergraduate research experiences in SURE's first year of use [Lopatto, 2008]). This large-scale implementation is facilitated by the online availability of the surveys.

Many of REEL's aims regarding student learning—such as having students improve their understanding of how scientists work on real problems, gain in ability to analyze data, or express a readiness for more demanding research–are considered in these surveys, with CURE being a very clear match. It would have been premature in REEL's initial years to compare REEL research experiences with those of other institutions. Now, however, is the start of an exciting period in which REEL's research model may be compared with those programs, and the benefits of REEL undergraduate research experiences may be examined with instruments common to many other sites.

Another instrument with widespread use and online availability that is relevant to REEL's aims is the Student Assessment of their Learning Gains (SALG) survey (Seymour et al., 2000). This instrument was originally developed to match the learning objectives and teaching methods for innovations introduced in chemistry classes at a large number of two- and four-year institutions. The SALG was designed to summarize the learning gains that students perceive they made, both as a consequence of classroom pedagogy and as a consequence of the teacher's pedagogical approach. The SALG website currently has users representing approximately 100 institutions, and an increasing robust dataset is being compiled to which REEL could be compared.

In addition to the SALG, several other noncontent assessment instruments have been introduced in chemistry. Examples include the Attitude toward the Subject of Chemistry Inventory (ASCI; Bauer, 2008) for measuring student attitudes regarding "chemistry" as a body of knowledge or practices; the Chemistry Self-Concept Inventory (CSCI; Bauer, 2005), the Chemistry Attitudes and Experiences Questionnaire (CAEQ; Dalgety & Coll, 2006), which examines self-efficacy, attitudes, and learning experiences; and CHEMX (Grove & Bretz, 2007), which probes cognitive expectations. This wealth of options highlights the importance of careful deliberation before choosing a particular instrument for program-wide use.

As noted by Barry Fraser, "Few fields of educational research have such a rich diversity of valid, economical and widely-applicable assessment instruments as does the field of learning environments" (Fraser, 1998, p. 7). Fraser's Science Laboratory Environment Inventory (SLEI) is an example of such an instrument that could be employed throughout REEL. Much of the work in classroom environment research has involved investigation of associations between students' cognitive and affective learning outcomes and their perceptions of psychosocial characteristics of their classrooms. Evaluation of education innovations has been another area of interest. Certainly REEL's dramatic departure from traditional expository instruction is one such innovation worth further investigation.

In addition to these established instruments, REEL would also benefit from instruments that probe content knowledge included in a research module. Concept Inventories (CIs) are an example of instruments that investigate a narrow subject area and, importantly, consider students' conceptual difficulties and misconceptions. Although CIs are growing in popularity (Libarkin, 2008; Richardson, 2004), instruments to probe topics prominent in REEL's research modules are currently lacking. For example, research modules investigating solid-state chemistry are popular at many REEL institutions. However, student learning of topics fundamental to solid-state chemistry, like X-ray diffraction and the periodic structure of solids, are not evaluated in a comparable way within REEL. The opportunity clearly exists for REEL faculty to collaborate with expert psychometricians and develop and implement CIs for topics included in research modules. This effort would provide insights into student gains in content understanding and should assist with the dissemination of REEL modules to other institutions.

The use of CIs to examine gains in learning content included in a research module should be accompanied by investigations examining student deficiencies in understanding for content that was removed to make room for REEL research. Research modules in general chemistry typically replace three to five weeks of traditional laboratory experiments in the final semester or quarter of the year-long course. It is noteworthy that several REEL institutions have chosen to reduce the time given to a common general chemistry assignment (i.e., a multiweek inorganic qualitative analysis experiment) in order to include a research module. An instrument measuring high school students' understanding of this topic has been developed (Tan, Goh, Chia, & Treagust, 2002) and could be modified to gather useful information as to what the REEL students are "missing."

Finally, it is important remember that many of the studies inspiring the creation of URCs were ethnographic ones, especially those completed by Elaine Seymour and her colleagues describing the benefits of undergraduate research. The REEL program and other URCs are ripe for similar ethnographic investigations. As this chapter suggests, many students are profoundly affected by undergraduate research. It is crucial to hear from these students, with ethnographic studies being a valuable way to enable their voices to reach a wide audience.

REFERENCES

Bauer, C. F. (2005). Beyond "student attitudes": Chemistry self-concept inventory for assessment of the affective component of student learning. *Journal of Chemical Education, 82*(12), 1864–1870.

Bauer, C. F. (2008). Attitudes towards chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of Chemical Education, 85*(10), 1440–1445.

Beane, R. J., & Urquhart, J. (2009). Providing research experiences to non-science majors in an introductory science course. In M. K. Boyd & J. L. Wesemann (Eds.), *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact* (pp. 223–226). Washington, DC: Council on Undergraduate Research.

Binkley, R. C., Ziepfel, J. C., & Himmeldirk, K. B. (2009). Anomeric selectivity in the synthesis of galloyl esters of D-glucose. *Carbohydrate Research, 344*(2), 237–239.

Blanton, R. L. (2008). A brief history of undergraduate research, with consideration of its alternative futures. In R. Taraban & R. L. Blanton (Eds.), *Creating effective undergraduate research programs in science. The transformation from student to scientist* (pp. 233–246). New York: Teachers College Press.

Boyd, M. K., & Wesemann, J. L. (Eds.). (2009). *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact.* Washington, DC: Council on Undergraduate Research.

Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86,* 175–218.

Clark, T. M. (2009). A statewide initiative for engaging undergraduates in research: Aligning efforts while fostering innovation. In M. K. Boyd & J. L. Wesemann (Eds.), *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact* (pp. 121–125). Washington, DC: Council on Undergraduate Research.

Clark, T. M., & Clark, C. T. (2008, July). *DIVINE chemistry and digital literacy: Student experiences with DIgital Video Inquiry NarrativE projects.* Biennial Conference on Chemical Education, Indiana University, Bloomington, IN.

Daempfle, P. A. (2003–2004). An analysis of the high attrition rates among first year college science, math, and engineering majors. *Journal of College Student Retention: Research, Theory & Practice*, *5*(1), 31–52.

Dalgety, J., & Coll, R. K. (2006). Exploring first-year science students' chemistry self-efficacy. *International Journal of Science and Mathematics Education, 4*(1), 97–116.

D'Cunha, D., Collins, D., Richards, G., Vincent, G. S., & Swavey, S. (2006). Dinuclear lanthanide(III) complexes containing b-diketonate terminal ligands bridged by a 2, 2'-bipyrimidine. *Inorganic Chemistry Communications*, *9,* 979–981.

Demaray, J. A., Thuener, J. E., Dawson, M. N., & Sucheck, S. J. (2008). Synthesis of triazole-oxazolidinones via a one-pot reaction and evaluation of their antimicrobial activity. *Bioorganic & Medicinal Chemistry Letters, 18,* 4868–4871.

Dolgos, M. R., Paraskos, A. M., Stoltzfus, M. W., Yarnell, S. C., & Woodward, P. (2009). The electronic structure of vanadate salts: Cation substitution as a tool for band gap manipulation. *Journal of Solid State Chemistry*, 182, 1964–1971. doi: 10.1016/j.jssc.2009.04.032

Domin, D. S. (1999). A review of laboratory instruction styles. *Journal of Chemical Education*, *76*, *543–547*.

Finster, D. C. (1991). Application of the Perry model to general chemistry. *Journal of Chemical Education*, 68(9), *752–756*.

Fraser, B. J. (1998). Classroom environment instruments: Development, validity and applications. *Learning Environments Research*, 1, *7*–33.

Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching, 32,* 399–422.

Gregerman, S. R. (2009). Filling the gap: The role of undergraduate research in student retention and academic success. In M. K. Boyd & J. L. Wesemann (Eds.), *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact* (pp. 245–256). Washington, DC: Council on Undergraduate Research.

Grove, N. P., & Bretz, S. L. (2007). CHEMX: An instrument to assess students' cognitive expectations for learning chemistry. *Journal of Chemical Education, 84*(9), 1524–1529.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies, 28,* 115–135.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28–54.

Kahle, J. B., Li, Y., & McFaddin, J. (2007). *Research experiences to enhance learning: Year 3 Annual Report*. Oxford, OH: Miami University, Evaluation & Assessment Center for Mathematics and Science Education.

Kahle, J. B., & Marks, P. A. (2007). *Research experiences to enhance learning: 2006–2007 Annual Report.* Oxford, OH: Miami University, Evaluation & Assessment Center for Mathematics and Science Education.

Kenny, S. S. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Stony Brook, NY: State University of New York.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching, 29,* 331–359.

Lederman, N. G. (2004). Laboratory experiences and their role in science education. In *America's lab report*. Washington, DC: National Academies Press.

Libarkin, J. (2008). Concept inventories in higher education. *National Research Council promising practices in undergraduate STEM education workshop* 2. Retrieved from http://www7.nationalacademies.org/bose/Libarkin_CommissionedPaper.pdf

Locks, A. M., & Gregerman, S. R. (2008). Undergraduate research as an institutional retention strategy: The University of Michigan model. In R. Taraban & R. L. Blanton (Eds.), *Creating effective undergraduate research programs in science. The transformation from student to scientist* (pp. 11–32). New York: Teachers College Press.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE Life Science Education, 6*(4), 297–306.

Lopatto, D. (2008). Exploring the benefits of undergraduate research experiences: The SURE survey. In R. Taraban & R. L. Blanton (Eds.), *Creating effective undergraduate research programs in science. The transformation from student to scientist* (pp. 112–134). New York: Teachers College Press.

Lunsford, S. K., Choi, H., Stinson, J., Yeary, A., & Dionysiou, D. (2007). Voltammetric determination of catechol using a sonogel carbon electrode modified with nanostructured titanium dioxide. *Talanta, 73,* 172–177.

Lunsford, S. K., Speelman, N., Stinson, J., Yeary, A., Widera, J., Choi, H., & Dionysiou, D., (2008). Electroanalytical and spectroscopic studies of poly (2,2′-bithiophene) modified platinum electrode to detect catechol in the presence of ascorbic acid. *Journal of Chemical Education, 85,* 128–130.

Lunsford, S. K., Stinson, J., & Widera, J. (2006). Electrochemistry and detection of organic and biological molecules such as catechol and ascorbic acid at conducting poly (2,2-bithiophene) electrode. *Analytical Science Digital Library*. Retrieved from http://www.asdlib.org

Lunsford, S. K., Yeary, A., & Nguyen, Q. P. (2008) Electrochemical detection of phenol utilizing a synthesized sonogel-carbon zirconium oxide. *Analytical Science Digital Library.* Retrieved from http://www.asdlib.org

Lunsford, S. K., Yeary, A., Stinson, J., Choi, H., & Dionysiou, D. (2007). Synthesis of a sonogel-carbon modified sensor electrode with titanium oxide (TiO_2) to detect catechol in the presence of common interferents by voltammetric studies. *Analytical Science Digital Library*. Retrieved from http://www.asdlib.org

Matthews, M. E., Atkinson, I., Presswala, L., Najjar, O., Gerhardstein, N., Wei, R., et al. (2008). Dielectric classification of D-and L-amino acids by thermal and analytical methods. *Journal of Thermal Analysis and Calorimetry, 93*(1), 281–287.

Matthews, M. R. (1998). In defense of modest goals when teaching about the nature of science. *Journal of Research in Science Teaching, 35,* 161–174.

National Science Foundation. (2003, March). *Exploring the concept of undergraduate research centers: Workshop final report.* Washington, DC: Author. Retrieved from http://urc.arizona.edu/

Novak, M., Wang, Y. T., Ambrogio, M. W., Chan, C. A., Davis, H. E., Goodwin, K. S., et al. (2007). A research project in the organic instructional laboratory involving the Suzuki-Miyaura cross coupling reaction. *The Chemical Educator, 12,* 414–418.

Osborn, J. M., & Karukstis, K. K. (2009). The benefits of undergraduate research, scholarship, and creative activity. In M. K. Boyd & J. L. Wesemann (Eds.), *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact* (pp. 41–53). Washington, DC: Council on Undergraduate Research.

Packard, B. W. (2004–2005). Mentoring and retention in college science: Reflections on the sophomore year. *Journal of College Student Retention: Research, Theory & Practice, 6*(3), 289–300.

Presswala, L., Matthews, M. E., Atkinson, I., Najjar, R., Wei, R., & Riga, A. (2008). Discovery of bound and unbound waters in crystalline amino acids revealed by thermal analysis. *Journal of Thermal Analysis and Calorimetry, 93*(1), 295–300.

Prince, M. J., Felder, R. M., & Brent, R. (2007). Does faculty research improve undergraduate teaching? An analysis of existing and potential synergies. *Journal of Engineering Education, 96*(4), 283–294.

Richardson, J. (2004). Concept inventories: Tools for uncovering STEM students' misconceptions. *Invention and impact: Building excellence in undergraduate science, technology, engineering, and mathematics (STEM) education.* Washington, DC: American Association for the Advancement of Science. Retrieved from http://www.aaas.org/publications/books_reports/CCLI/PDFs/02_AER_Richardson.pdf

Schwab, J. J. (1962). The teaching of science as enquiry. In P. F. Brandwein (Ed.), *The teaching of science* (pp. 1–103). Cambridge, MA: Harvard University Press.

Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Journal of Science Teacher Education, 88,* 610–645.

Seymour, E. (2001). Tracking the processes of change in U.S. undergraduate education in science, mathematics, engineering, and technology. *Science Education, 86, 79*–105.

Seymour, E., Hunter, A. B., Laursen, S. L., & Deantoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88,* 494–534.

Seymour, E., Wiese, D., Hunter, A., & Daffinrud, S. M. (2000, March). *Creating a better mousetrap: On-line student assessment of their learning gains.* Paper presented at the American Chemical Society 218th National Meeting, San Francisco.

Singer, S. R., Hilton, M. L., & Schweingruber, H. A. (Eds.). (2006). *America's lab report. Investigations in high school science.* Washington, DC: The National Academies Press.

Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychology, 39,* 97–110.

Tan, K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and applications of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching, 39*(4), 283–301.

Taraban, R., & Blanton, R. L. (Eds.). (2008). *Creating effective undergraduate research programs in science. The transformation from student to scientist.* New York: Teachers College Press.

Trosset, C., Lopatto, D., & Elgin, S. (2008). Implementation and assessment of course-embedded undergraduate research experiences: Some explorations. In R. Taraban & R. L. Blanton (Eds.), *Creating effective undergraduate research programs in science. The transformation from student to scientist* (pp. 33–49). New York: Teachers College Press.

Undergraduate Research Collaboratives. (2005). Program solicitation NSF 06-521. Retrieved from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=6675

Weaver, G. C., Varma-Nelson, P., Wink, D., Morris, R., & Lytle, F. (2006). Experiencing research in a first or second year science course: Preliminary results and reflections on the goals and progress of the Center for Authentic Science Practice in Education (CASPiE). *The Chemical Educator, 11,* 125–129.

Weaver, G. C., Wink, D. J., Varma-Nelson, P., & Lytle, F. E. (2009). Integrating research into the curriculum: Using laboratory modules to engage faculty and students. In M. K. Boyd & J. L Wesemann (Eds.), *Broadening participation in undergraduate research: Fostering excellence and enhancing the impact* (pp. 109–120). Washington, DC: Council on Undergraduate Research.

CHAPTER 3

HIGH SCHOOL CHEMISTRY STUDENTS' REPRESENTATIONS OF CHEMICAL REACTIONS AT THE ATOMIC/MOLECULAR LEVEL

Gillian H. Roehrig (University of Minnesota), Anne L. Kern (University of Idaho), Nathan Wood (North Dakota State University), and James M. Nyachwaya (University of Minnesota)

Many studies exist in the science education literature that explore student learning and the impacts of a variety of factors on student learning. These studies include the impact of factors such as curricular reforms (e.g., Kahle, Meece, & Scantlebury, 2000; Laguarda, 1998; Lockwood, 1995; Parker & Gerber, 2000; Roehrig & Garrow, 2007; Scantlebury, Boone, Kahle, & Fraser, 2001; Schneider, Krajcik, Marx, & Soloway, 2002; Von Secker & Lissitz, 1999), teacher variables (e.g., Huffman, Thomas, & Lawrenz, 2003; Roehrig & Garrow, 2007), and student gender (e.g., DeMars, 2000; Hamilton, 1998; Mullis, Dossey, Owen, & Phillips, 1993; Penner, 2003; Perie, Moran, & Lutkus, 2005) on student learning. Unfortunately, the majority of such studies rely on student test scores from standardized national or state tests (Schneider, Krajcik, & Blumenfeld, 2005). Such instruments assess knowledge of definitions and students' ability to solve mathematical and other rote procedural types of questions rather than conceptual knowledge of the discipline. In some fields, researchers have developed conceptual inventories in an effort to assess students' understanding of concepts central to the discipline. The most prevalent of such inventories is the Force Concept Inventory (FCI; Hestenes, Wells, & Swackhamer, 1992) that explores student understanding of force, the central concept of Newtonian mechanics, rather than students' ability to memorize and apply mathematics to Newton's laws.

In the chemistry education literature, three types of chemical representations have been discussed and are considered critical to developing a conceptual understanding of chemistry topics: macroscopic, particulate, and symbolic (Gabel, 1998). In *macroscopic* representations of chemistry, we expect that students can describe matter and changes in its properties in terms of characteristics that can be observed directly, via the senses, such as changes in state, color, temperature, etc. In *particulate* representations of chemistry, we expect that students can represent matter with constituent atoms and molecules, such as molecular models or particulate diagrams. Finally, in *symbolic* representations of chemistry, we expect students to be able to represent the interactions of atoms and molecules, and physical properties and observable changes in matter in symbols, such as mathematical equations, molecular formulas and chemical equations.

Studies have shown that students' conceptual knowledge regarding chemistry concepts, such as gas laws, is inadequate in comparison to algorithmic problem-solving skills. In fact, many students can successfully solve mathematical problems in chemistry yet do not understand the underlying chemical concepts (Gabel & Bunce, 1994; Nakhleh, 1993; Nurrenbern & Pickering, 1987; Sawrey, 1990). Unfortunately, this problem is not unique to high school students with limited exposure to chemistry concepts. Bodner (1991) investigated particulate-level understandings of gases held by incoming chemistry graduate students. In spite of the extensive time spent in laboratory and lecture during the undergraduate chemistry experience, a significant percentage of these students were unable to correctly explain what was in a bubble from boiling water, or how barometers, hot-air balloons, and pressure cookers work. At both the high school and college levels, chemistry curriculum, instruction, and assessment has traditionally focused on chemical and mathematical symbols and equations (symbolic representations) with little or no explicit discussion of the underlying concepts (Johnstone, 1991). In particular, traditional approaches to assessing chemistry assume that the ability to solve mathematical problems also reflects the students' understanding of the underlying concepts. The danger is that traditional chemistry assessments may fail to detect significant problems with students' conceptual understanding.

One particularly troubling area, in which students have been shown to successfully answer symbolic chemistry assessment items, while clearly not understanding underlying concepts, is balancing chemical equations. Yarroch (1984) showed that, while the high school chemistry students in his study could correctly balance simple chemical equations, most could not provide particulate drawings that were consistent with the notation of the chemical equations, even at the end of a year-long chemistry course. For example, the notation $3H_2$ could be appropriately represented as three pairs of hydrogen atoms (e.g., HH HH HH). However, many students in Yarroch's study instead drew a string of six connected hydrogen atoms (e.g., HHHHHH), indicating they did not understand the difference in meaning of subscripts and coefficients in chemical equation notation. In other words, these students successfully solved chemical equations following an algorithmic procedure, but could not demonstrate understanding of the atoms and molecules that the symbols in the equation were meant to represent.

Students' inability to appropriately represent the atoms and molecules in a simple chemical reaction are cause for concern, because modern chemistry (indeed, much of modern science as a whole) is predicated upon the notion of a particulate nature of matter. If students do not develop this central and fundamental concept during their chemistry classes, they will be unable to develop an understanding of any of the scientific ideas built upon it: bonding, chemical reactions, chemical thermodynamics, etc.

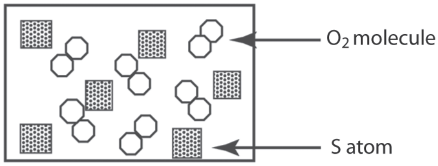
Assessing Students' Knowledge of the Particulate Nature of Matter

In the area of chemistry, there are very few instruments that provide the opportunity to assess students' conceptual understanding at the particulate level efficiently for a large number of students. The Symbolic, Application, Particulate (SAP) test was developed by Bunce and Gabel (2002) to assess the impact of a new teaching approach on students' knowledge of the particulate, macroscopic, and microscopic worlds of chemistry. The SAP consists of 30 multiple-choice questions covering 10 chemistry topics: states of matter, density, mixture/substance, conservation of mass, reaction type, moles, chemical reaction, solution, neutralization, and pH. A question using symbolic, macroscopic, or particulate representations is included for each of the 10 topic areas. Similarly, the Chemistry Concept Inventory (CCI; Mulford & Robinson, 2002) was developed to assess the conceptual knowledge of freshman college chemistry students. The CCI also covers a range of topics common to a first-semester general chemistry course: the particulate nature of matter; properties of atoms; bonding; gases; liquids and solutions; conservation of mass and atoms; symbols, equations, and stoichiometry; chemical reactions; heat and temperature; phase changes; and macroscopic versus atomic and molecular properties. The CCI is a 22-item multiple-choice assessment that draws on the particulate, macroscopic, and symbolic view of chemistry although not as explicitly as the SAP. Unlike the FCI, which focuses on a single concept (force) and topic (Newtonian mechanics), the SAP and CCI cover many topics and concepts. This likely reflects a major difference between physics and chemistry education at present. Whereas physics is structured around widely accepted organizational principles, such as force, motion, and energy, chemistry is more topic-driven. In other words, it is organized by major topics like stoichiometry, gas laws, atomic structure, solutions, the periodic table, states of matter, and the like, which are based more on historical traditions than underlying concepts. Since the SAP and CCI target such a large range of chemistry topics with such a small number of questions, they cannot target any one concept with a substantial number of items. Therefore, while both instruments may provide some indication of overall understanding of chemistry, they cannot assess student understanding of any single concept in the way that a more narrowly focused instrument, like the FCI, can.

Our instrument is intended to target a single concept, which we argue can be seen as one of the central, organizing ideas in chemistry: the particulate nature of matter. That matter is composed of discrete particles, which combine in predictable ways and rearrange during chemical reactions, is central to all of chemistry. Our instrument focuses on these ideas. It does not attempt to measure the breadth of topics and concepts represented in either the SAP or CCI.

Another issue with any forced-choice assessment is that none of the provided answers may match the students' actual representation of the chemistry addressed by the question. For example, consider the question in Figure 1 from the CCI designed to assess students' ability to translate from a symbolic representation of a chemical reaction to a particulate representation. The design of such questions is limited by an incomplete set of choices of the ways in which students attempt to make sense of particulate ideas. The existing research studies cited above detail some of the ways students attempt to represent their ideas about the particulate world but likely do not provide an exhaustive indication of the ways in which students attempt to represent the particulate nature of matter.

The diagram represents a mixture of S atoms and O^2 molecules in a closed container.



Which diagram shows the results after the mixture reacts as completely as possible according to the equation?

$$2S + 3O_2 \rightarrow 2SO_3$$

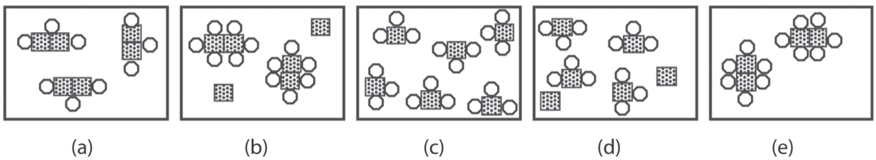


Figure 1. CCI sample item using particulate nature of matter.

Our study is intended to provide a broader depiction and more complete understanding of the different ways in which students attempt to represent the particulate nature of matter. Such an understanding aids the identification of specific areas of difficulty for which instructional interventions could be designed, and the present study's purpose is to provide a starting point to address that need. It supplies a report of the qualitatively different ways in which high school chemistry students attempt to represent a chemical reaction at the particulate or atomic/molecular level.

Data Collection

The data for the present study were collected as part of a larger evaluation study, whose purpose was to evaluate a high school chemistry curriculum. The sample, while not necessarily representative of all U.S. high school chemistry students, represents a very large and diverse group, so it is reasonable to believe that the results reported here would generalize widely. The overall sample comprises a total of 4,315 students from 61 high schools in 106 teachers' classrooms from across the U.S. The sample included slightly more girls than boys, with approximately two thirds self-identifying as Caucasian. About 20% of the students spoke a language other than English in their homes, and 90% reported they planned to attend and complete college.

The data that we present are based on student responses to our new assessment item focused on particulate knowledge related to chemical equations (see Figure 2), which appeared on one of the two forms of the posttest for the larger evaluation project. Of the students who received this form of the test at the end of their chemistry course, 1,337 students provided some response to the second part of the question, with 832 students providing no response. The reported data are drawn from the responses of these 1,337 students.

Methane gas (CH_4) reacts with oxygen gas (O_2) to produce carbon dioxide (CO_2) and water (H_2O). This reaction is represented by the unbalanced chemical reactions below:

$$\_\_CH_4(g) + \_\_ O_2(g) \rightarrow \_\_CO_2(g) + \_\_ H_2O(l)$$

1. Write the appropriate numbers in the blanks to balance the chemical equation.
2. In the space below, draw diagrams that represent what you think you might see if you were able to see the atoms and molecules involved in the chemical reaction above. Remember to draw the correct number of atoms and molecules for each reactants and each product.

Figure 2. The test item used to elicit students' particulate representations of a chemical reaction.

Data Analysis

Phenomenography (Marton, 1981) informs an analytical framework for identifying qualitatively different categories of description among the large number of responses (n = 1,337) considered in this study. As is common in phenomenographic analyses, the present study sacrifices some of the depth and richness in individual responses that is often associated with qualitative traditions in order to capture the breadth of variations in students' representations of a chemical equation at the particulate level (Pang, 2003). In other words, the large number of responses provided a very broad look at the variation among students' ways of representing the particulate nature of matter. The strength of this approach is that it achieves a breadth that would be impossible, practically speaking, with more in-depth qualitative approaches (e.g., involving in-depth interviews) or forced-choice items. However, it should be noted that only surface-level information (i.e., drawings without further explanation) was collected from each participant, so there is no way to do more than infer the ideas/conceptions/experiences which truly underlie the student's drawings. An experienced chemistry educator will likely be able to speculate about what ideas/conceptions underlie some of the categories of description detailed below. However, it is important to keep in mind that the present dataset does not contain any information that could be used to confirm or disconfirm such inferences, though that might be a fruitful avenue for future study.

Student drawings for the equation-balancing item were examined and discussed by the first three authors in order to develop a coding guide. Development of the coding guide was an iterative process. Initial categories were informed by previous research, as described above (e.g., instances of apparent subscript/coefficient confusion), but lengthy discussions among the authors, and several rounds of selecting and test coding subsets of the data, eventually led to consensus about themes, subthemes, and subcategories in the data, and the coding guide was structured around these.

Once consensus was reached on a final version of the coding guide, the entire set of 1,337 drawings was randomly distributed among the first three authors for coding. A randomly selected subset of about 10% of drawings (138) was scored by all three authors to check for consistency. Finally, as a further check on the validity of the scoring guide and the reliability of the three coders, the fourth author reviewed all of the coded drawings, noting any discrepancies. Any discrepancies noted in this final step were discussed by all four authors until consensus was reached about appropriate codes.

Finally, the coefficients in the balanced equation were scored for correctness. Three categories were identified: correctly balanced, balanced but coefficients not reduced, and unbalanced. Categories for the symbolic and particulate responses were cross-referenced to look for relationships between student responses in the two areas.

Results

Symbolic Representations

Of the students who provided codable drawings, 65.3% correctly balanced the equation (58.9% gave completely correct responses; 6.4% gave responses that showed proper ratios of all reactants and products, but did not reduce the coefficients), 32.4% gave incorrect responses, and 2.4% did not respond to the balancing/symbolic portion of the test item.

Particulate Representations

All diagrams were grouped into one of six broad themes: (a) Particulate Representations with Discrete Atoms, (b) Mechanistic, (c) Inappropriate Particulate Representations, (d) Quasi-particulate Representations, (e) Macroscopic Representations, or (f) Irrelevant Attempts. These themes are briefly described in Table 1, which also includes examples of student responses in each theme. In the next section of the paper, we describe each theme and its respective subthemes and subcategories in detail and include sample drawings and pedagogical inferences.

Particulate Representations with Discrete Atoms

Approximately one-third (31.1%) of student responses showed representations that correctly matched the given individual molecular formulas. These diagrams displayed scientifically relevant representations of all chemical species in the reaction, meaning all molecules were represented with the correct numbers of constituent atoms, and connectivity among atoms was consistent with the given molecular formulae. The two primary subcategories within this subtheme were: balanced and unbalanced (see Table 2).

Table 1
Description and Frequency of the Six Major Themes from Student Drawings

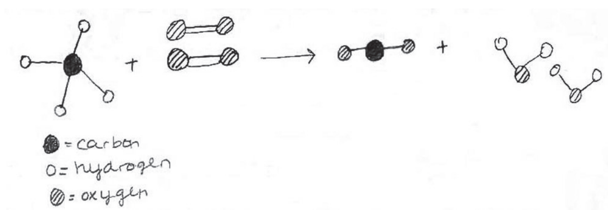
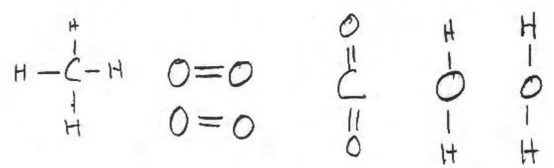
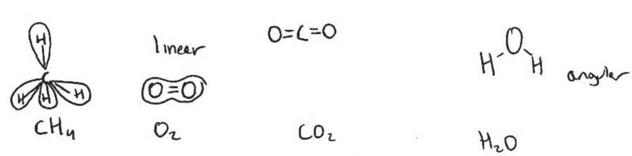
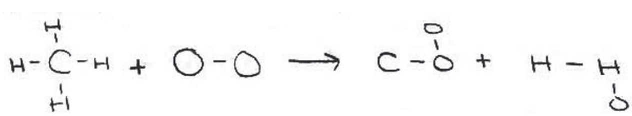
Theme	Description	Number of student responses*
Particulate Representations with Discrete Atoms	Responses included in this theme represented all of the species in the given chemical equation in a way consistent with the given molecular formulae. It is important to note that all of the representations in this theme are not necessarily correct in every detail. Details like central atoms (e.g., H_2O represented as H-H-O) and molecular geometry (e.g., linear water) were ignored for this broad categorization but are included in the discussion of student drawings.	31.1%
Inappropriate Particulate Representations	Responses in this theme included drawings with discrete atoms; however, the atoms were not grouped in a manner consistent with the chemical equation and/or the given molecular formulae within the equation.	46.8%
Quasi-particulate Representations	These representations sometimes contained elements of an appropriate particulate representation, but some parts did not show discrete atoms (i.e., they inappropriately showed multiple atoms as a single particle such as O_2-C-O_2).	5.9%
Mechanistic	In addition to elements that fit one or more of the other themes, mechanistic responses included some indication of a reaction mechanism (e.g., arrows depicting the rearrangement of atoms or intermediate steps in the reaction)—although students were not explicitly asked to do this.	4.0%
Macroscopic Representations	A number of representations depicted substances at a macroscopic level. Some of these were depictions of macroscopic properties (e.g., squiggly lines for gases and droplets for liquids). Others involved macroscopic containers (e.g., beakers or burettes).	3.8%
Irrelevant Attempts	A large number of responses represented no real attempt to address the question (note: this is distinct from leaving the response area completely blank). Some responses in this theme included a variety of scientific-looking representations that did not include details to represent any of the symbols in the chemical equation. For example, drawings included graphs or generic atoms represented as Bohr models. Other responses lacked any scientific detail, including silly drawings and written excuses for not addressing the question.	24.6%

*Note this column does not sum to 100% as some drawings were coded into multiple subcategories.

The distinction between balanced and unbalanced diagrams simply accounted for a student's overall attention to conservation of mass. It is interesting to note that such a large number of students did not conserve mass in their particulate drawings and that the drawings were mismatched with the coefficients from the balanced chemical equation. With our data, it is difficult to tell whether these students were displaying a misunderstanding of the coefficients in the equation or simply did not pay attention to the instructions to include appropriate numbers of each molecule. However, we speculate that the number of students in this category suggests it is more than an oversight. Possibly students are applying an algorithm to balance the equation at the symbolic level without recognizing the meaning of the coefficients in terms of the number of particles represented by the equation. This would be consistent with Johnstone's (1991) assertion that students have trouble moving between "worlds," so they apply one way of thinking to the symbolic world and a different way of thinking to the submicroscopic world, without connecting the two.

Not all of the representations shown in Table 2 are completely accurate; for example, we noted errors in central atom, bond order, and molecular geometry. While the question prompt did not ask for a level of detail that included correct bond order and molecular geometry, it is interesting that many students chose to incorporate these features. This provides both researcher and teacher with a secondary level of analysis from this prompt.

Table 2
Sample Responses Within the Discrete Atoms—Non-Mechanistic Theme

Balanced (16.3%)	i)	
	ii)	
Unbalanced (14.8%)	i)	
	ii)	

Inappropriate Particulate Representations

Forty-five percent of student responses were representations that incorporated discrete atoms within the individual diagrams; however, these representations did not reasonably match the given molecular formulae. In other words, the drawings indicate some appropriate understandings about the particulate nature of matter, but betray other, unscientific ideas or misunderstandings. Representations within this theme fell into four subthemes with seven subcategories (see Table 3) that displayed either inappropriate connections between atoms within individual molecules—formula errors (14.1%), or inappropriate groupings of atoms into individual molecules—amalgams (11.7%) and flocking (19.2%). The fourth subcategory included incomplete representations (1.8%). Examples of the subcategories within each of the four subthemes are described and illustrated in Tables 3 though 6 and in the text that follows.

Table 3
Description of Subthemes Within the Inappropriate Particulate Representation Theme

Subtheme	Subtheme description	Subcategories
Formula Errors	Drawings that showed clear particulate representations depicting discrete atoms bonded in discrete molecules, but with one or more species that clearly did not correctly match given molecular formulae. For example, drawings that represented water as HO_2.	Dyslexic Water (7.7%) Formula Mismatch (6.4%)
Flocking	These representations depicted individual species as molecules with discrete atoms, but their connectivity included groupings that did not match the molecular formulae. For example, $2H_2O$ depicted as H-O-H-H-O-H.	Molecular Flocking (13.1%) Atomic Flocking (6.1%)
	Flocking representations and amalgam representations both depict inappropriate groupings of species; however, flocking representations indicated segregated groupings of like species, whereas amalgam representations depicted combinations of unlike species.	
Amalgams	These representations showed discrete atoms but depicted groups of atoms not organized as distinct molecules from the chemical equation. In some cases atoms were grouped as distinct products and reactants; whereas, in other cases, all atoms were grouped into a single, large "molecule."	Quasi-Amalgam (2.2%) Morphing Amalgam (4.9%) Static Amalgam (4.6%)
Incomplete	A number of drawings represented part, but not all, of the given chemical equation. Some of these included only products or only reactants. Others included drawings of 3 of the 4 species involved in the reaction. Typically, what was included in these drawings appropriately matched parts of the given equation, but not the entire equation.	Incomplete (1.8%)

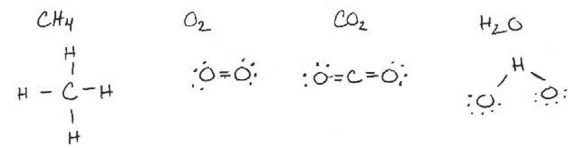
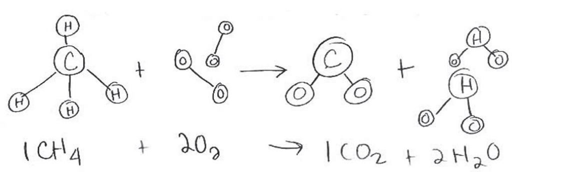
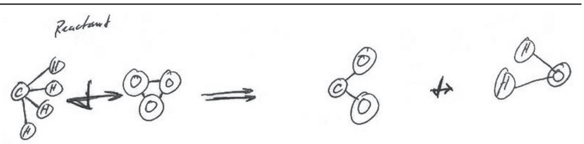
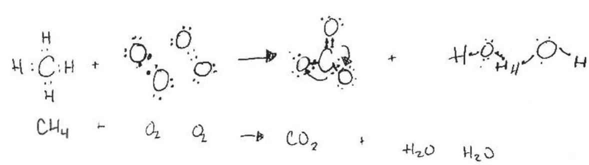
Formula errors. Subcategories such as dyslexic water and formula mismatch are illustrated in Table 4. These drawings showed clear particulate representations depicting discrete atoms bonded in discrete molecules, but with one or more molecules not matching the correct molecular formulae. One of the most striking of the subcategories depicted water with a dyslexic formula (i.e., as HO_2). This representation was too common to be simply attributable to sloppiness in labeling. It was also reported in a study by Keig and Rubba (1993), who referred to this as "formula error-ratio of atoms." Students are incorrectly associating the subscript with the following rather than preceding atom in the molecular formula for water. Students making this error do not appear to have problems with CH_4 and CO_2. When speaking the formula for water out loud, we say "H two O" which is possibly misinterpreted and students are mentally attributing the two to the oxygen.

Other drawings depicted molecules containing atoms, which matched those in the given formula, but there was a mismatch between the numbers of atoms in the molecule and subscripts in the given equation. For example, a number of students depicted O_2 as being composed of 3 particles, which were typically labeled as oxygen molecules (i.e., they drew O_3 instead of O_2). Other examples included representations of carbon dioxide as a carbonate ion. Students appear to be drawing on familiar examples seen across the duration of their course rather than transferring particulate ideas to the question at hand.

Flocking errors. This category was seen in those diagrams that depicted each species in the chemical equation as a set of discrete atoms; however, while the molecules were shown as separate entities, the connectivity among the atoms was inappropriate. The two common ways of depicting flocking included molecular flocking and atomic flocking (see Table 5).

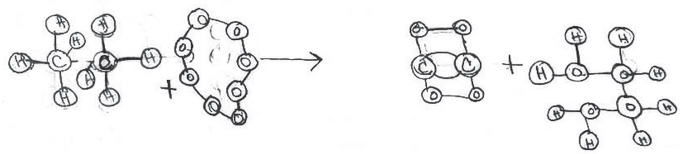
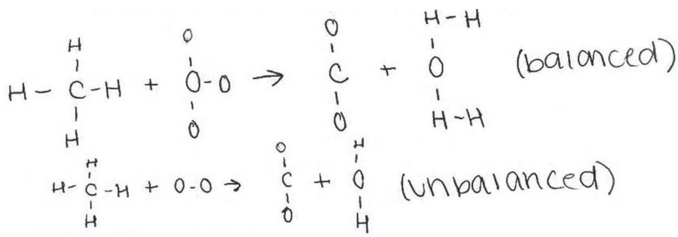
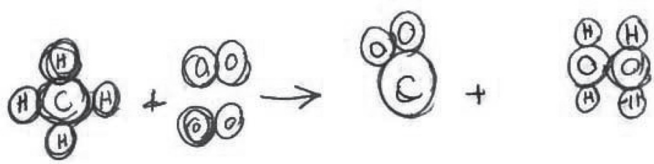
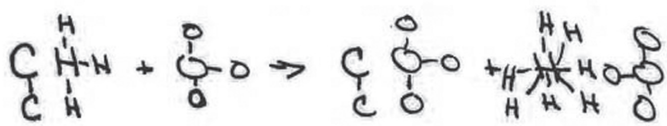
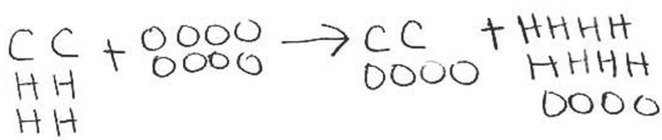
Table 4

Sample Responses from Formula Error Subtheme

Inappropriate Particulate–Dyslexic Water (7.7%)	i)	
	ii)	
Inappropriate Particulate–Formula Mismatch (6.4%)	i)	
	ii)	

Molecular flocking. These representations were characterized by all molecules of a given species being connected (e.g., $2H_2O$ depicted as H-O-H-H-O-H). Two prevalent types of molecular flocking exist: one in which all species are flocked and another in which only water is flocked (see Table 5). In the first type, we believe students may misunderstand the meaning of the coefficients in a balanced chemical reaction. In effect, students are treating the coefficients as another form of subscript, a misunderstanding also noted by Yarroch (1984). This may be a more sophisticated

Table 5
Sample Responses in the Flocking Subcategory

Inappropriate Particulate—Molecular Flocking (13.1%)	i)	
	ii)	
	iii)	
Inappropriate Particulate—Atomic Flocking (6.1%)	i)	
	ii)	

type of “atom accounting” error (see description of “Atomic Flocking” below), in which the student has some appropriate understanding of the particulate nature of matter, but is inappropriately reconciling it with equation-balancing algorithms. Second, we note that the molecule most problematic for students in all identified themes, not just molecular flocking, is water. Students made numerous errors with water that appeared to mimic actual chemical phenomena that teachers may have presented in class. For example, in some cases we hypothesize that students may be attempting to represent the intermolecular forces between water molecules.

Atomic flocking. These representations depicted all atoms of a given element as being interconnected, but segregated from other types of atoms (e.g., CH_4 depicted as C H-H-H-H). These representations appear to relate to the pedagogical technique of taking an “atom inventory” where molecules are broken into their constituent atoms for the purpose of counting “atoms” to adhere to the law of conservation of mass. Chemistry teachers commonly present equation balancing in this way: as simply a process of “counting up all of the Hs, Cs, and Os on one side of the equation and making sure they equal the numbers on the other side.” It is easy to see how this presentation would lead students into an algorithmic way of thinking about chemistry. Our results suggest this kind of teaching approach can also hinder students’ development of an appropriate understanding of the particulate nature of matter.

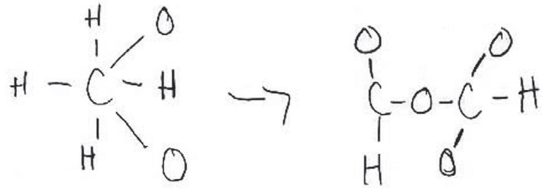
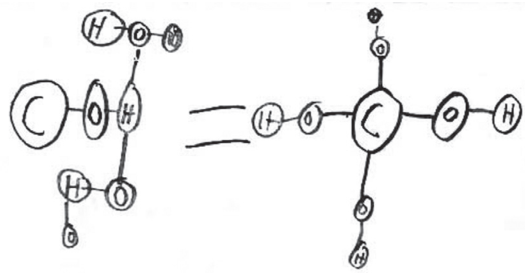
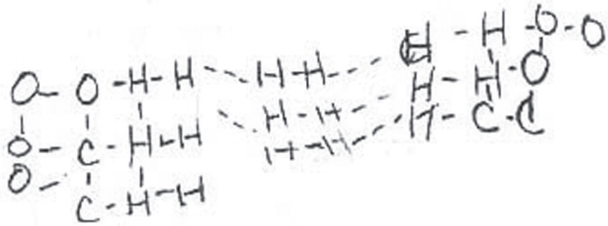
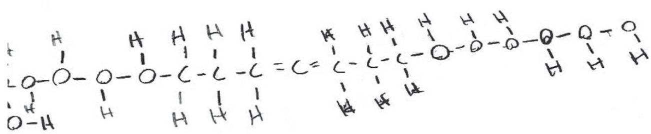
Amalgams. These representations showed discrete atoms in groupings of inappropriate or indiscernible molecules. Unlike flocking representations, where like species were shown as distinct groupings, amalgam representations depicted combinations of unlike species, indicating these representations may have originated from completely different understandings of how atoms and molecules combine in chemical reactions. Three different types of amalgam representations were evident: morphing, static, and quasi (see Table 6). In a morphing amalgam, all of the reactant atoms were combined in one bonded mass, and all the product atoms were combined in a separate bonded mass. Typically, the atoms were rearranged somewhat in space between products and reactants, whereas, in a static amalgam, only one bonded mass of atoms was depicted—so products could not be distinguished from reactants. A few morphing amalgams (about one-fifth of them) showed conservation of mass. However, most morphing and all static amalgams appeared to show a random number of atoms arranged artistically rather than with any chemical foresight.

A quasi-amalgam included depictions of some discrete molecules (often the reactant molecules) and some amalgamated mass of atoms (often all of the products bonded together in one large mass). As such, the quasi-amalgam incorporated elements of an appropriate particulate representation with discrete atoms as described above and elements of a morphing amalgam.

Implications of amalgam representations could be gleaned from students’ instructional experiences where chemical reactions and chemical kinetics are conflated. While it is understood that in “real life” single molecules and atoms do not simply change into a new species, chemical reactions are generally shown as single constituent molecules and atoms combining. There are even instances where

Table 6

Sample Responses in the Amalgam Subcategory

Inappropriate Particulate–Morphing Amalgam (4.9%)	i)	
	ii)	
Inappropriate Particulate–Static Amalgam (4.6%)	i)	
	ii)	
Inappropriate Particulate–Quasi-Amalgam (2.2%)		

molecular transition states are represented and a molecular species showing all molecules and atoms in a chemical reaction combined, indicating the reactant atoms joining and then separating into products. This combination of ideas can be confusing to students trying the make sense of the notion of the size of an atom and molecule and what is generally seen at the macroscopic level.

Quasi-Particulate

Quasi-particulate diagrams accounted for 5.9% of codes and included representations that displayed some form of particles that did not show discrete atoms (i.e., they inappropriately showed multiple atoms as a single particle). As shown in Table 7, representations in this theme fell into three distinct subthemes: diatomic particles (1.5%), multi-atomic particles (1%), and morphing particles (3.4%). In diatomic particle representations, students depicted diatomic elements as a diatomic "particle" within a molecule. In other words, instead of drawing molecular oxygen as a pair of oxygen atoms, it was depicted as a single particle labeled "O_2." In learning the rules of chemistry, some of the common general chemistry facts to memorize are the seven diatomic elements (i.e., Br, I, Cl, F, O, N, and H). Many times students are instructed to make up ways to memorize these elements so they can easily access them to predict the possible products when given the reactants and when writing out a chemical reaction. In these diatomic depictions, students appear to be extending the rule that oxygen and hydrogen are diatomic to compounds not just elements.

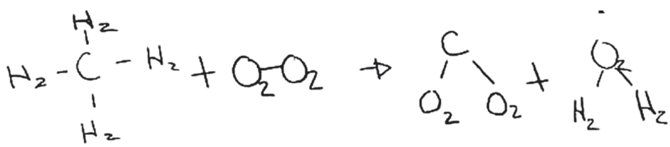
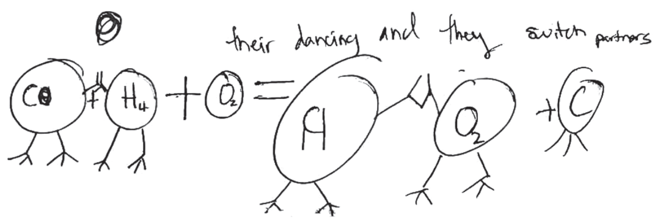
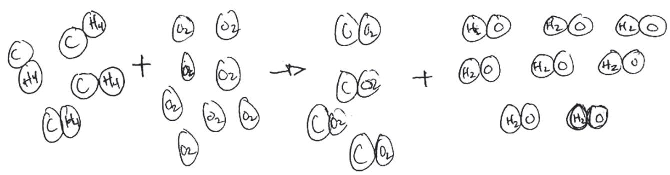
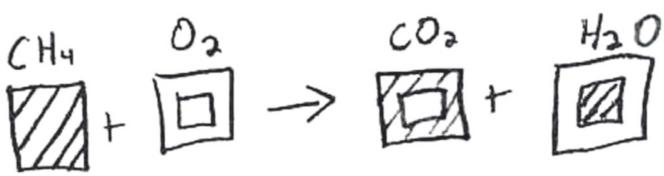
Less commonly depicted multi-atomic groups of atoms, such as H_4 or CH, were represented as single particles. It appears that students are inappropriately applying rules from ionic reactions to the example of methane combustion. Students may be forcing the particles into a configuration that matches a double displacement reaction.

In other cases, no discrete atoms were shown. Instead, polyatomic molecules were represented as a single particle, which were shown to change, or morph, into different polyatomic "particles." Drawings of these morphing particles depict products and reactants separately, but showed no difference between them except a different label. An inference made here is that these drawings are meant to depict one substance simply changing identity into another—like computer-generated morphing. We note that, in some cases (see Table 8), students identify a mechanism for this morphing. In these examples, students are showing thinking, although not always chemically accurate, at the macroscopic level rather than particulate level.

Mechanistic

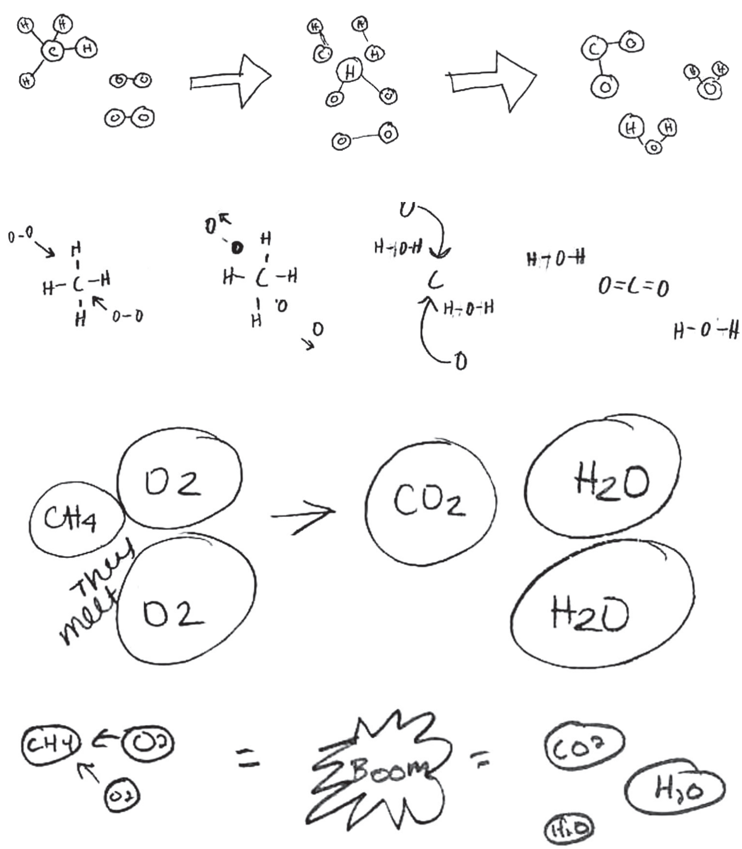
Four percent of students provided responses that included a mechanistic representation of the chemical reaction. All mechanistic drawings were double-coded, meaning that these drawings were coded as mechanistic and at least one of the other themes (overall only 10.7% of drawings fit more than one theme). Mechanistic drawings that were also coded as particulate mechanistic drawings invoked a process, using tools such as mechanistic arrows, movement or transition states, as exemplified by the drawings in Table 8. Our initial hypothesis was that we may have been observing a teacher effect and that these drawings would be grouped into a small number of classrooms in which the teacher had discussed mechanisms. However, these mechanistic representations are spread across classrooms and appear to show individual students' deeper thought processes in representing the chemical reaction.

Table 7
Sample Responses in the Quasi-Particulate Subtheme

Quasi-Particulate–Diatomic Particle (1.5%)	
Quasi-Particulate–Other Multi-Atomic Particle (1%)	 
Quasi-Particulate–Morphing Particles (3.4%)	 

Interestingly, almost one quarter of the morphing-particles representations also invoked some form of mechanistic thinking. In most cases, these mechanisms drew from the macroscopic realm invoking macroscopic processes such as melting and explosions. While such drawings do not demonstrate sophisticated views of the particulate nature of matter—and in some cases inaccurate macroscopic views of the reactions—these ideas should not be discounted. In a recent critique of assessment in science education, Russ, Coffey, Hammer, and Hutchison (2008) emphasize the importance of mechanistic thinking. Their critique of traditional assessment practices in the science disciplines is that too often teachers "judge the quality of ideas by comparing them to the canon as represented by the curriculum. In other words, making a judgment about whether the idea is right or wrong" (p. 876). Mechanistic drawings seem to represent students' personal attempts at sense-making, rather than simply a memorized "right answer."

Table 8
Sample Responses Within the Mechanistic Theme

Mechanistic (4%)	

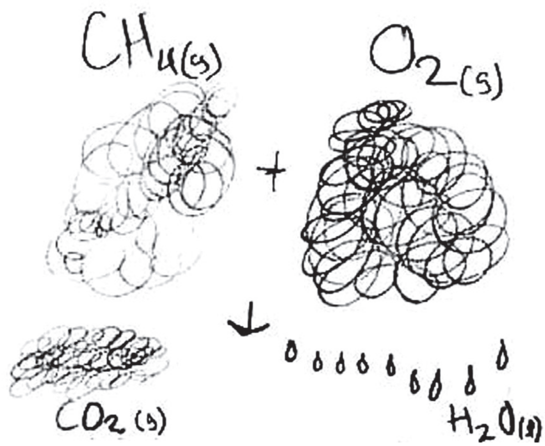
Macroscopic

Only 3.8% of responses were coded as macroscopic. Examples of the subthemes are provided in Table 9. Macroscopic representations depicted substances in various ways at a macroscopic level. Some of these were depictions of macroscopic properties (e.g., squiggly lines for gases and droplets for liquids).

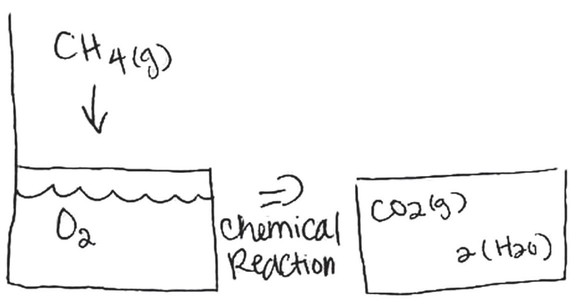
Table 9
Sample Responses in the Macroscopic Theme

Non-Particulate–Macroscopic (3.8%)

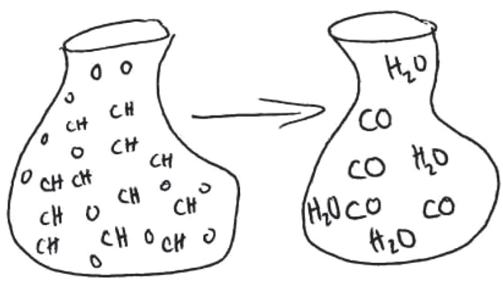
i)



ii)



iii)



Others involved macroscopic containers (e.g., beakers or burettes). Often inside of such containers were particulate drawings of atoms and molecules. In these drawings, students are blurring the macroscopic and particulate worlds of chemistry.

Irrelevant Attempts

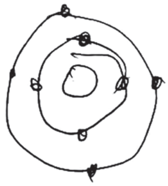
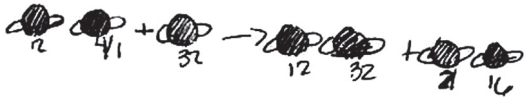
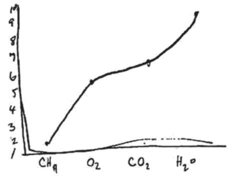
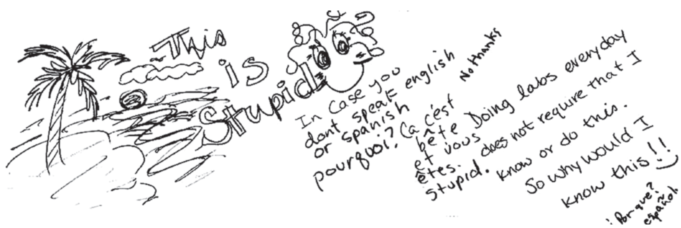
Almost one-quarter (24.6%) of responses represented no real attempt to address the question (note: this is distinct from leaving the response area completely blank). These responses (see Table 10) ranged in the representations that provided a scientific looking drawings, nonsensical depiction/drawing (e.g., a large "molecule" arranged into the shape of a question mark) to textual comments (i.e. "I love chemistry, not...) and excuses (e.g., "we did this several months ago and I don't remember how to do it anymore").

Scientific-looking responses included a variety of diagrams that contained scientific images but did not include details relevant to the chemical equation. Some were textbook-like and bore resemblance to Bohr-models or other atomic models, but had little or no connection to the given equation. Others depicted popular atomic models resembling the nuclear radiation symbol. Still others were graphs or charts (e.g., one resembled a titration curve). This variety of responses in the subtheme of scientific-looking drawings probably represents a number of subcategories. It was beyond the scope of the present report to code each of these subcategories separately, but most did not seem to represent a relevant attempt to address the question at hand. Most of these drawings appeared to be desperate attempts to generate some scientific-looking response. They are likely more a reflection of what students believe science is (i.e., the nature of school science) rather than their understandings of scientific ideas.

Several respondents provided a written excuse or commentary for not addressing the question (see Table 10). The students' comments provide a window into the classroom, and while they are the comments of an individual, they allow for some inferences about classroom practices as experienced by these students. Some students offered that they had a vague recollection of learning about the particulate nature of matter, but that it was too long ago to remember. Clearly the particulate nature of matter was not treated as an important and central concept in chemistry that was revisited throughout the course. Like other researchers before us (Bunce & Gabel, 2002), we argue that attention to the particulate nature of a matter is not a single activity or unit but that attention must be paid to the particulate nature of matter throughout the course, by making connections (intentionally) between various areas.

Another common practice in high school chemistry is the prevalence of laboratory activities. Yet unfortunately, student learning opportunities are often missed during laboratory work. As one student states, "Doing labs everyday does not require that I know this." Laboratory work holds a sacred place in the chemistry curriculum; indeed, few chemistry researchers and teachers would question the belief that students should experience a significant amount of laboratory work. However, as Hodson (1988) argues, "The actual performance of the experiments contributes very little" to

Table 10
Examples of Responses from the Irrelevant Attempts Theme

Scientific Looking	i)	
	ii)	
	iii)	
Irrelevant Commentary	i)	
	ii)	uh, we did this 8 months ago, I haven't a clue
	iii)	

student learning. Laboratory activities tend to focus on data collection with minimal or no attention to data analysis and the development of explanations from data (Domin, 1999), which is where significant learning should occur related to the particulate nature of matter. Unfortunately, most laboratory activities implemented in high school chemistry classrooms serve the purpose of "edu-tainment"—they provide the "wow" and excitement of science without contributing directly to the learning of conceptual knowledge (Hodson, 1988).

Multi-Representational Understanding

Table 11 shows the relationship between students' ability to correctly represent the reaction symbolically (insert the correct coefficients to balance the equation) and students' ability to represent the reaction appropriately at the particulate level. It can be seen that those students who provided reasonable particulate drawings were significantly more likely to correctly balance the equation, whereas those students who provided irrelevant drawings were significantly more likely to incorrectly balance the equation. There were no significant differences in balancing for drawings in any of the other themes. If students are able to appropriately represent the particulate nature of matter, they are more likely to be able to correctly balance a chemical equation. However, it is clear that students can correctly balance a chemical equation without an appropriate understanding of the particulate nature of matter. Therefore, instructors cannot assume that an ability to correctly balance a chemical equation implies any understanding of the meaning of that equation. As expressed by a student in Figure 3, it's just simple math to balance the equation, but this algorithmic skill does not extend to particulate representations.

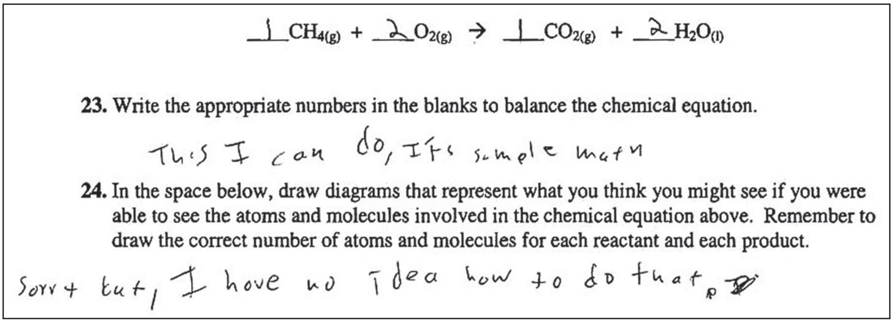
Conclusions and Implications

Unlike conventional equation-balancing assessment items, the equation-drawing task we presented here provides highly nuanced information about students' thinking

Table 11

Comparison of Performance on the Equation-Balancing Task Versus Themes into Which Drawings Were Coded

Theme	**Balanced Incorrectly**	**Balanced Correctly**
Particulate Representations with Discrete Atoms	60	352
Inappropriate Particulate Representations	192	426
Quasi-particulate Representations	38	38
Mechanistic Elements	9	43
Macroscopic Representations	24	20
Irrelevant Attempts	218	108



$\underline{1}\ CH_{4(g)} + \underline{2}\ O_{2(g)} \rightarrow \underline{1}\ CO_{2(g)} + \underline{2}\ H_2O_{(l)}$

23. Write the appropriate numbers in the blanks to balance the chemical equation.

This I can do, It's simple math

24. In the space below, draw diagrams that represent what you think you might see if you were able to see the atoms and molecules involved in the chemical equation above. Remember to draw the correct number of atoms and molecules for each reactant and each product.

Sorry but, I have no Idea how to do that.

Figure 3. Sample student response describing the mathematical simplicity of equation balancing.

and understandings about key chemistry concepts. While 65% of students correctly balanced the chemical equation, only 31% drew representations of the reaction that showed appropriate knowledge at the particulate level. For those whose drawings did not represent appropriate particulate understanding, it was apparent that many were struggling to make sense of information they had seen in class or in a textbook and were misapplying or misinterpreting it in a variety of ways. Others seemed to struggle to reconcile their (macroscopic) experience with particulate-level ideas, while a substantial number failed altogether to provide relevant drawings. This draws into question the value of many common experiences in chemistry classes. Of those students who did provide drawings that reasonably represented the meaning of symbols explicitly represented in the given chemical equation, many chose to include other implicit details in their drawings (such as bond-orders and molecular geometry) and, in many cases, they demonstrated misunderstandings of other, related chemical ideas. A traditional chemical equation-balancing task is not sensitive to most of these misunderstandings, nor can it reveal the fairly sophisticated ways some students attempt to make sense of chemical ideas on their own such as mechanistic thinking.

As a research tool, the equation-drawing task provides a window into ways students struggle to understand the particulate nature of matter. Many students seem to misapply particulate ideas in inappropriate contexts. In some cases, the misapplication is in a fairly closely related context, such as the "flocked" water molecules seen here, which seem to mimic intermolecular forces. Since the given chemical equation indicated liquid water was produced, it is not unreasonable to attempt to depict hydrogen-bonding here. However, other misapplications took chemical ideas much farther out of context, such as the incorporation of irrelevant chemical species like a carbonate ion. Based on details included in these drawings—like the indications of resonance in CO_3^{2-}—it seems clear students knew quite a bit about the species they chose to include, but did not understand it in a way that they could transfer conceptual understanding to new contexts. Even farther out of context were many of the "scientific looking" drawings, which seem to be more a reflection of students' perceptions about what is science or their feelings about how chemistry is presented in the classroom. Many students' comments are an indictment on common methods of teaching chemistry.

Students also appear to misinterpret common ways that teachers (and textbooks) present key ideas in chemistry. For one of the drawings in Table 7, it is easy to imagine a teacher explaining double-displacement reactions saying that it is like "they're dancing and they switch partners." For many of the "flocking" drawings, one can imagine students were taught to balance chemical equations using an atom-accounting algorithm (counting atoms on both sides of the equation to make sure numbers match), which leaves them unable to distinguish the meaning of the subscripts in a chemical equation from the meaning of the coefficients. There is also evidence that students do not fully understand chemical heuristics, such as rules about diatomic gases, because they apply them out of context. These are all examples of common ways teachers (and textbook authors) attempt to simplify the task of equation balancing, which may help students succeed at the symbolic level, but clearly do not promote appropriate understandings at the particulate level. While it seems clear more attention needs to be paid to helping students make connections among the symbolic, macroscopic, and particulate worlds, it should be noted that attempts to make these connections can be misinterpreted as well. Many drawings falling into the "amalgam" subtheme may have arisen from visualization materials included in many recent textbooks, which are intended to depict processes at the particulate level.

In completing the equation-drawing task, many students' drawings suggest they were working to make sense of chemical ideas on their own, rather than simply applying algorithms or reproducing information they had seen in their book or from their teacher. However, some of these attempts were more successful than others. Many "macroscopic" drawings did attempt to connect relevant, concrete experiences with components of the given chemical equation—as mostly, indications of states of matter. Likewise, many "morphing particle" drawings seem to arise from students' attempts to reconcile the notion of matter being composed of "particles" with their macroscopic observations of one set of substances combining, reacting, and becoming a new set of substances. While neither of these types of drawings demonstrated a sophisticated understanding of the particulate nature of matter, it is noteworthy that they do seem to represent students having taken the initiative to try to understand the material on their own rather than simply regurgitating memorized information. On the more successful end of the spectrum, some "mechanistic" drawings displayed not only an appropriate understanding of the particulate nature of matter, but also a sophisticated thinking about the interactions of atoms at the particulate level. (However, it should be noted that not all mechanistic drawings represent sophisticated understanding.)

As a classroom assessment tool, the equation-drawing task can provide teachers with deeper insights into their students' learning and may be especially suited to formative assessment. Students' drawing can be taken as indicators of specific ways in which they misapply or misinterpret the presentations of chemical ideas or struggle to make connections among the three "worlds" of chemistry. Teachers can then design instruction to address the specific ways in which their students struggle to understand the particulate nature of matter. Students' drawings can also provide a second layer of assessment, as many drawings also indicate students' level of understanding of related ideas, such as a bond-orders, molecular geometry, and central atoms.

This study offers the first phase of work needed to understand students' conceptions of chemical equations at the particulate level. With the present study, we provide a catalog of the many different ways in which students represent chemical equations at the particulate level, and we begin to infer students' ways of thinking and instructional issues that underlie these representations. Future research needs to be conducted to better understand students' conceptions through interviews. Research is also needed that includes a more detailed description of instruction that can provide guidance on instructional practices with promise in developing students' knowledge at the particulate level and students' ability to translate between representational levels.

REFERENCES

Bodner, G. M. (1991). I have found you an argument: The conceptual knowledge of beginning chemistry graduate students. *Journal of Chemical Education, 68,* 385–388.

Bunce, D. M., & Gabel, D. (2002). Differential effects on the achievement of males and females teaching the particulate nature of matter. *Journal of Research in Science Teaching, 39*(10), 911–927.

DeMars, C. E. (2000). Test stakes and item format interaction. *Applied Measurement in Education, 13,* 279–299.

Domin, D. (1999). A content analysis of general chemistry laboratory manuals for evidence of higher-order cognitive tasks. *Journal of Chemical Education, 76*(1), 109–112.

Gabel, D. L. (1998). The complexity of chemistry and implications for teaching. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 233–248). Dordrecht, Netherlands: Kluwer Academic Publishers.

Gabel, D. L., & Bunce, D. M. (1994). Research on problem solving: Chemistry. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 310–326). New York: Macmillan.

Hamilton, L. S. (1998). Gender differences on high school science achievement tests: Do format and content matter? *Educational Evaluation and Policy Analysis, 20,* 179–195.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141–151.

Hodson, D. (1988). Experiments in science and science teaching. *Educational Philosophy and Theory, 20,* 53–66.

Huffman, D., Thomas, K., & Lawrenz, F. (2003). Relationship between professional development, teachers' instructional practices, and the achievement of students in science and mathematics. *School Science and Mathematics, 103,* 378–387.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning, 7,* 701–703.

Kahle, J. B., Meece, J., & Scantlebury, K. (2000). Urban African-American middle school science students: Does standards-based teaching make a difference? *Journal of Research in Science Teaching, 37,* 1019–1041.

Keig, P. F., & Rubba, P. A. (1993). Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *Journal of Research in Science Teaching, 30*(8), 883–903.

Laguarda, K. G. (1998). *Assessing the SSI's impacts on student achievement: An imperfect science.* Menlo Park, CA: SRI International.

Lockwood, A. T. (1995). A synthesis of four reforms. *New Leaders for Tomorrow's Schools, 1,* 1–39.

Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science, 10,* 177–200.

Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education, 79, 739.*

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1993). *NAEP 1992 math report card for the nation and the states.* Washington, DC: U.S. Government Printing Office.

Nakhleh, M. B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? *Journal of Chemical Education, 70*(1), *52–55.*

Nurrenbern, S. C., & Pickering, M. (1987). Concept learning versus problem solving: Is there a difference? *Journal of Chemical Education, 64,* 508–510.

Pang, M. F. (2003). Two faces of variation—On continuity in the phenomenographic movement. *Scandinavian Journal of Educational Research, 47*(2), 145–156.

Parker, V., & Gerber, B. (2000). Effects of a science intervention program on middle-grade student achievement and attitudes. *School Science and Mathematics, 100,* 236–242.

Penner, A. (2003). International Gender X item difficulty interactions in mathematics and science achievement tests. *Journal of Educational Psychology, 95,* 650–*655.*

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics (NCES 2005-464).* U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Roehrig, G. H., & Garrow, S. T. (2007). The impact of teacher classroom practices on student achievement during the implementation of a reform-based chemistry curriculum. *International Journal of Science Education, 29,* 1789–1811.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2008). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education, 93,* 875–891.

Sawrey, B. A. (1990). Concept learning vs. problem solving: Revisited. *Journal of Chemical Education, 67,* 253–264.

Scantlebury, K., Boone, W., Kahle, J. B., & Fraser, B. J. (2001). Design, validation, and use of an evaluation instrument for monitoring systemic reform. *Journal of Research in Science Teaching, 38,* 646–662.

Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of enactments in reform classrooms. *Journal of Research in Science Teaching, 42,* 283–312.

Schneider, R. M., Krajcik, J., Marx, R. W., & Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of science achievement. *Journal of Research in Science Teaching, 39,* 410–422.

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching, 36,* 1110–1126.

Yarroch, W. L. (1984). Student understanding of chemical equation balancing. *Journal of Research in Science Teaching, 22*(5), 44.

CHAPTER 4

ASSESSMENT FOR TEACHING AND LEARNING IN A NONSCIENCE-MAJOR CHEMISTRY COURSE

Ruth E. Kinder and Teresa A. Johnson
Ohio State University—Lima Campus

Abstract

This chapter explores the assessment process and the impact it has made on both our own teaching and our students' learning. The results from this research were aimed at supporting and improving student learning via evidence-based course redesign and instructor development. Data were collected over several years from a nonscience-major, elementary chemistry course offered at a regional campus of a large, public research university. The assessment instruments used were administered within the university's online course management system (CMS) and have proven to be simple to maintain and easy to use. This assessment process has provided us new ways to understand whether or not we are successfully providing chemistry courses that allow our students to meet our set learning goals.

Introduction/Overview

This chapter explores a multiyear assessment project designed to support and improve student learning via evidence-based course redesign and instructor development. The course connected to the project was a nonscience-major, elementary chemistry course offered on a regional campus that enrolls approximately 1,400 students within a large university that enrolls approximately 55,000 students. The initial goal of this project was to collect and analyze data to determine whether a newly introduced course style (a flexible, hybrid, delivery with online lecture and face-to-face laboratory portions) was at least as effective as the more traditional face-to-face lecture and laboratory style. Recently, this project's goal expanded to include an analysis of whether the course was effectively meeting general education curriculum (GEC) learning goals. The project continues to evolve and grow as initial questions are answered and interesting new questions arise. Future goals include using these preliminary data for investigation of the effectiveness of specific teaching or course activities in both the traditional lecture and hybrid delivery style sections in an effort to directly improve student learning and attitudes.

The assessment instruments used to collect data were chosen so that they could be useful for both the smaller regional campus student numbers and the larger enrollments at the main campus. The university's online CMS quiz and survey tools were used to put the assessment instruments within easy reach of the enrolled students and faculty as well as in a secure data collection environment (http://www.desire2learn.com). The use of the CMS's tools provided easy storage, import/export, edit, and a few basic statistical functions. It also was relatively simple to share the assessment instruments throughout the university-wide system.

The processed data have been used to:

- demonstrate that students made significant gains in chemistry content knowledge;
- demonstrate that students completing the course through a hybrid delivery style ended with approximately the same amount and distribution of content knowledge as those completing the course through a traditional lecture delivery style;
- demonstrate that the students are aware of significant events/concepts/ideas in chemistry; and
- support that the course was meeting the university GEC learning objectives successfully.

Course Description

The course catalog describes Chemistry 101 as an "introductory general chemistry for nonscience majors, covering dimensional analysis, atomic structure, bonding, chemical reactions, states of matter, solutions, chemical equilibrium, and acids and bases" (Ohio State University Office of the Registrar, 2010). Students are asked to use logical reasoning to solve laboratory problems based on chemical data as well as to apply basic computational skills to laboratory data analysis. Students also draw conclusions from the analysis of their data. Chemistry 101 includes exposure to significant historical achievements in the field of chemistry such as the organization of the periodic table, the nature of chemical bonds, and atomic and electronic structure theories, as well as exposure to the methods and techniques of scientific investigation. Connections between the field of chemistry and societal issues are included, when feasible.

At our university, Chemistry 101 resides in the breadth category and natural sciences subcategory of GEC courses. The following GEC goals are listed for a course in this category and subcategory.

GEC Natural Science Breadth Goals

"Courses in natural sciences foster an understanding of the principles, theories, and methods of modern science, the relationship between science and technology, and the effects of science and technology on the environment." (Ohio State University Office of Academic Affairs, 2010).

GEC learning objectives:

1. Students understand the basic facts, principles, theories, and methods of modern science.
2. Students learn key events in the history of science.

3. Students provide examples of the interdependence of scientific and technological developments.
4. Students discuss social and philosophical implications of scientific discoveries and understand the potential of science and technology to address problems of the contemporary world.

With this course description and these general education learning goals, it is reasonable to expect successful Chemistry 101 students to have experienced an educational environment which allowed them to leave the course with a strong grasp of the fundamental content and some of the methods/techniques of elementary chemistry as well as the capability to describe significant achievements in the field of chemistry. Successful students also should have developed some ability to connect achievements in chemistry with societal issues and events.

The Assessment Process

The assessment process began with the content knowledge and attitude goals outlined for the course by the university and an examination of the current syllabi and student learning objectives written for the course. We focused first on measuring knowledge of the fundamental course content. However, it soon became clear that measuring student attitudes or changes in the affective domain would also be important (Seymour, 2002).

Our goal was to gather data to address these main questions: How do we know that students are learning? How can we know if, and what, they know? Student activities and outcomes shown in Table 1 were considered as potential target data (Suskie, 2009; Walvoord, 2004).

Measurements based on attendance were rejected because these measures most directly reflect dimensions such as effort and self-discipline rather than content knowledge. Measurements based on course letter grade or course total points were

Table 1

Student Learning Activities and Outcomes as Possible Data Targets

Are they learning when they…	Did they learn if they…
• show up to class?	• received a good grade in the course?
• listen to a lecture?	• gave the instructor a good rating?
• work alone?	• reported they liked the course?
• work together?	• reported they learned a lot?
• do things?	
• read about things?	
• use computers?	
• complete the assigned course activities?	
• practice?	
• look for more to explore?	
• look like they are having fun?	

also rejected because these are very broad measures and mix effort, self-discipline, and organization with content knowledge in a complex and interdependent manner, thereby making it very difficult to draw conclusions regarding causality due to the number of competing variables (Wiggins, 1998). Although an overall course grade may not be a reasonable assessment data target, it has been suggested that individual assignment grades can be used as part of assessment if explicit criteria are defined which allow the grade to be interpreted in terms of strengths and weaknesses in skills directly connected to clearly specified learning goals (Black & Wiliam, 1998; Seymour, 2002). No specific assignment grades have been used for this assessment project to date, but some were considered and will be used for future measurements once explicit criteria for rating them are established (Walvoord, 2004).

After possible activities, outcomes, grades, attendance, and other factors were considered, the overall assessment task was formulated to include two main components: core content knowledge gains and student self-reported attitudes about, and experiences with, the course.

Assessment instruments were selected to collect data to determine if the students:

- seemed to **know more** about elementary chemistry content after the course than they did before.
- reported **more confidence** in their abilities related to chemistry knowledge and skills after the course than they did before.

Four specific assessment instruments were used:

1. Core Content Knowledge
    - Before Course Content Check (BCC)
    - After Course Content Check (ACC)
2. Attitude and Experiences (levels of confidence in skills, expectations about, and experiences with, specific learning tools and course activities)
    - Before Course Attitude and Expectations Survey (BCA&E)
    - After Course Attitude and Experiences Survey (ACA&E)

Methods

Measuring Content Knowledge

The Content Checks provided a measurement of the core course content knowledge the students had when they entered (BCC), and again when they left (ACC), the course. Most instructors of elementary chemistry courses would be able to reach agreement about the basic content knowledge which should be included, but consensus on which more advanced content knowledge to include would be much more difficult to achieve. For example, all instructors of such a course would likely agree that stoichiometry problems and balancing reaction equations by inspection should be included in the core knowledge, and all would cover these topics in their courses; however, only some would include limiting reagent stoichiometry or balancing through redox half-reaction problems in their courses. With the content questions for this assessment instrument selected specifically to represent only core knowledge, concerns over whether or not a specific topic has been included in the course disappear and allow the assessment instrument to be used by different instructors.

The chemistry content questions were generated from course exams and collections of questions created mainly by one of the authors (Kinder). Over time, additional questions were added to the content question bank by collaborators. These questions would seem very familiar to any chemistry educator and represent the styles often found in textbook and standardized exam question banks. The majority of questions in the question bank would generally be rated as easy to moderate rather than difficult. Their difficulty level matches their defined purpose: to determine minimum, core, content knowledge levels rather than maximum achievement levels. The questions used in the BCC and ACC were organized into groups of course content within 14 basic categories. These categories were established by having an experienced instructor examine the syllabi, course description, textbooks, and exams typically used for this course as it has been taught on the regional campus for over 10 years. The content categories are identified in Table 2.

Table 2

Content Categories

Properties	Balancing Equations
Reaction Types	Acids and Bases
Bonding	Shape, Polarity, Properties, and IMFs (IFAs)
Measurements and Significant Figures	Solutions
Calculations with Formulas and Equations	Periodic Table and Trends
Atomic Structure	Names and Symbols
Electronic Structure	Vocabulary and General Information

Because the content check was meant to be a group sample rather than an absolute test of any one student's knowledge, it was set up to collect aggregate data. For example, all students were asked two questions related to balancing equations. However, there were many possible questions, and any two of these were pulled randomly for a specific student. Even though each student received only 20 questions and completed the content check in a short time period, pulling questions from the extended bank allowed collection of responses to a broad range of questions for each category.

Because both correct responses and responses of "I don't know" were important in the data analysis, special emphasis was placed on the portion of the instructions explaining the desired use of the choice "I don't know." The intent was to push students to make only educated choices and to be willing to select "I don't know" when it was the most appropriate selection. Students were allowed to use a standard periodic table and a calculator.

Measuring in the Affective Domain

The Before and After Course Attitudes and Expectations/Experiences surveys collected student self-reported information about their expectations for the course (BCA&E) or their actual experiences in the course (ACA&E), their general background, reactions to (or value to them of) specific learning tools, reflections on how they know that they understand something, and, lastly, their confidence in their ability to carry out a course-related action (Wiggins, 1998).

Many of the statements used in the surveys were taken directly from, or based upon, those found at the SALG website (Wisconsin Center for Educational Research, 2010) and in papers published by the initial developers of FLAG (National Institute for Science Education, 2010) and SALG (Seymour, Daffinrud, Wiese, & Hunter, 2000). SALG originally focused on chemistry course affective domain student self-reported learning gains, but has developed over time to be a resource for use in other disciplines as well (Wisconsin Center for Educational Research, 2010). Additional statements about attitudes, expectations, and experiences were drafted based on these survey models and also on student written and oral responses when specifically requested to provide examples of what did, or did not, work for them in the course, as well as to provide suggestions for improvements which would make any aspects which were not helpful to them more helpful. Current and former students tested and critiqued the statements for interpretation as well as the time required to complete the whole set. Modifications were made recently to include more statements directly related to GEC-type learning objectives. Statements used a 1–5 Likert-type scale and always included the option "not applicable."

Each category (except "Background") is shown with a small sample of its associated statements in Table 3 (the statements are all from the After Course Survey version).

Table 3

Samples of Statements by Category from the ACA&E Survey

Experiences
- The quizzes and exams accurately measured what I understood in the course.
- It was clear how the lab experiments fit into this course.
- I had specific opportunities during this course to make connections between chemistry concepts and current world issues.

Confidence in my ability to...
- understand key concepts in chemistry
- understand the methods used in chemistry lab experiments
- apply my knowledge of chemistry to the real world

In this course, my learning was enhanced by...
- doing homework and practice outside of class sessions
- responding to questions asked in the flow of a lecture presentation
- working computer-based practice problem sets

I know I understand when...
- I can work problems from a textbook
- I get a good grade on a quiz or exam
- I can explain the ideas/concepts to someone else

Students completed the Content Check and Attitudes Survey assessment events in a variety of locations and settings, assuming they could work online and access the CMS quiz feature. "Before course" assessment events (BCC and BCA&E) were completed within the first week of classes in an academic term. "After course" assessment events (ACC and ACA&E) opened during the last week of classes and remained open for several days beyond the course final exam. Students were not required to participate, yet were encouraged to do so and were awarded two course bonus points (out of a course total of 1000 possible points) for participation.

Results

Content Check Data Analysis and Outcomes

Participation was voluntary; of the original enrollment total of approximately 275 students, about 90% completed the BCC, and of the 255 students receiving final course grades, about 76% completed the ACC. A total of 191 students elected to participate in both the BCC and ACC over five terms (75% response rate compared to all students taking the course final exam). These 191 students were divided between the traditional lecture sections and the hybrid computer-taught sections with 99 traditional students and 92 hybrid students participating. For both delivery styles, and for each of the five terms, the difference in means between the BCC and ACC was significant at $p < 0.0001$. This result supports that the course environment allowed students to make significant gains in core content knowledge for both delivery styles and allowed us to answer our first question: *Yes, the students demonstrated gains in chemistry course content knowledge.*

The next question we wanted to answer involved comparing the content knowledge gains for the two course delivery styles. Figure 1 shows the BCC and ACC percent overall scores for eight course sections (five hybrid and three traditional) and represents the 191 students who elected to complete both events.

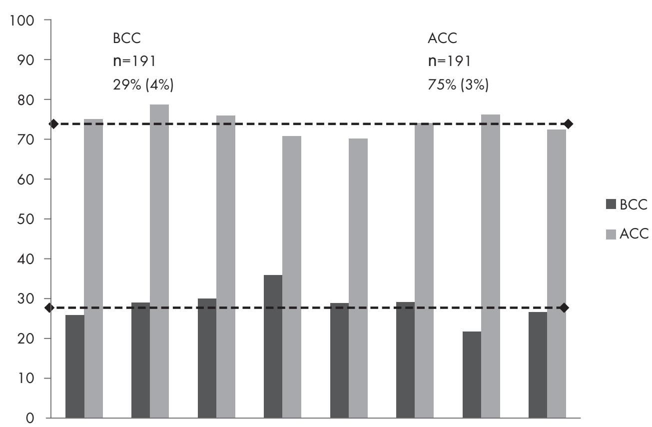


Figure 1. Overall percent correct score for eight baseline terms.

Note. The average over the time period is indicated using a dashed line.

In an effort to make collection and analysis the most straightforward, data were first included for all students who participated in a course event whether or not they also participated in another course event. So, any student who completed the BCC was included in that calculated average and sample count whether or not he or she also elected to complete the ACC. Removing students from the BCC data because they eventually left the course, or simply chose not to complete the ACC later in the term might have biased the BCC data toward successful students and would certainly not have represented the initial population of the class well. Participation in the ACC was available to all students who took the course final exam. Removing students from the ACC data because they had not chosen to complete the BCC at the start of the course might also have biased the data. Averages and other descriptive statistics for the BCC and ACC were first computed for the inclusive population which completed that event. This treatment seems most appropriate for reporting the descriptive statistics for any individual course event. However, it was also recognized that there could be value in treating the BCC and ACC data by including only those who completed both events. This data collection is less straightforward due to issues surrounding student identity, which generate data security concerns, and also due to the manner in which the data exports from the CMS. Also, with voluntary events, removing data from students who elect not to complete one or the other event may bias the data. For examination of gains or changes between course events, treatment of the data as pairs might be appropriate and was considered. Thus, a class could be described by at least two populations: (a) those who may or may not have taken the BCC and/or ACC and may or may not have remained in the class, and (b) those who remained in the course and took the final exam and completed both the BCC and ACC.

Each population has value and each may generate bias. Whenever possible, the data were examined in multiple ways. BCC and ACC results—when treated as inclusive of all students who participated versus only those students who elected to complete both events—were compared (see Table 4). There was no significant difference in the results for the two populations. So while Figure 1 shows the results for only those who completed both the BCC and ACC, the graph would be essentially identical if it showed inclusive data including all students who completed at least one of the events.

Table 4

Comparison of Content Check Results with Different Populations

	Inclusive BCC	**Only both CC BCC**	**Inclusive ACC**	**Only both CC ACC**
Average (SD)	29% (4%)	29% (4%)	75% (3%)	75% (3%)
Sample	*n* = 258	*n* = 191	*n* = 195	*n* = 191

When examined separately, the eliminated data for those students who did not complete the pair of events reflected a wide spread with both high-achieving students and low-achieving students represented. This pattern repeated term by term and implies that change or gain for the BCC and ACC can probably be well represented with the more straightforward inclusive data analysis.

Returning to Figure 1 and the data which underlie it, upon visual inspection and after comparing means and standard deviations, there appeared to be little difference between the content check results for the two styles. ANOVA results also showed no significant difference between the course styles for content check results (p values ranged between 0.082 and 0.333). Examination of Figure 1 shows that incoming groups on average for either course style can be expected to show about 29% content knowledge, and that outgoing groups will usually have made gains of about 45% (significant at $p < 0.0001$, $t = 34$, df = 380) ending up with about 75% content knowledge.

Unfortunately, a finding of "no significant difference" does not translate directly to a conclusion of similarity, so equivalence testing using confidence intervals was performed to support "equivalent for all practical purposes" (Clark, 2005). Equivalence testing was performed on all datasets, which through t-tests or ANOVA, showed "no significant difference." As an example of the process through which equivalence testing was completed and also how the decision to pool data was reached, one-way ANOVA ($p = 0.5259$) for the three sets of traditional-style ACC data indicated no significant difference in their mean values. Bartlett's test for equal variances also indicated no significant difference in their variances ($p = 0.7941$). And, Tukey's Multiple Comparison test also indicated no significant difference. Figure 2 shows equivalence testing for these three traditional sections' ACC overall score. The dashed lines indicate the selected zone of acceptable difference (or indifference); this limit was determined by considering what would be a tolerable margin within which small differences between means would lead to no discernable or practical difference in an outcome.

The differences of the means for all possible comparisons between the three sections are shown with their respective 90% confidence intervals. Since all three datasets fall entirely within the limits of the zone of acceptable (tolerable) difference, it

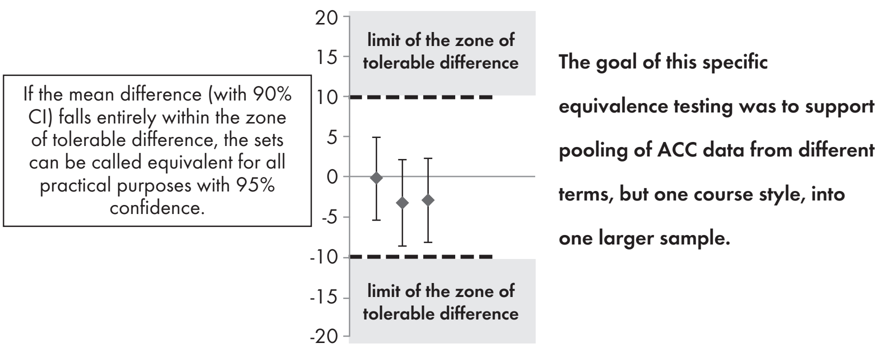
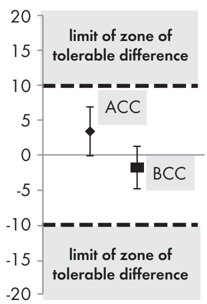


Figure 2. ACC mean differences (90% CI) for comparisons of three traditional course terms.

can be concluded that the three datasets are equivalent with 95% confidence (Berker, Luman, McCauley, & Chu, 2002). This specific equivalence test result supports pooling the different term ACC datasets into one larger data set for the ACC for the traditional sections. For all equivalence tests performed, the data were determined to be either acceptably equivalent or slightly ambiguous (no significant difference, and some, but not complete, support for equivalence). None of the compared data were completely outside the zone of acceptable difference (Rogers, Howard, & Vessey, 1993).

Figure 3 shows equivalence testing for the BCC and also for the ACC for the two course styles. The mean difference between the ACC for the pooled traditional sections from the ACC for the pooled hybrid sections and the mean difference between the BCC for the pooled traditional sections from the BCC for the pooled hybrid sections are shown. The dashed lines again indicate the selected zone of acceptable difference (or indifference).

If the mean difference with 90% CI falls entirely within the zone of tolerable difference, the ACC sets for the two course styles can be called equivalent for all practical purposes and hybrid can be pooled with traditional. The BCC sets for the two course styles can also be called equivalent to each other and can be pooled.



The goal of this equivalence testing was to show equivalence of ACC data between the two course styles. This also shows equivalence of BCC data between the two course styles. This does not show that the ACC and BCC were equivalent to each other.

Figure 3. ACC of pooled traditional mean difference (90% CI) from ACC of pooled hybrid and BCC of pooled traditional mean difference (90% CI) from BCC of pooled hybrid.

The goal for this instance of equivalence testing was to support that the traditional and hybrid sections had such equivalent outcomes that, based on the ANOVA results (no significant difference for each event) and equivalence testing (equivalent for practical purposes), we could answer our second question: *Yes, the students in the hybrid delivery style ended with approximately the same amount of content knowledge as those completing the course through a traditional lecture delivery style.* Figure 3 **does not show** that the ACC and BCC had equivalent results; equivalence of the BCC to the ACC was not something to expect, or test for, since ANOVA results indicated a significant difference between any BCC and ACC set ($p < 0.0001$). The data for the traditional and hybrid styles of Chemistry 101 were collected separately. However, because there was no significant difference for the overall percent correct score between the two course styles on either the BCC or ACC, and equivalence testing supported not only lack of significant difference but also practical equivalence, for the purpose of establishing baselines for future comparisons, these data have been pooled for each event rather than treated separately. The baselines for course content check scores were established by averaging the data gathered for five quarters, eight course groups, one instructor, one campus, and two course delivery styles. Weighted averages were used when they differed from standard averages for these datasets.

Another content check aspect which proved interesting to follow was the "I don't know" choice. Examination of Figure 4 shows that incoming groups can be expected to reply "I don't know" indicating no familiarity with the content knowledge at about 37% and that outgoing groups will show much greater familiarity with the course core knowledge by a drop in the "I don't know" response to about 4%. The sample *n* values differ because each represents all students who responded to the opportunity to complete that specific event. Analysis of the responses for the content check using only those 191 students who completed both the BCC and ACC produced essentially the same results.

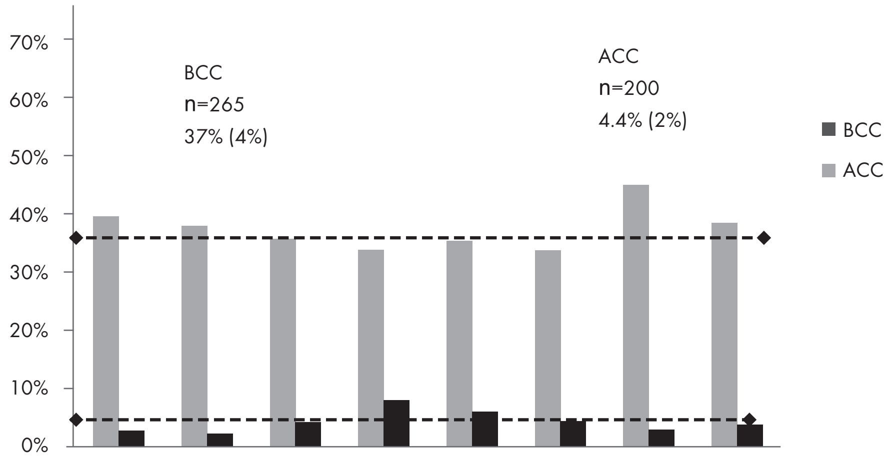


Figure 4. Overall "I Don't Know" percent score for eight baseline terms.

Note. The average over the time period is indicated using a dashed line.

As an alternate way to address whether the data obtained from the content checks was reasonable to combine over time, other data available from long-term tracking of exam scores and course total points were compared to the more recent five terms that were used to establish the baselines for the content checks. Table 5 presents the averages for four such measures. Student performance in the course on all these measures in the long term seems quite consistent with the five more recent terms. This suggests that the course environment has produced the same results for an extended period of time and that the five recent terms are probably quite representative of the course overall. This consistency of these measures over time establishes greater confidence in the new data from the five recent terms and the baselines established from them by combining data collected from the content checks.

Table 5

Course Scores over Time

Course Event	(long-term data) Average (SD)	(5 recent terms) Average (SD)
% Course Total	78.5 (3.9)	78.6 (4.2)
% Unit 1 Exam	82.8 (2.3)	82.6 (2.1)
% Unit 2 Exam	77.9 (3.2)	77.8 (3.0)
% Final Exam	75.1 (3.2)	76.3 (3.6)
Students	n = 480	n = 271

Correlation between course events provided another way to look at the new data. The correlation coefficients for percent score on the various course exams with the BCC and ACC are shown in Table 6.

Table 6

Correlation Between BCC, ACC, and Course Exams

	Final Exam	Unit 1 Exam	Unit 2 Exam	BCC
BCC	0.13	0.21	0.12	NA
ACC	0.65	0.67	0.53	0.17

The correlation between the Unit 1 Exam and the ACC (0.67) is somewhat stronger than that for the Unit 2 Exam (0.53) suggesting that, as intended, the ACC represents the basic, core, course content more closely than it does the more advanced course content. The correlation between the ACC and the Final Exam (0.65) is about as strong as that for the Unit 1 Exam. To some extent, this supports the use of the Content Check as a measure of course content knowledge, since the Final Exam for the course, if comprehensive, should certainly still include core course content knowledge. These correlation coefficients also suggest that there is no important relationship between a student's BCC knowledge level and successful course knowledge accumulation (correlations to all exams and to the ACC < 0.25), so use of the BCC to determine readiness for, or placement into, the course would not be supported by these data.

The next step following pooling the data from the content checks to generate reasonable baselines for future comparisons was to use the preliminary data for content knowledge to formulate criteria for evaluating the results of future course sections or styles. The first criteria considered were that

- choices of "I don't know" would decrease between the BCC and the ACC; and

- choices of "answered correctly" would increase between the BCC and the ACC.

These criteria were made more specific by using the pooled baseline data to

- establish an acceptable amount of change between the BCC and ACC for overall score (percent correct) and category scores; and
- establish an acceptable amount of change between the BCC and ACC for the "I don't know" response overall and category scores.

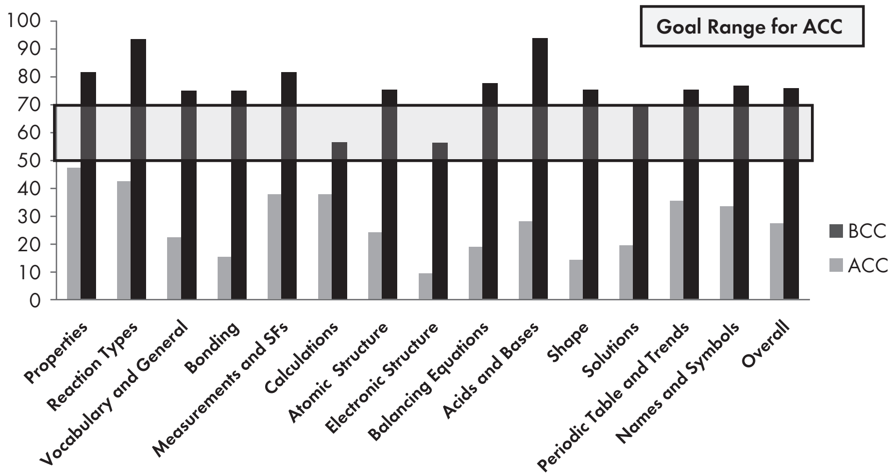


Figure 5. Percent correct score for BCC and ACC by content category with goal range.

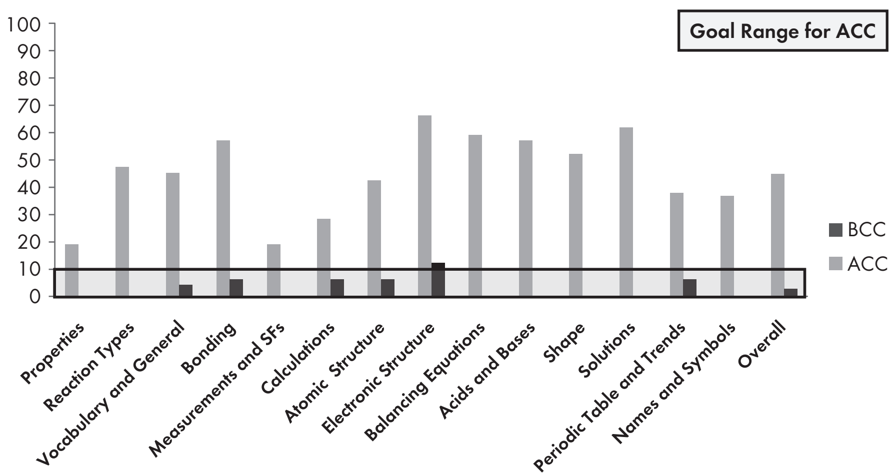


Figure 6. Percent "I don't know" for BCC and ACC by content category with goal range.

The decision was then up to us, as the instructors, to choose standards with which we were comfortable **and** which were supported by data. Figures 5 and 6 show the BCC and ACC results for a recent term drilled down to the content category for percent correct and for "I don't know" responses. Goal ranges for acceptable end scores (ACC) are shown as rectangles superimposed on each graph. These goals were chosen by us (as the course instructors) after examining the pooled baseline data for realistic results.

Another way we examined the results from a current course section was to use a spreadsheet developed to automatically show some criteria scenarios for the content category level. A portion of the spreadsheet output is shown in Table 7. Scenarios for improvement, end result, and one combination are shown in columns where the headings indicate the scenario being used. The specific course section sample used for Table 7 viewed in its entirety was evaluated as very successful, and the assessment goal of determining if the students seemed to **know more** about elementary chemistry content after the course than they did before was met in a data-driven manner.

Table 7

Criteria Scenario Spreadsheet Output Sample

	Improvement		**End Result**			**Combo**
	Improve > 25% minimum change	Improve > 40% high change	End > 50% satisfactory result	End > 60% good result	End > 70% high result	Begin < 50% End > 70% successful
Content Category						
Reaction Types	yes	yes	yes	yes	yes	yes
Measurements & SFs	yes	yes	yes	yes	yes	yes
Calculations	yes	No	yes	yes	No	No
Atomic Structure	yes	No	yes	No	No	No
Electronic Structure	yes	yes	yes	yes	No	No
Balancing Equations	yes	yes	yes	yes	yes	yes
Total (Overall Score)	good!	great!	satisfactory	good!	great!	successful

The criteria scenarios provide potential ways for us to view the data about content knowledge gains and to suggest points of entry to make changes if a specific aspect is not labeled as successful over time. These scenarios need to adapt to various people and groups and are not designed to be static or absolute. We found we could use such a system to identify categories of core knowledge which were not being learned as well as we would like; and, if the weakness remained when examined over time, this suggested where experiments with curricular alteration or delivery style change might be most influential. If, over time, our course samples tend to meet the selected scenario, then we can be more confident that our delivery style, content coverage, and course activities result in a course that helps students meet the content knowledge learning goal. If, over time, our course samples show patterns of content learning weakness in a few areas, then we may want to use that evidence to implement changes and then compare results following the change(s) to determine if the teaching changes have generated a desired learning improvement. If, over time, our course samples show patterns of content learning weakness in a lot of areas, then we may want to use the quantitative evidence to seek additional teaching improvement input from other instructors or instructional development specialists to test potentially effective changes.

This sort of data-driven scenario system has allowed us to use direct comparison to previously accumulated and analyzed data to evaluate specific changes in curriculum, supplemental resources, instructors, delivery, activities, and course style.

Attitudes and Experiences Survey Analysis and Outcomes

The data from the attitude surveys were used in analyses related to GEC-type learning objectives, but also provided broader information about student learning styles, student reflections on effective course activities, and other information which we hope will prove useful in continued course design. A total of 227 (84% response) students participated in the BCA&E and 120 (50% response) in the ACA&E over five quarters. Ninety-two traditional students and 135 hybrid students participated in the BCA&E, and 60 of each participated in the ACA&E. See Gutwill-Wise (2001) and Seymour, Daffinrud, Wiese, and Hunter (2000) for perspectives on the reliability and validity of student self-reported information.

For the Attitude and Experiences Survey, pairing BCA&E and ACA&E data was not possible because the survey was given anonymously during several terms and also because the survey data exports from the CMS without identity information. The data were analyzed to provide the average response and standard deviation on a scale of 1–5 with 5 being *strongly agree* or *positive* for most statements. Weighted averages were used when they differed from standard averages. A small sample of results for the BCA&E and ACA&E averages for the eight baseline course groups are show in Table 8. For the BCA&E, initial analyses showed no significant differences between the traditional and hybrid styles, so these data have been pooled to generate baseline attitudes and expectation levels.

The Before Course version of the survey was crucial to using the data from the After Course version. Running both versions of the survey allowed student attitudes and expectations at the beginning of the course to be compared to their attitudes and experiences after the course. Without a comparison, the After Course values would have been anchorless numbers.

Obviously, an ACA&E result which was significantly lower than the BCA&E result would indicate that the course was not meeting students' expectations for some aspect and might suggest a closer look be taken at any factors which could be influencing that aspect. When both surveys were directly compared, we set a minimum criterion of meeting the baseline attitude or expectation (from the BCA&E) as a successful result.

Table 8

Sample Results for BCA&E and ACA&E Averaged over Eight Class Groups (statements are worded from the BCA&E with the ACA&E wording in italics)

Expectations/Experiences	BCA&E Average (SD)	ACA&E Average (SD)
The structure of this course will enable *(enabled)* me to discover some of the ideas of chemistry for myself.	4.1 (0.2)	4.4 (0.2)
The quizzes and exams will accurately measure *(accurately measured)* what I understand in the course.	3.9 (0.2)	3.9 (0.3)
By the end of the course, I will feel *(felt)* able to apply the concepts presented.	3.9 (0.1)	4.0 (0.3)
In this course, my learning will be *(was)* enhanced by...		
viewing animations and video clips related to the concepts	3.9 (0.1)	3.9 (0.3)
reading about the concepts	4.0 (0.1)	3.8 (0.3)
giving explanations to others	3.3 (0.2)	3.1 (0.4)
being able to replay/repeat lecture material online	4.1 (0.2)	4.0 (0.5)
I understand when...		
I can apply ideas/concepts to new situations	4.7 (0.2)	4.8 (0.2)
I get a good grade on a quiz or exam	4.8 (0.1)	4.7 (0.3)

In an effort to avoid assessment overload, only the After Course Attitudes and Experiences Survey has been used for some current terms. Instead of meeting or exceeding the data-driven baseline levels as described above, an alternate criteria scenario was tested with minimum goals of 3.5 (a value firmly on the *positive* side of a 1–5 Likert range) set for all positive statements. Note that in many cases a

Table 9

BCA&E and ACE&A Results with Two Criteria Scenarios

Survey Statement	**BCA&E *n* = 227 Average (SD)**	**ACA&E *n* = 120 Average (SD)**	**Criteria 1 After ≥ Before**	**Criteria 2 After ≥ 3.5**
I will understand most of the ideas/concepts presented in this course.	3.8 (0.1)	4.0 (0.3)	met	met
Confidence in my ability to solve chemistry problems	3.0 (0.1)	3.6 (0.2)	exceeded	met?
Confidence in my ability to understand the chemistry of lab experiment	3.3 (0.1)	3.9 (0.2)	exceeded	met
I know I understand when I can connect concepts/ideas from one field to another	4.2 (0.3)	4.8 (0.1)	exceeded	met
Taking this course will help me understand newspaper and other articles I read about science.	3.5 (0.2)	3.5 (0.2)	met	met?
Confidence in my ability to apply my knowledge of chemistry to the real world	3.1 (0.1)	3.6 (0.1)	exceeded	met

value of 3.5 (see Table 8) **would be less than** the student expectations for the course (as identified through BCA&E), so this more arbitrary criteria scenario may not stand up to re-appraisal. Table 9 shows some BCA&E and ACA&E averages and their designation after applying both the data-based and more arbitrary criteria scenarios. For these samples the ACA&E data met or exceeded both criteria we set for "success."

The affective domain analyses to date have mainly focused on how the data indicated the course was, or was not, meeting noncontent knowledge mandated GEC learning objectives. Table 10 shows a connection between the four GEC learning objectives associated with this course with a few statements from the surveys.

Table 10

Survey Data and Desired Outcome Criteria Connected to GEC Learning Objectives

Meeting GEC Learning Objectives?	**BCA&E Average (SD)**	**ACA&E Average (SD)**	**Desired outcome? (compare to a goal ≥ Before survey)**	**Desired outcome? (compare to a goal of ≥ 3.5)**
GEC LO#1 Students understand the basic facts, principles, theories, and methods of modern science.				
This course improved my understanding of what chemists do.	4.1 (0.1)	4.1 (0.2)	met	exceeded
GEC LO#2 Students learn key events in the history of science.				
Confidence in my ability to understand key concepts in chemistry	3.2 (0.1)	3.8 (0.2)	exceeded	exceeded
Combined GEC LO #3 & GEC LO#4 Students provide examples of the interdependence of scientific and technological developments. Students discuss social and philosophical implications of scientific discoveries and understand the potential of science and technology to address problems of the contemporary world.				
This course encouraged me to apply chemistry ideas to everyday situations.	3.6 (0.3)	4.0 (0.3)	met	exceeded

Survey data such as that shown in Table 10 can be combined for all statements mapped to one GEC learning objective and then used to support that the course is, or is not, meeting the GEC goals set for that GEC category. Table 11 summarizes the results from a recent term for the combination of all statements mapped to each GEC learning objective and, if a goal value of greater than 3.5 is used, the data support that this course met the GEC goals successfully.

Discussion

Initially, this assessment project sought only to provide an outcome report that confirmed content knowledge gains were equivalent between the traditional and hybrid delivery styles. The data from the BCC and ACC have been used to confirm numerically

Table 11

ACA&E Statements Mapped to GEC Objectives Summary

GEC Objective	Average (SD)
Students understand the basic facts, principles, theories, and methods of modern science. (GEC LO #1)	3.9 (0.2)
Students learn key events in the history of science. (GEC LO #2)	3.7 (0.1)
Students provide examples of the interdependence of scientific and technological developments. Students discuss social and philosophical implications of scientific discoveries and understand the potential of science and technology to address problems of the contemporary world. (Combined GEC LO#3 &GEC LO#4)	3.8 (0.2)

that students do show equivalent gains in their overall core content knowledge as a result of completing the course through either the traditional or hybrid delivery style. Additional analysis drilled down to the content category level is in progress.

This assessment process also led us to the realization that we needed to measure more than content knowledge. Using student self-reported information, we have started to explore the affective domain of learning about chemistry and science (Seymour, 2002). Combinations of the attitudes and experiences survey data have allowed us to begin to address the question of how the course was meeting general education learning goals; this same process could let us explore how the students' attitudes about the nature of science in general are being changed by the course (Clough & Olson, 2008; Lederman, 1998; McComas, Clough, & Almazroa, 1998). The data are noisy enough that comparison over time is essential. The establishment of baselines through use of before course assessment events (content check and affective survey) over time provided us with comparison levels that take into consideration there will always be some term-to-term and section-to-section variability. Since these courses all had the same experienced instructor, this amount of variability must be attributed to variations in student groups, since it cannot be attributed to differences between instructors. It is important to be aware that this variability is present. Examining only one term would not necessarily be a good reflection of that course or that instructor, but patterns over time compared to baseline data provided useful information to make, or evaluate, course and delivery changes and student learning. These baselines allowed for data-driven criteria to be chosen for investigating our teaching and our students' learning.

As individual instructors, we decided what standards to use anchored by data-driven baseline possibilities. This body of baseline data will make future investigations of the effect of changes more straightforward by providing a point of comparison readily available to evaluate the effectiveness of that change.

The chosen assessment instruments shifted away from questions or statements that asked about the instructor directly, or even about "liking" any aspect of the course in favor of those which asked directly for content knowledge or for student self-reported ratings of aspects that enhanced their learning, confidence, or course experience. In this way, we have gained information about the alterable aspects of the course rather than about our personal characteristics (Tagamori & Bishop, 1995; Trout, 1997; Williams & Ceci, 1997).

Participating in ongoing assessment has influenced our teaching in both major and minor ways. As an example of the latter, an animated wizard had been used to narrate some online course material while instructor voice-over narrated the rest. Students who did not like the wizard were quite vocal about their dislike. The wizard was nearly removed entirely in a knee-jerk response to only the verbal input, but attitudinal statement results from the end of the course indicated that the students who did not like the wizard were actually in the minority and that the less vocal majority felt that the wizard and the instructor's voice-over were comparable with only a very, very slight preference for the instructor's voice. In this case, the data allowed an informed decision about the animated wizard's use to be made, and the wizard was saved. On a larger scale, we were able to draw a quantitatively backed conclusion that the hybrid delivery style was as effective as the traditional lecture style for student core content knowledge gains.

The sample data presented show only the beginnings of the possibilities for analysis. As we seek to verify the effectiveness of our course curriculum, delivery, and student experiences, or investigate the effects of a deliberate change made in any of these, analysis of student self-reported survey data gathered for categories of "Learning Enhancement" and "Confidence" could be quite interesting. For example, Table 12 shows the results for the seven measures of confidence that revealed significant gains between the BCA&E and the ACA&E. Although it is certainly reassuring that these data confirm the effectiveness of the course at maintaining or increasing student-reported confidence levels for these seven specific aspects, one might wonder why the other five confidence aspects were maintained more closely at the level the students expected upon entering the class and did not increase. Our data analysis seems to answer one question and create new ones as we proceed.

Table 12

Seven Measures of Confidence Increased

Confidence in my ability to...	**BCA&E Average (SD)**	**ACA&E Average (SD)**
understand key concepts in chemistry	3.2 (0.1)	3.8 (0.2)
solve chemistry problems	3.0 (0.1)	3.6 (0.2)
understand the chemistry of lab experiments	3.3 (0.2)	3.9 (0.2)
use computer- based materials to learn chemistry concepts	3.4 (0.2)	3.9 (0.2)
visualize key concepts of chemistry	3.3 (0.1)	3.8 (0.2)
apply my knowledge of chemistry to the real world	3.1 (0.1)	3.6 (0.2)
succeed in another chemistry course	3.3 (0.1)	3.7 (0.2)

In the future, we will attempt to answer questions related to other aspects of the different course delivery styles by continuing to collect data as well as analyzing further our current datasets. For example, we are curious to discover whether the delivery style influences student confidence in their course-related skills, or whether delivery style impacts how students attempt to connect the course material to the larger world. Additional analyses of the datasets to identify differences, if any, between the outcomes for the two course styles are ongoing and fascinating.

This project has allowed us to explore the assessment process and its value to our own teaching of chemistry and nonmajor students' learning about chemistry and the nature of science. One of the pleasantly surprising things we discovered is how easy it has been to use the same basic assessment instruments to explore increasingly complex questions. With minor modifications, we have been able to continue using the same instruments to frame new questions. What began as a project to obtain data to evaluate a new course delivery style has provided a pathway for us to explore not only course "quality" but also ways that students learn about chemistry and how nonscience majors become attuned to the nature of chemistry and science. We also have learned to trust the data collected as a valuable addition to our gut instinct regarding how things are going.

Perhaps the most important lesson we have learned from this process is that a few simple, easy-to-use instruments have given us ways to understand whether we are successfully providing courses that allow our students to meet the cognitive and affective learning goals we set. The time spent aggregating the data and analyzing the results is well worth the information it provides. The more we learn about how our students are learning chemistry, the more we want to know. As a result, we are certain our students will continue to benefit in terms of their learning as we continue developing as engaged researchers of our own teaching.

REFERENCES

Berker, L. E., Luman, E. T., McCauley, M. M., & Chu, S. Y. (2002). Assessing equivalence: An alternative to the use of difference tests for measuring disparities in vaccination coverage. *American Journal of Epidemiology, 156*(11), 1056–1061.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy, and Practice, 5*(1), *7–74*.

Clark, M. (2005, February). Equivalence tests. *RSS matters*. Retrieved from BenchmarksOnline http://www.unt.edu/benchmarks/archives/2005/february05/rss.htm

Clough, M. P., & Olson, J. K. (2008). Teaching and assessing the nature of science: An introduction. *Science and Education, 17*(2-3), 143–145.

Gutwill-Wise, J. (2001). The impact of active and context-based learning in introductory chemistry courses: An early evaluation of the modular approach. *Journal of Chemical Education, 78*(5), 684–690.

Lederman, N. G. (1998, December). The state of science education: Subject matter without context. *Electronic Journal of of Science Education, 3*(2). Retrieved from http://wolfweb.unr.edu/homepage/jcannon/ejse/ejsev3n2.html

McComas, W. F., Clough, M. P., & Almazroa, H. (1998). The role and character of the nature of science in science education: An introduction. *Science and Education*, *7*(6), 511–532.

National Institute for Science Education, University of Wisconsin-Madison. (2010). *Field-tested learning assessment guide (FLAG).* Retrieved from http://www.flaguide.org

Ohio State University Office of Academic Affairs. (2010). *Arts and sciences curriculum and assessment highlighted resources.* Retrieved from General Education Curricula http://asccas.osu.edu/files/asccas/GEC%20ELO%27s%20from%20Manual%202.26.10_0.pdf

Ohio State University Office of the Registrar. (2010). *Course catalog search for CHEM 101*. Retrieved from http://www.ureg.ohio-state.edu/courses/

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin, 113*(3), 553–565.

Seymour, E. (2002). Tracking the processes of change in U.S. undergraduate education in science, mathematics, engineering, and technology. *Science Education, 86*(1), *79–*105.

Seymour, E., Daffinrud, S., Wiese, D., & Hunter, A. (2000, March). *Creating a better mousetrap: On-line student assessment of their learning gains. Report to Exxon-Mobil Foundation.* Paper presented at the National Meetings of the American Chemical Society Symposium, San Francisco.

Suskie, L. (2009). *Assessing student learning.* San Francisco: Jossey-Bass.

Tagamori, H., & Bishop, L. (1995). Student evaluation of teaching: Flaws in the instrument. *Thought and Action: The HEA Higher Education Journal, 11*(1), 63–78.

Trout, P. (1997). What the numbers mean: Providing a context for numerical student evaluation of courses. *Change, 29*(5), 24–30.

Walvoord, B. E. (2004). *Assessment clear and simple.* San Francisco: Jossey-Bass.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* San Francisco: Jossey-Bass.

Williams, W., & Ceci, S. (1997). How am I doing? Problems with student ratings of instructors and courses. *Change, 29*(5), 12–23.

Wisconsin Center for Educational Research, University of Wisconsin-Madison. (2010). *Student assessment of their learning gains (SALG).* Retrieved from http://www.salgsite.org

CHAPTER 5

ASSESSMENT IN UNDERGRADUATE CHEMISTRY RESEARCH: ACCOMPLISHMENTS AT HAROLD WASHINGTON COLLEGE

Jeffrey S. Carver, Morna Brothers, and Thomas B. Higgins
Harold Washington College

Abstract

The following chapter describes the use of multiple assessment instruments utilized during the STEM-Engaging the Next Generation IN Exploring STEM Undergraduate Research Collaborative (STEM-ENGINES URC; National Science Foundation CHE-0629174). The nature of this project is unique in that it is the only NSF-funded Undergraduate Research Collaborative (URC) that focuses on the experiences of students at two-year community colleges (2YCs). Students participated in authentic undergraduate research experiences at the 2YC during the academic year, with half the students continuing to do research at a four-year college or university (4YCU) during the summer. Issues regarding 2YC implementation are discussed, including utilizing assessment instruments that focus on student learning of chemistry content and process skills, emotional intelligence, and potential pursuit of STEM careers. The use of these instruments, which were designed primarily for students at the 4YCU, allows us to analyze the 2YC experience with reference to other research experiences external to the program. Preliminary project results are shared to indicate the impact the experiences have had on the students involved thus far.

Introduction

Two-year colleges serve a large, diverse student population. Currently, of the 11.5 million students enrolled at 2YCs, the average age is 29, 60% are female, and 35% are minorities (American Association of Community Colleges, 2009). Recognizing the need for the United States to increase the number of students, particularly from underrepresented groups who pursue STEM degrees and careers, the STEM-ENGINES URC is dedicated to engaging 2YC students in authentic scientific research. This project, now in the fourth year of a five-year NSF grant (CHE-0629174) to explore the role of undergraduate research in engaging and retaining 2YC students in the sciences, grew out of the shared vision of 2YC chemistry professors at separate institutions devoted to authentic research as a method of teaching. The STEM-ENGINES, composed of participants from 10 Chicago metropolitan area 2YCs and three Midwestern 4YCUs, not only has the potential to increase the number of students who pursue STEM degrees and careers, it also offers a significant opportunity to explore existing assessment strategies in novel ways. This chapter provides a description of an assessment project in progress and reports on preliminary findings concerning the role of undergraduate research in enhancing student learning of chemistry content and process skills, emotional intelligence, and influencing student decisions to pursue STEM careers. Further reporting of finalized datasets will occur in future manuscripts as we continue to investigate the role of undergraduate research in engaging and retaining community college students in the sciences, specifically, chemistry. We have begun to collect empirical evidence to support the abundant anecdotal evidence that indicates research experiences are crucial to engaging and retaining students, and in particular 2YC students, in the sciences. As we discuss below, the 2YC student experience often differs from a 4YCU student experience, especially with respect to greater student mobility and perhaps also in terms of interest in participating in scientific research; therefore, the program evaluation must take these factors into account.

One of five URC projects sponsored by NSF, STEM-ENGINES is the only URC to focus solely on 2YC students. Begun as a project devoted to students in chemistry, STEM-ENGINES has grown to include students from other disciplines, including geology and biology. However, the assessment practices and some of the instruments used in this project were developed specifically for chemistry students. The STEM-ENGINES URC project provides a significant opportunity to examine the benefits that may accompany student participation in research at the community college level.

This project draws from the talented pool of 2YC students at the partner schools, engages them in scientific research, and supports their transfer to baccalaureate-granting institutions. During the first three years of implementation, 168 students have been STEM-ENGINES fellows, taking part in year-long research experiences, at a 2YC during the academic year and at a partner 4YCU during a 10-week summer internship. Of the 168 students, 93% have completed the academic-year program, 50% have participated in summer research experiences at partner institutions, and 52% have transferred. Although this information is being used as one indicator of success, this chapter describes a far-reaching evaluation program that considers the possible benefits of student participation in undergraduate research in several different ways, including the Survey of Undergraduate Research Experience (SURE; Lopatto, 2004), a content assessment instrument modeled after an American Chemical Society measure of content knowledge, the Bar-On Emotional Quotient Instrument (Bar-On EQ-i; Bar-On, 2006), and a modified version of the Kardash survey (Kardash, 2000). Before going on to a more detailed description of assessment methods and a discussion of the assessment efficacy, it is beneficial to give a narrative description of the STEM-ENGINES project.

Program Description

The STEM-ENGINES URC is based on the premise that authentic research experiences will engage 2YC students and encourage them to persist in STEM education and careers. Multiple studies (Bauer & Bennett, 2003; Hathaway, Nagda, Gregerman, 2002; Nagda, Gregerman, Jonides, von Hippel, & Lerner, 1998; Russell, Hancock, & McCullough, 2007; Seymour, Hunter, Laursen & DeAntoni, 2004) support the notion that a higher percentage of students native to the baccalaureate institutions who participate in undergraduate research may persist in STEM education and careers relative to baccalaureate-native students who do not participate in research. However, investigations of 2YC student research participation and its effect on student learning of chemistry content and process skills, potential pursuit of STEM careers, and emotional intelligence are limited. This study project seeks to fill a void in the current knowledge base by documenting effective and established assessment practices that speak to the unique situations of the 2YC, which in the case of the STEM-ENGINES URC program center around a multi-institutional partnership with year-round research opportunities for a diverse population of early undergraduate students.

Given the importance of 2YCs in higher education and the call for more authentic learning experiences earlier in the curriculum, this project lies at the intersection of two important themes in science education: (a) it informs teaching and learning at all levels, and (b) it holds promise for reaching a population of students traditionally underrepresented in STEM professions. The primary goals of the project revolve around providing 2YC students with academic experiences similar to those found at baccalaureate institutions and include (a) identifying and recruiting promising young scientists from 2YCs into the STEM disciplines; especially from traditionally underrepresented groups; (b) training 2YC students to become effective practitioners of science; (c) instilling in 2YC students the confidence to pursue science as a profession; (d) encouraging 2YC students to complete their undergraduate and graduate STEM education; (e) enriching academic preparation by enhancing student knowledge in STEM areas; and (f) transforming the cultures of participating 2YCs by embedding intensive research experiences during the academic year and summer into their curricula and their courses. This chapter focuses on the evaluation methods utilized for the first five goals. Recruitment of students is measured through a demographic survey, which new students are required to complete upon entrance and returning students are required to complete every fall semester. Effective training of future scientists is measured by tallying the number of student products each year, including formal presentations, co-authored publications, and patent applications. Participants are also required to complete a content exam every semester. The impacts of the research experience in terms of career choice and education plans are measured through the Kardash survey, the SURE, and a tracking survey.

Located in the Chicago metropolitan area, the collaborative consists of 13 partner institutions. Two-year colleges include the seven City Colleges of Chicago (CCC), College of DuPage, William Rainey Harper College, and Oakton Community College. Hope College (Holland, MI), Illinois State University (Normal, IL), and Youngstown State University (Youngstown, OH) comprise the original baccalaureate-granting partner institutions. Students who have completed at least a semester of research at the 2YC are eligible to apply for a 10-week, paid summer research experience at one of the three original partner baccalaureate institutions or any one of the other formal or informal educational partner institutions that have joined the project, including the Brookfield Zoo, Chicago Botanic Garden, Chicago State University, DePaul University, Dominican University, and University of Illinois at Chicago.

Participating institutions currently employ two models of practice for students engaging in research during the academic year, with a third proposed model that has not yet been realized (Brothers & Higgins, 2008). Students who participate in only one type of research will likely preferentially engage in that model of research in their future STEM careers. The advantage to the design of the STEM-ENGINES URC is that students have the opportunity to participate in several styles of research practice at multiple institutions throughout their participation in the program by taking advantage of both academic year and summer research experiences. The first model is a traditional mode of undergraduate research where the student is mentored by an individual faculty member. The second model of instructional practice for undergraduate research involves the student engaging in an elective course in which an interdisciplinary team of faculty teaching the course engages the students specifically in instruction revolving around conducting research, presenting results, and technical writing skills associated with dissemination of the research. The third model of instructional research practice has not yet been fully realized by the collaborative but involves cross-group and cross-college collaborations of both physical and human resources.

Through these models, faculty mentors seek to engage a diverse, "non-traditional" population that tends to differ from that of a 4YCU. The community colleges participating in the STEM-ENGINES URC collectively serve 106,000 undergraduates annually, 42% of whom come from underrepresented groups. During the first three years of the project, STEM-ENGINES has engaged a heterogeneous population underrepresented in STEM. Ranging in age from 17 to 63 years old, the 168 student participants have diverse demographics with 54% being female, 19% Asian, 25% Black, 14% Hispanic, 41% White, and 1% Native American. This student population includes first-time college students, reverse transfers, and career changers, and reflects varied histories and cultures.

Student diversity also extends to the level of science education. STEM-ENGINES seeks to engage first- and second-year students who may not have considered a career in science, and the only prerequisite for participation in the project is an interest in science. Mentors at the 2YCs recruit students they believe have the potential to be successful. By recruiting them early and introducing them to the fundamentals of research, the program aims to promote the development of students' confidence as scientists and to encourage them to pursue their STEM education beyond the 2YC. In many regards, recruitment efforts focus heavily on finding the "diamonds in the rough" at any stage in their 2YC experience. As such, students range considerably in their scientific preparation. Some students may have taken a sequence of chemistry courses in order to receive an associate of science degree. Others may be taking their first chemistry class ever. Unlike many traditional research experiences at a 4YCU, which typically engage only juniors and seniors (Brown, 2006; Russell et al., 2007), the STEM-ENGINES program aims to immerse the students in the research experience regardless of their previous educational experience.

This immersion is possible through a series of programmatic elements that span the different research models and institutions: project development, high standards, and community support. Students are engaged from the beginning as they work with the faculty mentors to identify interesting and novel research questions. Mentors agree that students will not be engaged by investigating questions that have already been solved. As such, projects focus on current and relevant topics such as fuel cell research, water quality issues, nanoparticle toxicity, and analysis of non-FDA-regulated supplements.

As primary participants, students are encouraged to develop experimental designs to research their questions over the course of the academic year.

Students are held to rigorous standards, requiring five to ten hours of research per week while taking a full course load. Participating students receive course credit for their independent research in addition to receiving a small stipend of around $100 per week as a student research assistant. Students, working with mentors from 2YCs and 4YCUs, are required to present their research to the STEM-ENGINES community and are encouraged to co-author publications and presentations in national forums. Some of our students are depicted in Figures 1–6 below.[1]

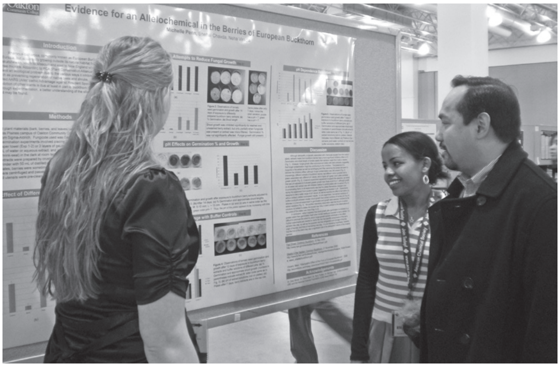


Figure 1. STEM-ENGINES students presenting their research at the American Chemical Society meeting. Photo taken by Morna Brothers.

Figure 2. Students at work in the Truman College laboratory. Photo taken by Robert Shretzman.

Figure 3. STEM-ENGINES summer research fellows at Illinois State University. Photo taken by Linda Ferrence.

Figure 4. Summer fellow running a column. Photo taken by Donieka Burris.

[1]STEM-ENGINES students signed an agreement allowing for the use of their photographs.

Figure 5. URC student in the lab during summer research experience at Hope College. Photo taken by Melissa McMahon.

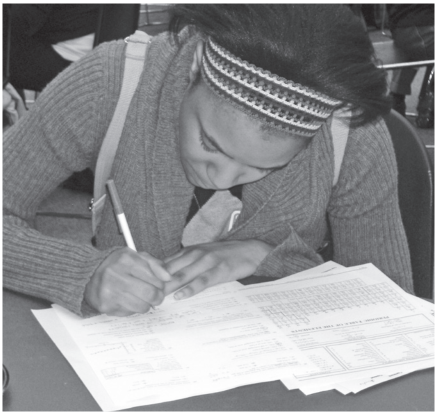


Figure 6. Student completing an assessment exercise. Photo taken by Morna Brothers.

At the campus level, community support is promoted by the use of peer mentors, returning students who work with new students and help them acclimate to the research culture. In addition to helping with proper research techniques such as keeping a lab notebook and using instrumentation, peer mentors alert newer students to the benefits of belonging to STEM-ENGINES, both informally through conversation and formally through oral presentations. All returning students are expected to present their experiences at summer research sites and at national meetings to the entire URC community. According to the meeting evaluations, these presentations are inspiring to the other students and help them realize some of the possibilities open to them through their research.

STEM-ENGINES includes multiple partners, linking 2YCs and 4YCUs. Connections across institutions are sustained by community-building activities including workshops (such as instrumentation training and presentation practice), field trips, and our tri-annual research symposiums. At the symposiums, hosted at least once a semester, faculty and students from all the STEM-ENGINES partners convene. Representatives (students and mentors) present their research, mingle, and discuss science. Participants get to know each other and each other's research as students are required to present their projects at least once a semester. Repeated presentation experiences are meant to instill confidence in the students' knowledge of subject matter. Additional benefits of regular meetings include increased access to instrumentation, opportunity for cross-group and cross-college collaboration, and a forum for student research assessment activities. Students also have an opportunity to meet and hear from working scientists, who are invited to discuss their work with the students. It is important for first- and second-year students to hear about the range of jobs available for professional scientists as well as potential opportunities for funding advanced education.

This community support extends through the summer research experiences, where students devote themselves to research full-time. The STEM-ENGINES community helps provide transitional support for those students who are eligible to apply for a 10-week paid summer research experience at our partner 4YCUs and informal education organizations. Some students have never left the Chicago metropolitan area before and become more willing to travel to a campus where they have an established connection. Each year, the program supports over 30 summer students. Participating students have the opportunity to become familiar with a different campus, meet and network with the institution's faculty and students, have access to equipment unavailable at the 2YC, and become part of a research team. Beyond the benefits of the research experience, students have the chance to identify themselves as students at a particular 4YCU, and mentors encourage students to consider transferring to partner institutions where they have already established connections. Not all of the students choose to participate in the 4YCU experience for a variety of reasons, including being place-bound by familial responsibilities. However, for those who do not choose to participate in a 4YCU experience, summer opportunities are available at other local informal education organizations as well as at their home 2YC institutions.

The assessment practices were designed to meet these challenges in a number of ways. In addition to recording the number of participant presentations, publications, degrees, and transfers, the evaluation team distributes assessments that address emotional intelligence, chemistry content, and potential of fellows to continue pursuit of a STEM career. Measures also aim to acquire participant feedback from students as well as faculty mentors by distributing surveys after the research experiences.

Assessment Methods

Literature Review

With an understanding of the types of activities being conducted as part of the STEM-ENGINES URC described, we can turn our attention to the focus of this chapter—assessment of 2YC student learning of chemistry content and process skills, emotional intelligence and propensity toward STEM careers, and the value related to those students' research experiences. While 2YC education has educational strengths in comparison with traditional 4YCU experiences, the opportunity for 2YC students to engage in research has been limited. Similarly, investigations into the role of undergraduate participation in research at community colleges have been limited. This is not to say that research at 2YCs is not being conducted. Brown (2006) describes the future "Winds of Change" in the prominence of 2YC research activities stemming from the large number of Ph.D.-educated faculty being available for hire at 2YCs because of the highly competitive nature of tenure-track faculty positions at 4YCUs. However, while other NSF-funded projects have focused on 2YC participation in research activities (Gaglione, 2005), the predominance of the current literature available focuses more on the development of 2YC faculty scholarship (Palmer & Vaughn, 1992) and less on student participation and subsequent potential gains.

Most recently, studies have examined student retention and success rates in science after participation in novel research activities as part of their normal laboratory instruction (Marcus, Hughes, McElroy, & Wyatt, 2010) or as part of an early college introduction-to-research type of class (Behar-Horenstein & Johnson, 2010). However, these studies focus on early college students at 4YCUs and not those at 2YCs. Thompson (2001) reports that 2YC students are more likely to put more effort forth in their science and math classes if they participate in informal student-faculty interactions and, subsequently, the more likely they are to report gains in science and math performance. However, the study did not specifically list research activities as an informal student-faculty interaction nor did it report on gains measured by any assessment instrument. Rather, the gains reported were student perceived gains in science and mathematics courses.

The culture of the 2YC is such that faculty are chosen predominantly for their teaching ability, and research is rarely, if ever, a part of a faculty member's responsibility, even in the sciences where research has dominated the growth of knowledge of the disciplines. In fact, Smolkin (2003) describes 2YCs as "Hidden Bastions of Chemical Education." While this may be the case from a teaching perspective, the fact remains that, in general, 2YCs are not providing the kind of scientific research experiences that support a first- or second-year undergraduate's development and interest in a STEM field. The STEM-ENGINES URC was designed to provide a research basis for inclusion of undergraduate research experiences at the 2YC as well as provide several models for implementation that can be disseminated and replicated at other 2YC campuses. The project evaluation makes an important contribution to the limited research documenting the impact of undergraduate research on 2YC students' learning gains in chemistry content and process skills, emotional intelligence and intent to pursue STEM careers (Bleicher, 1996; Brown, Bolton, Chadwell, & Melear, 2002; Etkina, Matilsky, & Lawrence, 2003; Richmond & Kurth, 1999; Ritchie & Rigano, 1996; Ryder, Leach, & Driver, 1999; Schwartz, Lederman, & Crawford, 2000; Seymour et al., 2004; Westlund, Schwartz, Lederman, & Koke, 2001).

Assessment Practices of the STEM-ENGINES URC

In an effort to measure the effectiveness of our program, goals have been listed below with corresponding research questions and subsequent assessment instrument or activity used to measure those goals related to undergraduate research in chemistry at the 2YC. With the focus on the outcomes of the goals and the potential to answer the defined research questions, the evaluation design of the STEM-ENGINES URC project was developed both to assess student learning in chemistry content and process skills and to examine the benefit of research as an integral part of the scientific culture of the 2YC system. If students at the 2YC have never had a first-hand opportunity to work with a scientist who engages in the practices of science through authentic research, they will not have the ability to model scientific behavior as they continue their STEM education at baccalaureate institutions. In the following sections, we approach the various assessments used as they specifically relate to the achievement of the program goals described above: identifying and recruiting promising students, developing 2YC students' skills in scientific practice, instilling confidence to pursue STEM fields, encouraging degree completion, and enriching academic preparation. Although the program includes additional goals related to the culture of conducting research at a community college, discussion of those will be omitted here. Figure 7 shows the various pathways in which a student can participate in the STEM-ENGINES project and when the various assessments are administered.

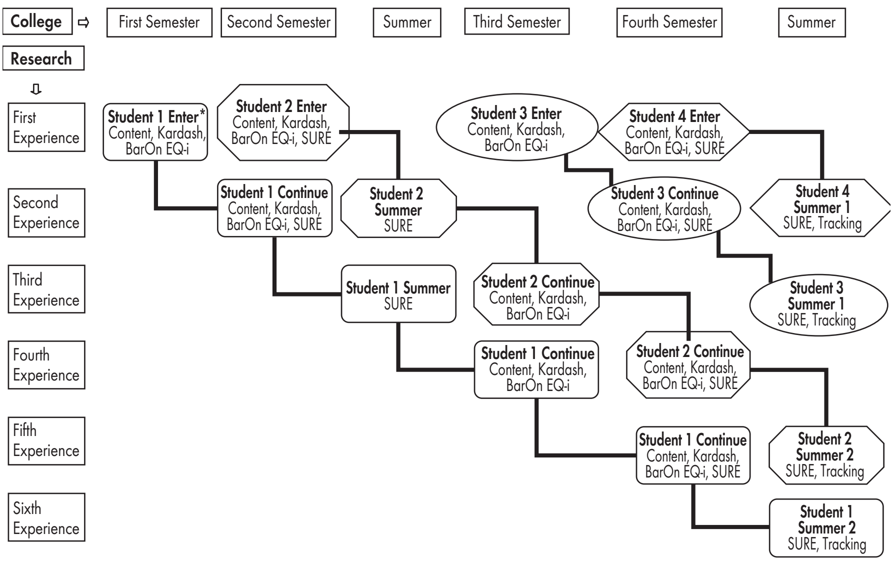


Figure 7. Multiple pathways through STEM-ENGINES URC including student entry point and the timing of the assessments. *Student 1 enters the URC as a first semester freshmen, Student 2 enters as a second semester freshmen, Student 3 enters as a first semester sophomore, and Student 4 enters as a second semester sophomore.

Goal 1—Identify and recruit promising students.

The research questions related to Goal 1 are:

- Is the emotional intelligence of a student a useful predictor of success in the STEM-ENGINES program, and can the results of this assessment be considered when identifying potential participants?
- Do those students engaged in research demonstrate gains in emotional intelligence?

Bar-On Emotional Quotient Inventory (EQ-i)®.[2] Given the diverse demographics of community college students, it is useful to collect baseline data

[2]The Emotional Quotient Inventory (EQ-i) is a registered trademark of Multi-Health Systems Inc.

describing the emotional and social intelligence of the students. Bar-On (2006) describes his model of social and emotional intelligence as "a cross-section of interrelated emotional and social competencies, skills and facilitators that determine how effectively we understand and express ourselves, understand others and relate with them, and cope with daily demands" (p. 15). The Bar-On EQ-i, available commercially from Multi-Health Systems Inc., provides a determination of the level of the emotional maturity the student holds and the ability of that individual to engage in intellectual and social interactions through being aware of others as well as being aware of self. Its use here is significant because it does not measure either of the two domains of science—science content and science process skills. The EQ-i instrument's reliability and validity has been established in previous studies involving university students (Dawada & Hart, 2000). The intent for the use of this particular instrument as an assessment in this project is two-fold. First, it is used to examine the baseline emotional maturity of the students entering the program. Second, it examines whether students exhibit a change in emotional intelligence after participating in a research experience. Also, if it is determined that emotional intelligence is a reasonable predictor of success in a research experience, then a gain in emotional intelligence may indicate improved preparation for subsequent research experiences at both the 2YC and 4YCUs. While the second use of this instrument is not directly related to Goal 1—Recruitment, both uses of the Bar-On EQ-i instrument will be discussed here for ease of organization.

An important theme of the NSF URC approach is that research experiences should be extended to include many students who traditionally do not participate in research, be they students in introductory chemistry courses at research universities or students at 2YC sites. However, some students may still be ill-suited for such a research experience. While the original use of the EQ-i in the STEM-ENGINES URC was not intended to eliminate individuals from participating in certain activities, it has the potential to serve that function. Individuals whose scores fall in the Area for Enrichment realms of the EQ-i are viewed as having less of an ability to socially interact with other individuals. Since participation within a research group requires abilities of social interaction for the purpose of advancing the knowledge base, the Bar-On EQ-i could be used as a screening process for participants potentially establishing the EQ-i instrument as a useful predictor of student success in the STEM-ENGINES program.

The instrument is administered each semester as the students progress through the program. The outcome of the instrument establishes emotional maturity in three realms: Areas for Enrichment, Effective Functioning, and Enhanced Functioning. Implementation of this particular instrument was intended to study the extent to which participants engaged in undergraduate research experience a gain or show growth in their emotional maturity as a result of participation. Once the baseline emotional intelligence for each student was determined, a quasi-experimental design was followed to determine growth of emotional intelligence within the groups, though a pretest/posttest analysis may also prove useful to determine growth of an individual's emotional intelligence.

Goal 2—Develop 2YC students' skills in scientific practice.

The research question related to Goal 2 is: Does participation in the STEM-ENGINES program affect student views of self-efficacy as it pertains to conducting research?

Kardash survey. Since part of the students' training involves applying scientific research methodologies, the Kardash survey (Kardash, 2000) was implemented to provide a measure of students' self-efficacy as it pertains to conducting a scientific investigation. The reliability of the instrument comes from the dual nature of the administration of the survey. Kardash reports that faculty mentors indicate very similar student ability to engage in scientific processes in comparison to the student reported data. The survey consists of 14 items on a 5-point Likert-type scale and assesses student perception of their abilities in areas including but not limited to understanding concepts in the discipline, designing a research study, collecting and analyzing data, and disseminating the results of a study in various formats. Kardash (2000) developed and implemented this survey with junior and senior undergraduate university students involved in academic-year or summer research experiences. Essentially, the survey assesses students' self-reported ability to conduct a scientific investigation. The survey is administered to the students at the STEM-ENGINES program-wide meetings that typically occur annually in January, May, and October, when members of the assessment team have access to all program participants. This allows the data to be analyzed in a quasi-experimental design using control and treatment groups from within the program. In addition, the Kardash survey has been modified and administered to the faculty mentors of these students so that the mentors' perceptions of how their students have grown in these areas may be measured as well. This model of comparing faculty perceptions and student perceptions of a shared experience has been utilized successfully in the past (Kardash, 2000; Taylor, Fraser, & Fisher, 1997) and provides a more complete characterization of the shared research environment.

Goal 3—Instill confidence to pursue STEM fields, and Goal 4—Encourage degree completion.

Since results from the following instruments address the questions related to both Goal 3 and Goal 4, those questions are listed here together for ease of organization:

- Does participation in research affect student interest in pursuing STEM degrees?
- How do the benefits of research differ for community college students versus students at four-year institutions regarding pursuit of STEM fields and degree completion?
- Do the summer experiences affect student transfer decisions?
- Are STEM-ENGINES students earning associate degrees in STEM disciplines and continuing to pursue a STEM degree after transferring?

The focus on degree completion as well as pursuit of STEM education and careers addresses how a URC project like STEM-ENGINES can have an impact on the potential of students from underrepresented populations in STEM careers. Drawing upon work previously done by Kardash (2000), discussed above, and Lopatto

(2004) regarding undergraduate research experiences at 4YCUs, program personnel have incorporated the use of the Kardash Survey and the SURE to examine student self-efficacy toward scientific practice and their intent to pursue STEM careers. After completing research experiences during the academic year and the summer, STEM-ENGINES students are required to complete the surveys and report on how the experiences affected their confidence levels as well as professional and career goals. An annual tracking survey is also distributed to exiting students in an effort to record the actual educational and professional decisions made by students.

Survey of Undergraduate Research Experiences (SURE). The STEM-ENGINES URC utilized the SURE in order to examine whether students' career goals are altered by participating in a research experience (Lopatto, 2004). The original survey consists of 44 questions, including demographic information, designed to elicit students' interest in further pursuit of STEM education and careers as well as their self-perceived achievement gains in areas related to scientific research. Since the inception of the program, the SURE has been through several revisions, including the original SURE and the SUREay which includes items relative to the academic year experiences. In its current version, the SURE III incorporates aspects of the original SURE survey, designed for summer research students, and subsequently includes additional questions appropriate to students participating in academic-year research experiences as well. The SURE III survey is given to each cohort upon the completion of each summer research experience and upon completion of each academic-year research experience. These surveys are administered in an online format, analyzed by Dr. David Lopatto at Grinnell College, Grinnell, IA, and an aggregate summary report is distributed to the project evaluation team. There is no cost to complete the survey at this time. The SURE assessment and its original development occurred as a result of Howard Hughes Medical Institute and National Science Foundation funding; it is available at http://web.grinnell.edu/sureiii/. The value of this particular instrument is that it provides a viable method for comparing the 2YC students to students conducting undergraduate research at 4YCUs. This comparison provides an opportunity to determine if the 2YC research experience is having similar effects on 2YC students' decisions to pursue STEM education and careers and students' growth as scientists as on students at 4YCUs.

Tracking survey. Since the STEM-ENGINES URC is a five-year NSF-funded program being conducted at a partnership of 2YCs with a high student turnover rate, with many students potentially transferring from the program in three or four semesters, it is desirable to maintain communication with the graduates of the program and track their academic progress and movement into the workplace. A tracking survey was, therefore, developed internally and modeled after a tracking survey designed and implemented for a Summer Undergraduate Research Experience program at West Virginia University. The information gathered by the tracking survey includes, but is not limited to, where the participant attended 4YCU, student's major, and whether he or she continued participation in undergraduate research at their 4YCU. Information gathered from the tracking survey helps establish trends of student participants after they leave the 2YC.

Goal 5—Enrich academic preparation.

The research question related to Goal 5 is: Do our activities provide strong training in chemistry content and process domains?

Content Assessment. In order to measure academic performance of the students, we used a multiple-choice assessment instrument containing questions similar to those on nationally standardized general chemistry content assessment instruments such as those published by the American Chemical Society. Once per semester, STEM-ENGINES URC sets aside one hour for the students to conduct the various assessments and surveys that are being administered as part of the program. Therefore, due to the limited time available for administering all of the variety of assessments used in this program, utilizing a nationally standardized assessment such as an American Chemical Society General Chemistry Assessment instrument in its current form was unfeasible. In order to assess what knowledge students may have gained from their undergraduate research experiences, a modified version of a nationally standardized general chemistry content assessment instrument was designed by the evaluation team to measure general chemistry content knowledge. The assessment consists of 20 multiple-choice questions targeting nine topics including atomic structure, molecular structure, gas laws, thermodynamics, kinetics, kinetic molecular theory, acids and bases, oxidation-reduction reactions, and Henry's law. Table 1 shows a summary of each program goal and the assessment, as described above, that was used to assess achievement.

Table 1

Summary of Instruments Used to Measure Goal Attainment

Goal	Instruments Used
Identify and recruit promising students	Bar-On EQ-i
Develop student research skills	Kardash Survey; Modified Kardash Survey
Instill confidence to pursue STEM fields	Kardash Survey; SURE III Survey
Encourage degree completion	SURE III Survey; Tracking Survey
Enrich chemistry content preparation	General Chemistry Content Assessment

Method

Research Design

The evaluation utilizes each of the assessment instruments discussed above. With the exception of the SURE III, all instruments are administered three times per year, roughly corresponding to the beginning or end of any given semester (January, May, and October) of each semester of participation in the STEM-ENGINES URC. While this has resulted in multiple administrations of the instruments to many individuals, it has also given us the opportunity to collect longitudinal data as students progress through the program. Recognizing that participants in the STEM-ENGINES program have expressed an interest in pursuing undergraduate research as part of the academic career, this evaluation examines student development within the program instead of comparing STEM-ENGINES' students with other 2YC students that may or may not have an interest in undergraduate research. Undergraduate research opportunities are not the norm at the 2YC level, and those students who express a strong interest in research and enroll in the STEM-ENGINE program are not necessarily comparable to traditional 2YC students for analysis of these data. Participation in the STEM-ENGINES program is voluntary, and faculty members recruit students from within their classes. Many 2YC sites are characterized by high student turnover. In addition, students choose to enroll in the STEM-ENGINES program at different points in their academic career (e.g., some begin research after one semester of college, others after two or three semesters). For these reasons, it is desirable to group students into cohorts based on the number of semesters they have been enrolled and the number of semesters they have participated in research. By collecting student data throughout the year, many different comparisons may be made across cohorts, and comparisons within a particular cohort may also be completed. In this manner, it may be possible to more fully describe the benefits of participating in research at the 2YC level.

A quasi-experimental design was utilized for all but the data from the SURE III instrument. In order to establish if the research participation has had an influence on their science content knowledge and process skills, emotional intelligence, and interest in science careers, various cohorts of students from within the program were compared. An example of a treatment and comparison group would be where students who have completed two semesters of college and one semester of research are compared to students who have (a) also completed two semesters of college and (b) also have elected to participate in research but have not yet done so. This group would be enrolling in the program at the beginning of their third semester of college. Bauer and Bennett (2003) utilized a similar idea of identifying a comparison group that has had similar experiences, though potentially at a different time. However, Bauer and Bennett use department alumni as the comparison group only allowing for reflective perspectives to be compared.

In addition to these internal comparisons, the effect of the STEM-ENGINES program will also be examined by comparing evaluation data obtained via instruments common to other investigations, most importantly, the SURE III results. It is noteworthy that, at this time, the STEM-ENGINES URC is unique as it represents the only significant 2YC program contributing data to the SURE III program. The potential to compare students based on other demographic characteristics such as age, amount of general chemistry completed, total amount of research completed, race, and gender also exists.

Discussion

The current study is midway through a longitudinal study, and a plan to report final outcomes in future publications exists. Therefore, the present discussion focuses on our initial findings in each of the stated goal areas with the recognition that future findings will be more inclusive and potentially more conclusive. The report will also address the effectiveness of the assessments utilized in this project and their potential for use as viable methods of determining student gains at the community college level.

Currently, 19% of all students involved in the URC are still attending a 2YC. Additionally, 52% of students have transferred to 4YCUs with nearly 42% continuing to pursue academic and career paths in STEM disciplines. Table 2 describes the products that have come from the first three years of the project. Students have co-authored 64 presentations at regional and national meetings, four journal publications, and one patent application. Given the diverse nature of the demographics from which these students come, that in and of itself is a powerful testament to the influence that a research program like this can have on students. At the time of publication, some of the assessment instruments were still in the process of being administered for the first time (Mentor Survey, Tracking Survey). As such, there is little to report on these two instruments in terms of providing either program assessment or student assessment.

Table 2

Number of Student Co-authored Presentations, Publications and Patent Applications

Total of 168 STEM-ENGINES Fellows	Year 1	Year 2	Year 3
Local Presentations	45	75	90
Regional Presentations	2	14	40
National Presentations			8
Co-authored Publications		2	2
Number of Patent Applications	1		

Note: Specific references for co-authored presentations, publications and patents can be found at the project web site (http://stemengines.com).

The chemistry content assessment of student learning was originally designed for a project involving research specifically in chemistry. As the program has expanded, and now includes disciplines of biology and earth science, the original intent of the chemistry content assessment may no longer be as applicable. However, in order to maintain consistency in the evaluation of the project, the assessment continues to be used as the measure of growth of participants' content knowledge of general chemistry content. It has, in many regards, provided a useful model for longitudinal statistical analysis of the various groups of individuals who have participated in the project. Many of the research projects on which students are working, particularly during the summer experiences at the 4YCUs, involve predominantly organic chemistry or entirely different disciplinary content of biology and earth science. However, by using the general chemistry content assessment, we have the ability to look at another aspect of learning science—that of knowledge transfer across disciplines.

It has been discussed that the earlier students participate in undergraduate research, the greater the impact on their advancement of general chemistry content knowledge (Carver et al., 2010). In fact, preliminary results from this study show that students who participate in their first semester of research experience during their second semester of college show the greatest gains in general chemistry content knowledge as compared to (a) students who are at the same semester in college and (b) students who show an interest in research but have not yet participated in a research experience. As the comparison and treatment group comparisons have taken into account the students' chronological academic progress, the result is quite promising, particularly since the second semester of college is also when many students traditionally take general chemistry classes.

The Bar-On EQ-i (Bar-On, 2006), as used in this study, has shown that participants do not seem to increase their level of emotional functionality as a result of participating in this study. While that may seem like a negative result at first glance, it is not. All students who have participated in an administration of the Bar-On EQ-i have scored into the Enhanced Functioning range, indicating sufficient ability to understand and express themselves as well as relate to others and cope with daily demands. As such, there is little room for growth in the area of emotional intelligence. The EQ-i instrument may still be useful as a determinant of student readiness to participate in research, though at this time there is no data to support that, since all students are scoring into the Enhanced Functioning range. Furthermore, the STEM-ENGINES URC has been able to accommodate all students who have applied to the program to participate in undergraduate research, rendering the necessity of eliminating students based on their emotional intelligence scores unnecessary.

The modified Kardash survey (Kardash, 2000) has proven quite useful in assessing students' perceptions of their abilities to perform research functions. However, the modified Kardash survey has not yet established itself as a reliable tool in the assessment of student learning of chemistry process skills in this project by comparing student perceptions to faculty observations. One potential issue has arisen with the use of the modified Kardash survey. The modifications were such that the original survey was altered so that it could be administered to the faculty mentors, in addition to the students, during the tri-annual meetings of the STEM-ENGINES URC. Upon early administrations of the survey, it became clear that giving only one survey to each mentor was unrealistic, since each mentor may have multiple students working with him or her in the research group. In order for this analysis to be effective, each faculty member would need to fill out a survey for each student, and the student survey and faculty survey for that particular student would need to be paired in the dataset. As provisions for this had not been made at the onset of the program, the modified Kardash survey has not provided the richness of data for which it was intended, at least not from a comparative perspective.

Preliminary results for the use of the Kardash survey for student self-perceptions indicate no statistical increase in participants' perceptions of their knowledge of science process skills (ability to conduct an investigation). Similar to the results of the implementation of the Bar-On EQ-i instrument, students self-reported their knowledge of science process skills as being very high during the initial administrations of the modified Kardash survey, thus providing for very little room for improvement. This effect of higher-than-expected pretest scores is referred to as response-shift bias. A potential correction includes an additional modification of the Kardash survey to utilize a retrospective pretest model proposed by Howard et al. (1979), where participants complete a retrospective presurvey in conjunction with their postsurvey. Very often, participants report higher than realistic knowledge on an initial survey but it is only upon gaining new insights that they potentially realize that their initial impression of their skills and knowledge may have been inflated. Further revisions to the modified Kardash survey are underway, though are not likely to be implemented in the remaining year of the project funding.

The SURE survey has been through several iterations since the project inception. Now on version III, the SURE has allowed us to compare our students' perceived academic and career goals to other students in 4YCUs. At the time of publication, the STEM-ENGINES URC was the only significant 2YC project to be utilizing the SURE III survey, providing for a unique ability to study the impact of research at a 2YC. In fact, the use of the SURE instrument may provide the most compelling reason to include research programs at the 2YC. If 2YC students show more promise for change of academic and career goals than 4YCU students, then a renewed emphasis on recruitment and retention of 2YC students has the potential to make a substantial impact on the future of chemical education and, subsequently, the number of students majoring in STEM disciplines. In fact, preliminary results suggest that participants in the STEM-ENGINES URC are reporting higher knowledge of all but two of the 21 research-benefit items as measured by the SURE, as well as a stronger intent to pursue STEM disciplines in the remainder of their college careers than 4YCU students nationwide. Table 3 describes the comparison of the most recent group of 2YC participants to the national averages of all students taking the SURE survey.

Table 3

Anticipated Majors of STEM-ENGINES Students

Anticipated Major	STEM-ENGINES Participants	National Average
Biology	20%	48%
Chemistry	30%	8.5%
Biochemistry	20%	10.5%
Engineering	5%	4%
Education	10%	0.4%

While the percentage of participants reporting biology as a potential field of study is lower than the national average, the percentage of participants indicating chemistry and biochemistry are clearly higher than the national averages, potentially indicating that a focused research experience in chemistry may convince otherwise undecided students to pursue study in chemistry rather than biology. We are struck with one comment that we received on the most recent administration of the SURE, which we think encapsulates the entire theme and goal of our project.

> I evaluate the mentor as average but I am not sure because this was my first research experience and I didn't really know what to expect and I didn't want to exaggerate either. The mentors were pretty helpful and the experience overall was fine, but I surely know that I want to experience more of it.

The participant is clearly conscious of his own lack of reference for rating the mentor since this was his first experience. However, he is clear that the experience itself was such that he wanted more.

The current chapter has discussed the assessment of the students participating in the STEM-ENGINES URC. However, what has proved to be somewhat elusive is the assessment of the sixth goal of the URC—transforming the cultures of participating 2YCs by embedding intensive research experiences during the academic year and summer into their curricula and their courses. The culture of the community college is that of a teaching institution with research comprising almost none of a faculty member's workload. This URC set as one of its goals the task of establishing an institutional culture within the participating 2YCs which supports undergraduate research in STEM disciplines as a normal part of the faculty members' teaching load. While institutional changes have begun to take place, institutional culture has a large inertia and is slow to change. For instance, Harper College has built a new multimillion dollar science teaching and laboratory facility complete with research instrumentation (HPLC, NMR, FT-IR, UV-VIS, GC-MS, IC, AA). However, researchers from the other 2YCs are just beginning to utilize the facility. Additionally, it has been difficult for faculty members from different campuses to collaborate on research projects due to the time requirements established by the heavy teaching loads required at the 2YC level. Several of these cultural shifts are in the process of being studied using a newly developed Survey Undergraduate Research Faculty Mentors with faculty who have been involved in the program. Results of this study will be disseminated in future writings.

Several additional aspects of the program will require further assessment practices not described in this chapter, and many of which are still under development. One such assessment will probe the long-term effects of students who have participated in this program as they progress through their education and, ultimately, into their chosen careers. Additionally, since the program was designed as a chemistry-focused research program, but the research interests of participating faculty have broadened to include areas of research in biology and geology, additional assessments in biology and geology and perhaps a revision of the chemistry assessment may be in order to assess future academic enrichment of the participants involved in those originally unintended research areas.

Given the overall percentage of students being educated at the 2YCs across the country, the emphasis on providing a valuable learning environment must be re-examined to include educational experiences that have typically not been a part of a 2YC education, including, but not limited to, undergraduate research experiences. In order for that to occur, studies like the ones underway with the STEM-ENGINES URC need to become more prominent in describing the value of the opportunities available, the cultural shifts that can occur, and the student learning outcomes that are associated with authentic experiences such as undergraduate research in the STEM fields.

REFERENCES

American Association of Community Colleges. (2009). *AACC research and statistics*. Retrieved from http://www2.aacc.nche.edu/research/index.htm

Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicothema, 18*(Suppl.), 13–25.

Bauer, K. W., & Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. *Journal of Higher Education, 74,* 210–230.

Behar-Horenstein, L. S., & Johnson, M. L. (2010). Enticing students to enter into undergraduate research: The instrumentality of an undergraduate course. *Journal of College Science Teaching, 39*(3), 62–70.

Bleicher, R. E. (1996). High school students learning science in university research laboratories. *Journal of Research in Science Teaching*, *33,* 1115–1133.

Brothers, M. R., & Higgins, T. B. (2008). Research models that engage community college students. *Council on Undergraduate Research Quarterly, 29*(1), 12–17.

Brown, D. R. (2006). Undertaking chemical research at a community college. *Journal of Chemical Education, 83*(7), 970–972.

Brown, S., Bolton, K., Chadwell, N., & Melear, C. (2002, January). *Preservice secondary science teacher apprenticeship experience with scientists*. Paper presented at the annual meeting of the Association for the Education of Teachers of Science, Charlotte, NC.

Carver, J. S., Brothers, M. R., Dowd, T., Edler, K., Ferrence, G., Harris, Y., Persson, S. (2010, March). *Student learning gains through undergraduate research experiences with two-year college students.* Paper presented at the international meeting of the National Association of Research in Science Teaching, Philadelphia, PA.

Dawada, D., & Hart, S. D. (2000). Assessing emotional intelligence: Reliability and validity of the Bar-On emotional quotient inventory (EQ-i) in university students. *Personality and Individual Differences, 28(*4), *797*–812.

Etkina, E., Matilsky, T., & Lawrence, M. (2003). Pushing to the edge: Rutgers Astrophysics Institute motivates talented high school students. *Journal of Research in Science Teaching, 40*(10), 958–985.

Gaglione, O. G. (2005). Underground existence of research in chemistry in two-year college programs. *Journal of Chemical Education, 82*(11), 1613–1614.

Hathaway, R. S., Nagda, B. A., & Gregerman, S. R. (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: An empirical study. *Journal of College Student Development*, *43*(5), 614–631.

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement, 3*(1), 1–23.

Kardash, C. M. (2000). Evaluation of an undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology, 92*(1), 191–201.

Lopatto, D. (2004). Survey of Undergraduate Research Experiences (SURE): First findings. *Cellular Biology Education, 3*(4), 270–277.

Marcus, J. M., Hughes, T. M., McElroy, D. M., & Wyatt, R. E. (2010). Engaging first-year undergraduates in hands-on research experiences: The Upper Green River Barcode of Life Project. *Journal of College Science Teaching, 39*(3), 39–45.

Nagda, B., Gregerman, S. R., Jonides, J., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student-faculty research partnerships affect student retention. *The Review of Higher Education*, *22*(1), 55–72.

Palmer, J. C., & Vaughn, G. B. (1992). *Fostering a climate for faculty scholarship at community colleges.* Washington, DC: American Association of Community and Junior Colleges. Retrieved from ERIC database. (ED350048)

Richmond, G., & Kurth, L. (1999). Moving from outside to inside: High school students' use of apprenticeships as vehicles for entering the culture and practice of science. *Journal of Research in Science Teaching*, *36*(6), 677–697.

Ritchie, S., & Rigano, D. (1996). Laboratory apprenticeship through a student research project. *Journal of Research in Science Teaching*, *33*(7), 799–815.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science, 316*(5824), 548–549.

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching, 36*(2), 201–219.

Schwartz, R. S., Lederman, N. G., & Crawford, B. (2000, January). *Making connections between the nature of science and scientific inquiry: A science research internship for preservice teachers.* Paper presented at the annual meeting of the Association for the Education of Teachers of Science, Akron, OH.

Seymour, E., Hunter, A. B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates: First findings from a three-year study. *Science Education, 88*(1), 493–594.

Smolkin, R. (2003, Spring). Hidden bastions of chemical education. *Chemistry*, 22–27.

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research, 27*(4), 293–302.

Thompson, M. D. (2001). Informal student-faculty interaction: Its relationship to educational gains in science and mathematics among community college students. *Community College Review, 29*(1), 35–57.

Westlund, J. F., Schwartz, R. S., Lederman, N. G., & Koke, J. R. (2001, March). *Teachers learning about nature of science in authentic science contexts: Models of inquiry and reflection.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis, MO.

AUTHOR NOTE

The work included in this chapter was funded in part by the National Science Foundation (CHE-0539214 and CHE-0629174). The authors would like to acknowledge the following colleagues for their assistance in conducting the study: Thomas Dowd, Roger House, Yvonne Harris, Raymund Torralba, William J. F. Hunter, Gregory Ferrence, Kate Edler, Stephanie Persson, Catherine Han, Rowena Misayah, Liliana Marin, Dana Perry, Olu Akinbote, Ramon Castro, Julienne Thomas-Hall, Gitendra Paul, Austin Ferguson, Bridget Stuckey, Mahesh Gurung, Doris Joy Espiritu, Warren Menezes, Joseph Oyugi, Helen Rarick, Richard Jarman, Andy Kidwell, Julie Ellefson-Kuehn, Dan Stanford, Joseph Wachter, Melodie Graber, Michael Carr, Carol Ward, Gary Mines, Terry Trobec, Ken Brown, Marge Jones, Allen Hunter, Matthias Zeller, and David Lopatto. This manuscript was written primarily by primary author, Jeffrey S. Carver, and secondary author and project director, Morna Brothers, with strong contribution from project PI, Thomas Higgins.

CHAPTER 6

OPENING THE GATEWAY: THE REDESIGN OF A FRESHMAN CHEMISTRY COURSE AT THE UNIVERSITY OF MARYLAND EASTERN SHORE

Jennifer L. Hearne, Joseph M. Okoh, Yan Y. Waguespack, Amelia G. Potter, James R. Hayes, Gladys G. Shelton, Charles Williams, and Nancy Shapiro
University of Maryland Eastern Shore

Abstract

Principles of Chemistry I, the first-semester course in a two-semester sequence chemistry regimen designed for freshman science and health professions program majors at the University of Maryland Eastern Shore (UMES), was redesigned on the basis of the National Center of Academic Transformation's Replacement Model. The course was selected for redesign, as it suffered from the following issues: (a) inconsistent knowledge of incoming students; (b) an average 55% student retention rate in regard to students eligible for enrollment into the second part of the freshman chemistry regimen; and (c) a lack of coordination among the faculty members teaching the course, leading to course drift and inconsistent learning outcomes.

During the spring 2008 academic semester, the UMES offered Principles of Chemistry I in two formats: the traditional format and the redesigned format. Following the Replacement Model, the redesigned Principles of Chemistry I class encourages individualized, active learning; offers ongoing and prompt assessment and feedback through technology-based exercises; and offers on-demand, personalized assistance by introducing a variety of course personnel. A comparison of final grades (A–F, W, I) reveals an 11.1% increase in students' successful completion of the redesigned course in comparison to students in the traditional course. Cost analysis reveals a 60% decrease in institutional costs associated with offering the redesigned course relative to the traditional offering.

Opening the Gateway: The Redesign of a Freshman Chemistry Course

Public higher education in Maryland, as throughout the nation, is challenged by the need to increase access, improve the quality of teaching and learning, and control or reduce costs (University System of Maryland, n.d.a., n.d.c.). Historically, improving quality and/or increasing access has meant increasing costs, while reducing costs has generally meant reducing quality and/or access (University System of Maryland, n.d.c.). To sustain vitality while serving a growing and increasingly diverse population, the University System of Maryland, and higher education institutions in general, must find a way to resolve the familiar trade-offs amongst quality, cost, and access (University System of Maryland, n.d.c.).

Maryland Course Redesign Initiative

University System of Maryland (USM)

In the past, USM institutions adopted methodologies for the infusion of technology to enhance the teaching and learning process and to extend access to new and underrepresented groups of students (University System of Maryland, n.d.b.). The USM has been widely recognized for its successes in applying technology; however, the USM institutions, like most other higher education entities, have not fully harnessed the potential of technology to improve the quality of student learning, increase retention, address teaching efficiency, and reduce the costs of instruction in courses that have the broadest impact (Twigg, 2003; University System of Maryland, n.d.b., n.d.c.).

To address the use of technology in the enhancement of teaching and learning, as well as to absorb a projected 7% increase in student population over the next decade, the USM launched a system-wide, multiyear efficiency and effectiveness initiative aimed at maintaining or increasing quality of instruction while containing costs, the Maryland Course Redesign Initiative (MCRI; University System of Maryland, n.d.b., n.d.c.). The MCRI, launched in 2006, was planned and implemented by the USM System Office under the leadership of the Office of the Senior Vice Chancellor for Academic Affairs and Associate Vice Chancellor for Information Technology in cooperation with the National Center for Academic Transformation (NCAT).

Studies have shown that undergraduate enrollment in the United States is highly concentrated in introductory courses: at the baccalaureate level, the 25 largest courses generate about 33% of student enrollment at four-year institutions (Twigg, 2005; University System of Maryland, n.d.c.). These introductory gatekeeper courses are good prospects for technology-enhanced redesign projects, because they serve as foundation studies that help students make the transition to more advanced study (University System of Maryland, n.d.c.). Choosing such a course to redesign ensures a high impact on the curricula, specifically the rate of successful completion leading to increased retention and more efficient utilization of resources. Successful learning experiences in these courses can influence students to persist (University System of Maryland, n.d.c.), while high failure rates (typically 15% at research universities, 22–45% at comprehensives, and 40–50% at community colleges) can lead to significant dropout rates (Twigg, 2005).

Through the restructuring of large enrollment (relative to institutional size), multisection courses using technology-supported and active learning strategies, the goals of the MCRI were to (a) adopt methods to improve student learning outcomes; (b) demonstrate these improvements through assessment; (c) reduce institutional costs; (d) release instructional resources for other purposes; and (e) develop the internal capacity of USM faculty members to continue this process (University System of Maryland, n.d.c.).

The University of Maryland Eastern Shore (UMES)

The University of Maryland Eastern Shore (UMES), a member institution of the USM, is the state's historically black 1890 land-grant institution. UMES has made significant strides towards academic excellence. For the first time in its history, the university was ranked among the first tier of historically black colleges and universities in the nation. UMES is one of the four doctoral degree-granting institutions in the USM and is the only doctoral degree-granting institution on the Eastern Shore of Maryland.

During the spring 2008 academic semester, the semester in which the pilot phase MCRI data were collected, student enrollment at the UMES totaled 4,086 (University of Maryland Eastern Shore, n.d.).

Principles of Chemistry I

Throughout the past decade, several reports have been published regarding the improvement of undergraduate science, technology, engineering, and mathematics (STEM) education (Boyer Commission on Educating Undergraduates in the Research University, 1998, 2001; Committee on Undergraduate Science Education, 1999; Kuenzi, Mathews, & Mangan, 2006; Project Kaleidoscope, n.d.; Spellings, 2007). To address the need for STEM education reform and the goals of the MCRI, UMES redesigned Principles of Chemistry I.

At UMES, Principles of Chemistry I is the first course in a two-semester chemistry regimen designed for freshman science and health professions program majors. It is a requirement for approximately 31% of the academic degree programs offered by the university.

During the pilot phase of the redesign, the student composition of the traditional class was 82.8% freshman, 13.8% sophomore, and 3.4% senior students, while that of the pilot section was 52.6% freshman, 42.1% sophomore, 1.8% junior, and 3.5% senior students (self-reported). Regardless of classification, the classes catered to mostly traditional college students: 72.4% of the traditional section and 71.9% of the pilot section population (self-reported).

The goal of Principles of Chemistry I is to introduce students to the basic concepts in chemistry, which include density, basic atomic and molecular theory, chemical nomenclature, reaction stoichiometry, and the gas laws. The expected learning outcomes of Principles of Chemistry I are exhibit a working knowledge of the chemical principles; exhibit mastery of critical thinking, problem solving, and data analysis skills; and appreciate the interconnected nature of chemistry and mathematics.

Prior to the spring 2008 semester, UMES offered Principles of Chemistry I as a lecture course with a maximum enrollment of 50 students per section, a large class size for UMES. Professors used a combination of PowerPoint presentations and chalk talks. While the department encouraged the usage of the Blackboard Learning System, it was rarely used. Individual faculty members, up to three per semester, presented the material they desired, teaching and assessment methodologies were not standardized, and the enforcement of pre- and corequisite mathematics courses was lenient.

Establish a Team

The high level of success achieved in NCAT's course redesign programs can be attributed to dedicated participants teaching them the planning methodology and actively supporting them as they developed their redesign plans (University System of Maryland, n.d.c.). The redesign team should comprise faculty members who teach the course, administrators directly and indirectly involved in the course offering, technology professionals, and an assessment expert. By choosing such a diverse, dedicated team of people, issues encountered in the planning and execution of the redesigned course will be minimized.

Faculty. Faculty engagement is a critical factor in the success of a redesign project (University System of Maryland, n.d.b., n.d.c.). To ensure consistency, course redesign requires that faculty members collaboratively identify and agree on the course's desired learning outcomes, course content, materials, modes of delivery, and assessment methods (University System of Maryland, n.d.b, n.d.c.). Faculty members participating in the UMES redesign included a professor of chemistry, laboratory coordinator, and assistant professor of biochemistry.

Professors involved in the delivery of the redesigned course should agree on the materials employed in the pilot phase and should fully participate in all aspects of course development (Twigg, 2003). It is not always possible to make existing materials fit the redesign goals. The review and selection of materials is a tedious and time-consuming process. Prior to searching for materials, those involved should determine the most important features necessary to facilitate the redesign. In this case, a book was sought that (a) progressed in a logical order to facilitate the construction of knowledge; (b) was written with the freshman students' limited knowledge-base in mind; (c) was accompanied by a web-based program that consisted of exercises, coached problems, guided simulations of concepts, instant feedback for students, automated grading, grade book features, and a friendly user interface (Cengage Learning, n.d.); and (d) was sold to students at a reasonable price. After reviewing several texts, *Chemistry, The Molecular Science,* was selected for use (Moore, Stanitski, & Jurs, 2008). This text was supported by the web-based program CengageNOW (Cengage Learning, n.d.).

Administrators. Because a redesign impacts multiple sections, large numbers of students, as well as academic policies and practices, it is of great importance to include administrators on the redesign team (University System of Maryland, n.d.c.). The level of these administrators depends on the organization and size of the institution. The redesigned course presented in this chapter is offered by the Department of Natural Sciences (DNS), a multidisciplinary component of UMES with an enrollment of approximately 14% of the university's student population. As such, the immediate administrator, chairperson of the Department of Natural Sciences, secondary administrator, Dean of the School of Agricultural and Natural Sciences, and the primary administrator, Vice President for Academic Affairs, were deemed appropriate for the UMES redesign effort.

Technology professional. Since one goal of a course redesign is to incorporate technology, a technology professional is an integral part of the redesign

The combination of these issues caused the instruction of Principles of Chemistry I to suffer from the following academic issues: (a) inconsistencies in the knowledge of chemistry, English language skills, and mathematic skills of the incoming students; (b) low mastery of course content resulting in an average 55% pass rate; and (c) course drift and inconsistent learning outcomes due to lack of coordination among the professors teaching the multiple course sections.

The course was redesigned to alleviate these difficulties through the use of technology, coordination amongst the professors, introduction of mixed-course personnel, and standardization of assessment measurements. Assessment of the students' final grades and costs savings to the institution became an inherent part of the redesign initiative.

National Center for Academic Transformation (NCAT)

Several years ago, Twigg (2005) observed that there were several introductory courses offered by nearly every institution of higher education, and almost every student in the United States took at least one of those courses. This group tended to be the lower division general education courses and included courses in biology, chemistry, English, mathematics, and physics (Twigg, 2005). Typically, they were high-enrollment courses and were served through meeting in a large lecture section or in multiple smaller sections thereby requiring institutions to make large investments (Twigg, 2005). As a result, improvements made to these courses could have a significant impact on teaching efficiency and learning effectiveness (University System of Maryland, n.d.b.).

In 1999, Twigg (2003, 2005) proposed to the Pew Charitable Trust that there were likely models for improving learning outcomes and reducing the per-student cost for high-enrollment courses. The implementation of a model became known as the process of course redesign.

Five models for implementing a course redesign were captured: supplemental model, emporium model, fully online model, buffet model, and the replacement model (National Center for Academic Transformation, n.d.). With emphasis on selecting a model that would aid in the achievement of course consistency, decrease course preparation and delivery time, provide latitude in learning activities, and allow for a cost savings, the Replacement Model was selected for the redesign of Principles of Chemistry I at UMES. The Replacement Model reduces the number of in-class meetings; replaces some in-class time with out-of-class, online, interactive learning activities completed in a computer laboratory; and makes significant changes in remaining in-class meetings.

Course Redesign Implementation

The strategies for implementation of a course redesign have been distilled by NCAT into a broad set of processes regardless of the model chosen (Twigg, 2003; University System of Maryland, n.d.b.). The characteristics of course redesign that emerged aligned with characteristics proposed by Chickering and Gamson (1987).

important in student development: students share meanings and grow through interaction (Jacob, 1987; Kahveci, Gilmer, & Southerland, 2008; Leonard, 2000; University System of Maryland, n.d.c.; Zare, 2000). During the pilot phase, students were asked to form peer learning groups, and these groups were formed through student initiative instead of forced interaction and ranged in size from two to five students. The peer learning groups worked together in the classroom to complete assignments at the conclusion of class meetings, and groups were encouraged to study together.

Mastery of Learning

In a course redesign, more emphasis should be placed on learning and giving students more responsibility for their learning rather than lecturing (Millar, Kosuik, Penberthy, & Wright, 1996). Principles of Chemistry I is a foundation course in which students learn the skills (taking notes, asking questions, completing homework, and translating abstract knowledge into workable knowledge) needed to master the material presented (Barrow, 1994). These skills are essential in the successful navigation of the students' college career.

The pilot section incorporated two components to encourage the development of such skills and mastery of learning. The first component was the development of a series of notes to assist students in learning how to be successful in chemistry and in university study in general. These notes were posted on Blackboard after each lecture session. Reviews of mathematic concepts were incorporated into these notes.

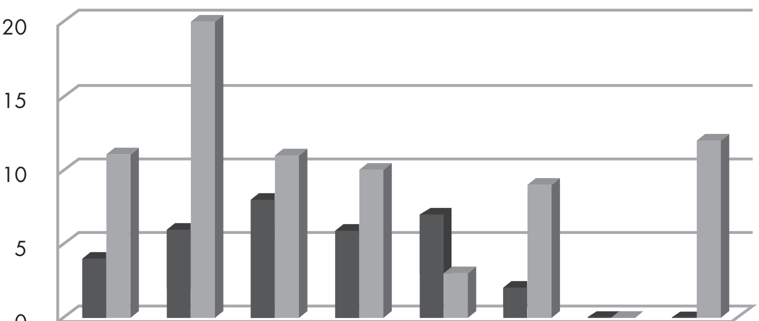
The offering of a recitation session was the second component in the Principles of Chemistry I redesign. The recitation session was mandatory for students with a cumulative average of less than 75%, yet all were encouraged to attend. The recitation session reviewed concepts, through written exercises, presented in that week's lecture.

Assessment

Student Assessment

It is important to note that students enrolled in either the traditional or redesigned course without knowledge as to which instructional methods they would receive. The students completed the same work but in different formats. The students in the pilot section completed their work online and were provided with two chances to successfully complete the work. Students in the traditional section were given the same assignment; however, it was paper-based and these students only had one chance to complete the work. Students were only given one chance to do their homework because there was a time lag of at least two days to return their work. The lag time which the students experienced was reduced from two days to instantaneous with the usage of CengageNOW.

A comparison of the percentage of successful students (students earning a grade of C or better) enrolled in Principles of Chemistry I was made between the pilot section and the traditionally taught section. Grading in each course was absolute (e.g., all exams were included and there was not a curve): examinations accounted for 39.5%, homework assignments for 30%, in-class assignments for 5.5%, a project for 8%, and the final examination for 17% of the final grade. Analysis of all participating students who were enrolled in the course after the Add/Drop period, revealed an 11.1% (65.6%) increase in the students earning the grade of C or better in the pilot section in comparison to the traditional section (Figure 1). Of the students who completed the course (e.g., those student who did not withdraw), 58.5% earned the grade of A–C in the traditional section, and 76.4% earned the grade of A–C in the pilot section. The average percentage of these students in the pilot section was 78.1 ± 12.33% and for the traditional section was 68.0 ± 17.9%.



	A	B	C	D	F	W	I	NP
Traditional	4	6	8	6	7	2	0	0
Pilot	11	20	11	10	3	9	0	12

Figure 1. Comparison of grades earned in the Traditional and Pilot sections of Principles of Chemistry I offered during the pilot phase of the MCRI at UMES.

Trends in student success were examined on the basis of gender as well. The traditional section was 30.3% male and 69.7% female. Analysis of grades earned in the traditional section reveals that 15.2 ± 2.1% male students earned the grade of A–C, whereas 39.4 ± 5.2% females earned the grade of A–C. Thus, female students outperformed their male counterpart. The pilot section male to female ratio was 53.1% male to 46.9% female. Here, the male students outperformed the female students: 35.9 ± 4.3% males and 29.7 ± 5.8% females successfully completed the course.

Cost Assessment

During the 2005/2006 academic year, UMES recorded a growth rate of 6.7%, the highest of any member institution in the USM, thus straining the limited resources of UMES. Increased enrollment and budget constraints at the state level, as well as federal and state implementation plans for technology access to all students, has led to a reduction of funds for courses. Thus, the implementation of this project was viewed as a creative way to cut costs at the departmental level while maintaining efficiency and implementing the state technology mandate. Costs associated with the course were examined relative to (a) time invested in course offering and presentation; (b) personnel costs; and (c) tasks associated with preparation and delivery of the course.

During the pilot phase offering, the pilot section catered to 76 students, thereby reducing the number of sections of the course offered and professors required to teach the course. Additionally, one 75-minute lecture session replaced the traditional three 50-minute lecture sessions per section per week. This decreased the amount of time required for course delivery.

Another cost reduction strategy employed in the redesign was the creation of two low-cost staffing positions. Not all tasks associated with a course require a faculty member's time (National Center for Academic Transformation, n.d.). These positions decreased the time the professor spent grading exams, recording attendance, grading in-class assignments, and the number of one-on-one contact hours with the students through the provision of on-demand assistance in the chemistry computer laboratory.

The final cost reduction strategy incorporated into the redesigned course was the introduction of the CengageNOW program. The use of CengageNOW provided for the substitution of human monitoring and hand-grading for automated monitoring and grading.

Appreciable savings were achieved by the university through the redesign of Principles of Chemistry I. Using the NCAT Course Planning Tool, the average cost per student in the traditionally offered course was calculated to be $268. The redesigned class decreased the cost per student to $151, a 44% decrease. The cost of the Chemistry Computer Lab is not included in this calculation.

Lessons Learned

Buy-In

Support from all interested parties is critical in the redesign process. Open dialogue should occur before deciding to undertake a substantial redesign.

Students

A few noteworthy observations regarding student population were made during the pilot phase:

1. The class population of the traditional class was 33, and that of the redesigned class was 76 (12 nonparticipating; nonparticipating students were defined as students who attended no more than two classes over the entire semester). Despite the increased student population, the students in the redesigned course outperformed those in the traditional course. This negates the idea that a smaller class population favors success.

2. In comparing the two classes, it is worth noting that 70.2% of the students in the pilot section (75% of the enrolled population at the onset of the semester, self-reported) and 44.8% of the students in the traditional course (87.8% of the enrolled population at the onset of the semester, self-reported) were participating in a college-level chemistry course for the first time.
3. Students in the redesigned course were more likely to seek assistance from the learning assistant rather than the undergraduate learning assistant or the professor. Therefore, it was of utmost importance to ensure that the assistance provided was well-coordinated and reflected the methods the students were exposed to during class. This was accomplished by inviting the learning assistant to attend the class meeting. Students in the traditionally taught course were unlikely to seek assistance.

The fact that male students outperformed female students in the pilot section was not surprising. It was informally observed that female students were more likely to work through the web-based problems as peer learning groups and were more likely to seek assistance. These actions may have contributed to an inflated sense of learned skills.

Students in both classes reported that Blackboard tools were helpful throughout the course and in preparation for exams. Those in the pilot section cited that the completion of assignments in the chemistry computer lab was crucial to their success in the course.

Course Structure

When the redesigned course was discussed with other participating USM faculty members, it was suggested that the recitation session could be viewed by students as punishment and most likely caused students to feel stigmatized and demoralized. Consequently, the course was restructured to include two 50-minute lectures during which recitation activities were incorporated.

Implementation Issues

During the pilot phase of the UMES course redesign, the team encountered several implementation issues, some anticipated and others not. The reluctance of some faculty and students to participate in the effort was not anticipated. Clearly, students who did not participate failed. Faculty members were initially skeptical but eventually became convinced that teaching method appreciably influences student learning outcomes.

Faculty

Other than reluctance of faculty members to modify their teaching methods, faculty were concerned about their workload increasing relative to their full-time equivalent fulfillment. Prior to the redesign, it was typical for faculty to instruct four classes of approximately 30 students to fulfill their full-time equivalent obligation. Post redesign, it is likely that faculty members will fulfill 25% of their full-time equivalent team. During the redesign of Principles of Chemistry I, the technology professional provided expertise so that the redesign goals were accomplished in ways that made the technology as easy for students to use as possible (University System of Maryland, n.d.c.). The UMES technology professional was particularly helpful with examining the requirements of the web-based program to ensure the program did not require new, or surpass current, system capabilities and would function given UMES's security and network requirements.

Assessment expert. Another important aspect of the MCRI was assessment. NCAT suggested straightforward methods that enabled student learning in the redesigned course to be compared to that of the traditional course. It is, however, useful to include someone who is knowledgeable about assessment, particularly if your institution seeks to measure additional facets of the redesign such as performance in downstream courses (University System of Maryland, n.d.c.).

Redesign the Whole Course

The long-term goal of a course redesign is to redesign the whole course; however, at the beginning, it is wise to test redesign ideas through the offering of a pilot phase. During the UMES pilot phase, one section of the traditional and one redesigned section (pilot section) of Principles of Chemistry I was offered. By redesigning one section of this course, the number of lectures per week was reduced from the traditional three 50-minute lectures on Monday, Wednesday, and Friday of each week to one 75-minute lecture on Monday of each week. This 75-minute lecture was used to introduce concepts that were explored in the web-based exercises assigned for the week.

The pilot phase provided the opportunity for the collection of baseline data: baseline data were obtained through the offering of the traditionally taught section in parallel with the pilot section. Alternatively, baseline data founded on institutional statistics can be used. Those involved in the UMES redesign felt more confident in the validity of the baseline data obtained during the pilot phase rather than using averages of data collected in the past due to the (a) use of similar materials; (b) comparison of a subsequent "trailer" course (Chemistry 111) offered out of sequence in the spring for students who were ill-prepared for the fall course or who failed the fall course; and (b) similarity in course delivery methods.

Encourage Active Learning

The goal of a redesign is to adopt methods to improve student learning outcomes. To address this goal, the teaching-learning enterprise needs to become appreciably more active and learner-centered (Polik, 2006; Twigg, 2003; University System of Maryland, n.d.c.). Active learning requires students to become engaged in the course—with course materials, staff, and each other. The more diverse the interactions, the more effective the teaching and learning processes are (Kahveci, Gilmer, & Southerland, 2008; Van Sickle & Spector, 1996). To do so, lecture time is replaced with a variety of learning experiences that move students from a passive, note-taking role to engaging exercises (Twigg, 2003; University System of Maryland, n.d.c.).

Incorporate technology. Three aspects of technology were included in the UMES redesign to encourage active learning: (a) use of the Blackboard Learning System; (b) use of CengageNOW, a web-based program published by Thomson; and (c) the establishment of a dedicated computer laboratory.

During the pilot phase, the Blackboard Learning System was used to provide students with instant access to the course syllabus, announcements, calendar of important dates, lecture notes, review exercises, and, most importantly, grades. Grades were posted weekly so students could track their progress.

The second technology component incorporated into the redesign was the use of CengageNOW, a web-based program published by Thomson, which provides tools for both the student and the professor (Cengage Learning, n.d.). For students, the program offers exercises, coached problems, guided simulations of concepts, a tutorial feature, instant feedback for students, and a friendly user interface. Students were regularly assigned work through the use of CengageNOW to facilitate repetition, probe preparedness and conceptual understanding, and increase the frequency and specificity of feedback to students.

CengageNOW is also beneficial to the professor. The program provides grade book features, automated grading, and a friendly user interface. The grade book is composed as students registered for CengageNOW, and it was maintained by the program. The option is available to analyze an individual's work, or the class as a whole through examination of the distribution of questions answered incorrectly, or the amount of time required to answer each question and/or complete each assignment.

The third element of technology incorporated into the redesigned course was the establishment of a dedicated computer laboratory. This was necessary to provide (a) students with an atmosphere that was conducive to learning; (b) individualized, on-demand assistance; (c) supplemental instruction to small groups; and (d) flexibility in scheduling study time. Although students enrolled in the pilot section could spend unlimited time in the chemistry computer laboratory, it was mandatory that they spend at least two productive hours per week in the laboratory completing CengageNOW assignments or receiving tutoring.

Alternative staffing and the provision of on-demand assistance. In an effort to increase diverse interactions, learning assistants and undergraduate learning assistants were introduced to expand the support system for students in the redesigned course (Twigg, 2003; University System of Maryland, n.d.c.). These individuals were accessible for up to 50 hours per week to provide students with the added flexibility to encourage learning in the chemistry computer laboratory (Young & Langford, 1971). The learning assistant aided in the delivery of recitation, proctoring of examinations, development of the course notes, and the grading of examinations. Most importantly, the learning assistant provided on-demand assistance in the chemistry computer laboratory. The undergraduate learning assistant served as a liaison among faculty and students involved in the redesign and provided on-demand assistance to students in the chemistry computer lab.

Meaningful connections are made apparent not only by the professor but also via peer interactions. Collaborative, peer-influenced learning is effective and teaching 120 students. The administrators on the team should plan for issues such as this one well in advance of the course redesign process.

Students

Despite the daily use of technologically advanced items, students were not proficient in the use of the web-based program. They struggled with following on-screen prompts and directions. It is highly recommended that one spend adequate time demonstrating the features of the program if such a web-based program is incorporated. This is easily accomplished by a learning assistant during out-of-class hours.

Resources

Due to the increased section size, the university was able to provide the Principles of Chemistry I course to 30% more students per semester than before. The increased student population enrolled in the course significantly strained the resources for the offering of the laboratory corequisite.

Sustainability

In support of these results, the university continues to support the effort of course redesign through the provision of infrastructure to redesign the second semester of freshman chemistry, Principles of Chemistry II, and the first semester of freshman biology, Principles of Biology I.

REFERENCES

Barrow, G. M. (1994). General Chemistry and the basis for change. *Journal of Chemical Education, 71*, 874–878.

Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* New York: Stony Brook State University.

Boyer Commission on Educating Undergraduates in the Research University. (2001). *Reinventing undergraduate education: Three years after the Boyer Report*. New York: Stony Brook State University.

Cengage Learning. (n.d.). *Cengage Learning.* Retrieved from http://www.cengage.com/about/

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *The American Association for Higher Education, 39,* 3–6.

Committee on Undergraduate Science Education, Center for Science, Mathematics, and Engineering Education National Research Council. (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology.* Washington, DC: National Academy Press.

Jacob, E. (1987). Qualitative research traditions: A review. *Review of Educational Research, 57*, 1–50.

Kahveci, A., Gilmer, P. J., & Southerland, S. A. (2008). Understanding chemistry professors' use of educational technologies: An activity theoretical approach. *International Journal of Science Education*, *30*, 325–351.

Kuenzi, J. J., Mathews, C. M., & Mangan, B. F. (2006). *Science technology, engineering, and mathematics (STEM) education issues and legislative options*. Washington, DC: Congressional Research Service.

Leonard, W. H. (2000). How do college students best learn science? *Journal of College Science Teaching, 29*, 385–388.

Millar, S. B., Kosiuk, S., Penberthy, D., & Wright, J. C. (1996). *Faculty assessment of the effect of a freshman chemistry course.* Madison, WI: University of Wisconsin Learning through Evaluation, Adaptation, and Dissemination Center.

Moore, J. W., Stanitski, C. L., & Jurs. P. C. (2008). *Chemistry, the molecular science* (3rd ed.). Pacific Grove, CA: Thomson Brooks/Cole.

National Center for Academic Transformation. (n.d.). *Six models for course redesign*. Retrieved from http://www.thencat.org/PlanRes/R2R_ModCrsRed.htm

Polik, W. F. (2006). Report of an ACS presidential symposium. *Journal of Chemical Education*, *83*, 17–18.

Project Kaleidoscope. (n.d.). *What works visions*. Retrieved from www.pkal.org/collections/whatworks.cfm

Spellings, M. (2007). *Report of the Academic Competitiveness Council*. Washington, DC: U.S. Department of Education.

Twigg, C. A. (2003). Improving learning and reducing costs: Models for online learning. *EDUCAUSE Review*, *28*, 28–38.

Twigg, C. A. (2005). *Increasing success for underserved students: Redesigning introductory courses.* Saratoga Springs, NY: The National Center for Academic Transformation.

University of Maryland Eastern Shore. (n.d.). *Institutional research, planning, and assessment*. Retrieved from http://www.umes.edu/IEAO/

University System of Maryland. (n.d.a.). *About University System of Maryland*. Retrieved from http://www.usmd.edu/about_usm/

University System of Maryland. (n.d.b.). *Efficiency and effectiveness reports 2006*. Retrieved from http://www.usmd.edu/usm/workgroups/EEWorkGroup/eeproject/index

University System of Maryland. (n.d.c.). *MCRI: Call to participate*. Retrieved from http://www.usmd.edu/usm/academicaffairs/courseredesign/?zoom_highlight=2006+mcri+call+to+participate

Van Sickle, M., & Spector, B. (1996). Caring relationships in science classrooms: A symbolic interaction study. *Journal of Research in Science Teaching, 33*, 433–453.

Young, J. A., & Langford, C. H. (1971). Recognition of individual students in large classes. *Journal of Chemical Education, 48*, 795–796.

Zare, R. N. (2000). On the love of teaching and the challenge of online learning: A few reflections. *Journal of Chemical Education, 177*, 1106.

AUTHOR NOTE

Jennifer L. Hearne, Department of Natural Sciences, University of Maryland Eastern Shore; Joseph M. Okoh, Department of Natural Sciences, University of Maryland Eastern Shore; Yan Y. Waguespack, Department of Natural Sciences, University of Maryland Eastern Shore; Amelia G. Potter, Department of Natural Sciences, University of Maryland Eastern Shore; James R. Hayes, Information Technology Department, University of Maryland Eastern Shore; Gladys G. Shelton, School of Agricultural and Natural Sciences, University of Maryland Eastern Shore; Charles Williams, University of Maryland Eastern Shore; Nancy Shapiro, University System of Maryland System Office.

CHAPTER 7

CLASSROOM ASSESSMENT IN SUPPORT OF BIOCHEMISTRY COURSE REFORM AT SEATTLE UNIVERSITY

Jennifer Loertscher
Seattle University

Abstract

In 1997, Seattle University made the transition from a traditional, lecture-based biochemistry course to a course based on guided-inquiry activities completed in structured small groups. In 2007, we were awarded National Science Foundation funding to test and disseminate, on a national level, biochemistry active learning materials developed at Seattle University. The move from making curricular changes in our own classroom to collaborating with biochemistry instructors nationwide was driven by assessment that began informally and has become progressively more formal. Immediately after switching to an active learning format, we observed that students were able to answer more difficult, open-ended exam questions than students in our previous lecture-based class. Over the next several years, we collected data from student exams to support this observation. For the past three years, we have been collecting data on student perceptions of their learning in the course using a Student Assessment of their Learning Gains (SALG) survey (http://www.salgsite.org). Finally, during the academic years 2009/10 and 2010/11, we will be piloting a pretest/posttest to assess student conceptual gains that result from using our materials. This assessment will be piloted at Seattle University and six collaborating institutions at which our materials are being tested. Data collected from all of these assessment efforts were used to refine our active learning materials in order to improve student learning. This case study of our experiences could act as a model for chemistry instructors interested in using evidence from their own classrooms to shape course design and to elevate the quality of instruction in their courses.

Introduction

Good teachers create an environment in which continual improvement in student learning is a priority. A well-established method to elevate student performance in the classroom and on a programmatic level is to use an assessment cycle that includes setting clear goals for student learning, assessing how well instructional design helps students reach those goals, and responding to assessment with appropriate changes. Yet, some faculty members do not have a clear idea of how assessment could be used to improve student learning in their own classrooms. The goal of this chapter is to provide a context for the discussion of assessment as it relates to chemical education using a case study describing the development and assessment of biochemistry active learning materials.

Assessment in the Chemistry Classroom

The term *assessment* has different meanings in different contexts. *Summative assessment,* sometimes called evaluation, involves a judgment based on information collected up to a given time (Taras, 2005). A midterm exam in a course would be an example of summative assessment. In contrast, *formative assessment,* which will be the focus of this chapter, is a systematic process that uses evidence to make decisions about changes in instructional design that could result in improved student learning (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Walvoord, 2004). Assessment, therefore, relies on a clear definition of desired outcomes and implies an intention to make changes based on findings. Taken this way, assessment can be thought of as action research, aimed at making changes in a given classroom or laboratory in response to data and observations (Mettetal, 2001). A complete formative assessment cycle, therefore, starts with definition of desired learning outcomes, uses a plan to collect relevant data, and ends by applying findings to make appropriate changes, which better support learning. The amount of time needed to complete one cycle varies depending on instructor and student needs. For example, data could be gathered in a course with the intention of making changes the next time the course is offered. Alternatively, the whole cycle could be completed within one class period using data gathered to make changes in real time.

Using assessment as part of course design should come naturally to scientists, because it allows us a way to use the principles of scientific inquiry to investigate our own classrooms (Handelsman, Miller, & Pfund, 2007). Just like science, classroom action research demands that investigators pose intentionally designed questions, design a method to gather evidence, collect data, and use the results to direct the next course of action (Savory, Burnett, & Goodburn, 2007). One practical consequence of classroom inquiry is backward course design (Dick, Carey, & Carey, 2005; Wiggins & McTighe, 2005). Using this approach, an instructor designs a course by first considering the desired student learning outcomes. Once outcomes are defined, the teacher can choose instructional approaches for the course that have been shown through classroom inquiry to support the desired learning outcomes.

Some instructors experience a significant "activation barrier" when considering an assessment project for their classroom because they have the notion that high-quality assessment projects are necessarily large in scope. In fact, experienced teachers often know intuitively that one of the greatest strengths of formative assessment is that it can be used to identify and address problems in student understanding in real time. For example, in a lecture-based course, clicker questions can be used periodically to monitor student understanding of central concepts related to that day's lecture. If student responses indicate that understanding is low, a teacher may choose to respond immediately with further clarification or additional examples. Active learning classrooms provide even more opportunity for real-time assessment, since the instructor can see first-hand how students are processing information given in a daily activity. If common problems in understanding are observed, an instructor may choose to interrupt a classroom activity with a short just-in-time lecture aimed at clarifying troublesome points.

Assessment is a Scholarly Endeavor

Assessment allows instructors to take a scholarly approach to teaching. Classroom assessment meets many of the traditional expectations of scholarship including the thoughtful definition of research questions that serve a specific purpose, the collection of data, and the use of evidence to make claims and plan action. Communicating the results of classroom investigations at professional conferences or in peer-reviewed journals is the final step needed for many to view assessment as true scholarship. Although the term *scholarship of teaching and learning* can be used in a number of ways, it generally refers to classroom assessment that intends to go one step beyond improving teaching and learning in isolated classrooms. Teachers who intend classroom assessment results to be applied to the scholarship of teaching and learning start with good classroom assessment techniques, and then connect to a wider scholarly context by making use of the literature and publishing results (McKinney, 2007; Savory et al., 2007).

The chemistry community has a long and rich history of engagement with education. Comfort with conversations about chemical education sometimes leads people to unwittingly conflate two distinct scholarly areas–chemical education research and the scholarship of teaching and learning. The following definitions aim to clarify the differences and alleviate confusion. Chemical education research and the scholarship of teaching and learning operate under two different paradigms with different motivations. Chemical education research is an interdisciplinary field that applies theories and methods from education, psychology, and sociology to teaching and learning in chemistry (Bunce & Cole, 2008b). It is motivated by a desire to create or support theories of learning in chemistry. Therefore, chemical education research asks experimental questions that are grounded in accepted theory, and data collection is performed using metrics that adhere to standards established in social science research. In comparison, the scholarship of teaching and learning has its own growing literature and also poses questions that are grounded in accepted theory. It is most often motivated by a desire to make local changes in student learning. As a consequence, the lived experience of teachers and students is central to study design, and the time from posing a question to making changes based on results should be short. Both chemical education research and the scholarship of teaching and learning are valued by the community as evidenced by the inclusion of both forms of scholarship in prominent disciplinary journals like *The Journal of Chemical Education*.

A scholarly approach to teaching has the power to do more for the field of chemistry than improve student learning in isolated classrooms. The American Chemical Society (ACS) recognizes the transformative power of using evidence to support changes in chemistry education. The current ACS scholarship statement embraces a broad definition of scholarship that includes and emphasizes the importance of the scholarship of teaching and learning:

> The scholarship of teaching and learning is still perhaps the least understood and recognized of all the forms of scholarship, but has the potential to transform chemical education. It must be encouraged and its role in preparing scientists for the new millennium must be recognized. (American Chemical Society Legislative and Government Affairs, 2010)

Likewise, the National Research Council argues that the guiding principles of scientific research should also be applied to education in the sciences (National Research Council, 2002). The biochemistry community is also beginning to recognize the importance of using evidence to inform decisions in the classroom, as seen by a recent series in *Biochemistry and Molecular Biology Education* entitled "Bridging the Educational Research-Teaching Practice Gap" (Anderson, 2007a, 2007b).

A Case Study—Use of Assessment in the Biochemistry Classroom

Over the past six years, I have been involved in an effort to write and use active learning materials in the biochemistry classroom. This project began as an effort to improve student learning in my own classroom and evolved into a broader effort to improve, disseminate, and assess materials at colleges and universities nationwide. This case study of my experience provides a model for instructors interested in using assessment to improve learning in their own classrooms and describes one approach for making the transition from classroom-action research to formal chemical education research. The following case study describes three distinct phases: (a) the move from lecture to active learning; (b) the use of classroom assessment activities to support and monitor instructional changes at one institution; and (c) the development of large-scale assessment projects to support national dissemination of instructional innovations.

Case Study: Assessment of Undergraduate Biochemistry Curriculum

Part 1—From Lecture to Active Learning

Biochemistry at Seattle University is taught over three quarters in a small group, active learning format using materials written and refined by the biochemistry instructors (Loertscher & Minderhout, 2009; Minderhout & Loertscher, 2007). Students in the course are mostly seniors and are a mix of chemistry, biochemistry, biology, and general science majors. The activities, which drive the content focus for each class period, cover the standard range of topics expected in an upper-division biochemistry course including macromolecule structure and function, enzyme kinetics and inhibitions, and metabolism. Nucleic acid biochemistry is taught in a literature-based course, which was not included in assessment projects described here.

The transition from lecture-based courses to active learning occurred in 1997 after colleagues attended a workshop about process-oriented classrooms (Hanson & Wolfskill, 2000). This type of instruction is based on the constructivist theory of learning, which holds that knowledge is constructed in the mind of the learner (Bodner, 1986). A consequence of this theory for teaching is a shift in the role of the teacher from an authority, who transmits knowledge, to a facilitator, who guides students to construct their own knowledge. Instructional methods used in process-oriented classrooms are informed by research aimed at determining which classroom practices best support learning and transfer of knowledge to new settings. Cooperative learning, guided inquiry, and metacognition (the process of thinking about thinking) have been shown to promote deep and lasting understanding (Bransford, Brown, & Cocking, 2000) and all are incorporated into process-oriented learning (Hanson & Moog, 2007).

Since transitioning from lecture to active learning, biochemistry at Seattle University has been taught using process-oriented guided inquiry learning (POGIL). POGIL, which has grown to be a major force in chemistry undergraduate education, uses guided-inquiry activities to help students build content knowledge and classroom structures to help students gain content-independent skills (Process Oriented Guided Inquiry Learning, 2009; Spencer, 1999). Therefore, at the beginning of each quarter, students are given not only a list of expected biochemistry content outcomes but also an additional set of skills-based outcomes that are not specifically linked to a particular area of biochemistry content (Table 1).

Part 2—Examples and Outcomes of Classroom Assessment

In the biochemistry classroom at Seattle University, assessment is used to make changes on a variety of time scales. Some assessment data provide useful information about how changes in the overall structure of the course can be improved from year to year to better support student learning of course objectives. Other assessment results can be used to make changes within a given quarter or semester as a course

Table 1

Learning Outcomes for a Senior-Level Biochemistry Course

Goal Type	Goal Statements
Cognitive	• Further acquire and master the vocabulary of biochemistry • Improve problem-solving skills using methodology • Improve ability to read primary journal articles • Analyze and interpret data • Improve visualization and modeling skills • Improve the ability to ask questions, examine assumptions, and solve problems • Apply knowledge to new and different situations • Strengthen critical thinking skills • Associate new understanding with prior knowledge • Understand the big picture
Social	• Work cooperatively • Listen to and learn from peers • Value others • Demonstrate commitment to a group
Affective	• Obtain a belief in one's ability to learn and apply the material • Advance intellectual tolerance and integrity • Set personal goals for improvement • Ask for help
Metacognitive	• Become self-directed—initiate the learning process • Become self-reflective—review goals, purposes, and outcomes • Become a self-assessor—assess your own progress for strengths, areas for improvement and insights into your learning process to continually improve

is ongoing. Still other assessment methods are aimed at gathering data and making changes in real time to better support student understanding of key concepts or performance of essential skills during a given class period. Assessment in all three time frames can be managed by designing, using, and modifying daily facilitation plans (see Figure 1 for an example; Minderhout, 2007). Facilitation plans can be used to direct the instructors' activities in the classroom and can promote continual improvement in instruction by linking assessment data from prior years or class periods to upcoming instruction. The subsequent paragraphs describe some approaches that have been used to accomplish short- and long-term assessment goals.

Prior to using a facilitation plan, the exam was the primary means of collecting data on student understanding of topics covered in class. For example, instructors

1. Learning Outcomes:

- Identify peptide bond and structural features of bond (review bonding)
- Determine charge on AA at given pH (use pKa)
- Determine pI of small peptide (use pI)

2. Activity Type: Guided discovery

3. Roles: Manager, Spokesperson, Recorder, Reflector

4. Student Preparation Assessment Plan:

- Examine assignment for correct lysine curve. Identify one in each team that is correct to act as a resource.
- Check that all are complete, regardless of correctness

5. Activity Set-up (Time: 5 minutes): Application of pKa in a new context

6. Group Work (Time: 40 minutes): Following are misconceptions or support that may be needed.

- Cognitive: Incomplete analysis of coplanar nature of peptide bond (ignoring resonance)
  - o Prior knowledge issue—remind students to think of organic chemistry
  - o Encourage students to draw resonance structures
- Cognitive: Inaccurate analysis of amino acid side chain charge
  - o Student titration curve should help, are they using it?
  - o Encourage students to consider charge
- Cognitive: C and N terminal have altered pKa values in peptides as compared to free amino acids. Be sure students have noticed this in the text since it is not in the table.
  - o Book discusses peptides on page 84
- Social: Poor communication, not listening
  - o Have manager solicit input on listening
- Affective: Members withdraw
  - o Engage withdrawn students in conversation

7. Closure (Time: 20 minutes):

- Have reflector complete report to teams
- Report group answers to questions
- Spokesperson shares discoveries

Figure 1: Sample Facilitation Plan—Introduction to Proteins.

had noticed that many students were unable to draw titration curves for peptides, and many of them had a fundamental misunderstanding of peptide bond structure. After observing these problems on exams, during the subsequent year, instructors took detailed notes about student performance on activities related to these concepts. Instructors listened to students during group work and read written responses. Based on these classroom observations, the activity was modified to better guide students to desired learning outcomes, and the facilitation plan shown in Figure 1 was created for use in subsequent years. In this way, problems identified on a midterm exam were used as a prompt to make changes in classroom instruction, and effective changes were recorded as part of a facilitation plan to be used in subsequent years.

Active learning classrooms provide more natural places for real-time classroom assessment than a traditional lecture-based classroom. Daily assessment in the active learning classroom is guided by use of a facilitation plan. Identifying difficulties that students have in real time can allow for immediate instructor intervention as well as long-term changes in the course to better support targeted learning outcomes. For example, part 6 of the facilitation in Figure 1 alerts the instructor to anticipated challenges that students might have with a given activity. These ideas were noted and recorded in previous years. The prompts in part 6 account for cognitive, social, and affective learning outcomes. Such prompts work best when instructors list not only anticipated challenges, but also what behaviors would be observed in the classroom when a specific challenge arises. Finally, it is also helpful if the facilitation plan suggests possible instructor interventions that could help students overcome listed challenges. Such facilitation plans also proved to be useful when instructors outside of Seattle University began to use POGIL biochem activities.

Student perception data can provide another window into the student learning experience, which may not be readily apparent through observing students while they work. For example, students could be asked to complete a midterm course assessment once or twice throughout the quarter. These surveys could include questions such as, "What aspects of the course support your learning well?" or "What changes could the instructor make to better support your learning?" In my classroom, I summarize student responses and present them to the class. I always choose several of the suggested changes I am willing to make as well as identifying those aspects of the course I am not going to change and the reasons why. In addition to midterm course assessments, students could be asked to complete periodic self-assessments. The primary reason for asking students to complete these exercises is to promote metacognition, which has been shown to improve learning (Bransford et al., 2000), but the self-assessments also provide instructors with more student perception data.

In order to track student response to recent curricular changes, biochemistry instructors at Seattle University have asked students to complete a Student Assessment of their Learning Gains (SALG) survey asking them about their perceived acquisition of content knowledge and process skills. These data have been useful in identifying both content and skills areas that are either well-developed or deficient in students who have completed the biochemistry course. A description of some of these previously published results is given below (Minderhout & Loertscher, 2007). For example, SALG survey results suggested that students lacked understanding of basic biochemical techniques, with only 63% of students stating that that they understood this topic "a great deal" or "a lot." These findings led to development of new and improved activities related to biochemical techniques. Questions aimed at understanding skills development showed great progress as a result of the course with 92% of students reporting substantial gains in taking responsibility for their own learning and respect for the opinion of others and 79% reporting substantial gains in confidence in their ability to learn complex material and ability to think through a problem. Finally, 83% of students strongly agreed that activities increased their understanding of the concepts, and only 21% of students strongly agreed that they learned better in lecture courses. These results assured instructors that most students had accepted and perceived a benefit from learning biochemistry in an active learning setting.

Part 3—Dissemination of Instructional Innovations and Ongoing Assessment Efforts

After several years of conducting the above-described assessment of active learning materials at Seattle University, it became clear from conversations with colleagues at national meetings that there was widespread interest in a complete set of biochemistry active learning materials. With support from the National Science Foundation (NSF), POGIL biochem activities designed and refined at Seattle University are now being used and assessed at a number of diverse institutions nationwide. Assessment data gathered from beta-testing institutions is being used to improve activities in preparation for broad dissemination.

Instructors who are using POGIL biochem materials in their classrooms (beta-testers) have agreed to provide a variety of assessment data. Two questions are driving the assessment efforts in beta-testing classrooms: (a) Are POGIL biochem activities appropriate for use at diverse institutions, and (b) Are students who are using POGIL biochem materials meeting defined expectations for learning in biochemistry? A variety of data addressing these questions are currently being collected including written formative feedback from instructors, faculty interviews, student perception (SALG) surveys, and answers to a common embedded exam question that beta-testing faculty include on their final exam. Funding from NSF also supports annual workshops in which beta-testing faculty are given feedback about assessment results. Thus, assessment data not only support ongoing improvement of POGIL biochem activites, but they also give beta-testers the opportunity to use assessment data to make improvements in their own classrooms.

The NSF-funded project has also enabled initiation of a chemical education research project. In collaboration with chemical education research experts, an instrument is being designed to assess students' conceptual gains as a result of using POGIL biochem activities. A pre- and posttest will be administered to determine whether student conceptual understanding of topics from general chemistry and general biology is improved as a result of using a common set of POGIL biochem activites. The process of designing an effective instrument is challenging and ongoing, but should provide new data indicating how learning in biochemistry, an upper-division course, relates to learning in foundational courses.

Using Classroom Assessment to Support Teaching and Professional Goals

Many instructors who collect assessment data in the classroom are motivated by an intuitive sense that knowing more about what their students experience in the classroom could provide the basis for making rational changes in their courses. Some also intend to publish articles about classroom innovations. However, few chemistry instructors carefully plan their intended use of assessment data in advance, and even fewer consider how such activities relate to their overall professional goals. The following section summarizes issues to consider before beginning assessment projects and strategies to use classroom assessment efforts to help achieve instructional and professional goals.

Define Goals Within a Local Context

Before beginning a classroom assessment project, it is important to clarify the pedagogical *and* professional goals of the endeavor. Classroom assessment can remain fully contained within the teaching arena, but with increased demands on faculty for publication even at primarily undergraduate institutions, the desire to translate classroom activities into publishable work can be strong and sometimes strategic. Ideally, well-planned classroom assessment can result in the dual benefits of improved student learning and discipline-appropriate publications. In order to obtain the largest return from time invested in assessment projects, it is important to set short- and long-term goals that integrate institutional culture with individual professional goals.

Expectations for faculty scholarship depend on the specific culture of each institution and department. Publication in peer-reviewed journals is the major measure of scholarly productivity in most chemistry and biochemistry departments, but the number of expected publications and the preferred journals for publication vary widely. As faculty consider whether results of planned assessments may be appropriate for publication, it is important for individuals to first become familiar with the culture of their home departments. Faculty should consider whether all publications in peer-reviewed journals are given equal weight at their institutions or whether some particular journals or subdisciplines are considered more prestigious than others. College or university scholarship statements can provide useful information, but conversations with trusted colleagues are also crucial.

Develop Connections with Like-Minded Colleagues

Being part of an engaged intellectual community is vitally important in any academic undertaking. As scientists, we are trained to work as part of a team; frequent collaboration with colleagues is the norm. In our laboratory research, we rely on interactions with other scientists to help us improve the quality of our work in a number of ways. For example, conversations with colleagues can help us focus research questions or interpret the significance of data within the context of the field. Despite the clear advantages of collaboration, teaching has traditionally been a relatively isolated undertaking. Yet, teachers interested in using a scholarly approach to their instructional design could clearly benefit from the same sort of interactions. In many instances, establishing a supportive and intellectually challenging community of peers interested in teaching can be more important—but also more challenging—than making similar connections in projects related to laboratory research.

A consideration of some of the differences between laboratory research and teaching can be informative when thinking about how to initiate conversations about classroom assessment. Most scientists received little formal training in teaching. As a result, colleagues may feel less comfortable being seen as experts in teaching than they are being viewed as experts in the laboratory. Furthermore, teaching is a human endeavor, and therefore problems in the classroom often feel more deeply personal than those in the laboratory. It is easy to view our students' academic shortcomings as personal failures. For these reasons and others, conversations about teaching often occur in the "back stage" of our professional lives (Goffman, 1959). When trying to connect to a network of colleagues interested in discussing teaching and learning, it is useful to determine whether teaching is "front stage" or "back stage" at a given institution. Starting conversations in the appropriate context can lead to more fruitful discussions and productive relationships. The scholarly approach to teaching described in this chapter and elsewhere can be used in the classroom without making it a part of one's professional scholarly endeavors. When deciding whether to keep assessment projects firmly linked to the classroom or to communicate findings to others as part of the scholarship of teaching and learning, it is important to keep in mind that not all institutional cultures support the scholarship of teaching and learning. Therefore, each individual should consider the prevailing institutional culture when making decisions about the role of classroom assessment in scholarly life.

Sometimes it is necessary and enriching to look beyond one's home institution to interact with peers about teaching and learning. Not everyone is fortunate enough to have colleagues interested in scholarly discussions about teaching within their own departments, but conversations with those from other departments and institutions can be equally valuable. Networking at regional and national meetings and using a campus teaching center to make local connections can be effective ways to meet colleagues engaged in the scholarship of teaching and learning. Finally, workshops, many of which are part of NSF-funded initiatives, can be excellent settings in which to work collaboratively with like-minded peers while also building teaching skills and developing resources for use in the classroom.

Seek Out the Help and Advice of Experts

Over the past 10 years, there has been a significant increase in the number of publications available to help college and university professors, scientists in particular, to become more scholarly about their teaching. Handelsman et al.'s excellent book, *Scientific Teaching* (2007), provides context, motivation, and resources for scientists interested in using assessment and the principles of backward course design to improve student learning in their own classrooms. The book *Classroom Assessment Techniques* (Angelo & Cross, 1993) as well as a book chapter on assessing learning in the POGIL chemistry classroom (Cole & Bauer, 2008) provide practical, off-the-shelf methods for collecting data about student learning. Although such books describe how to apply a scholarly mindset to teaching through the use of assessment, they do not provide guidance on how to conduct scholarship of teaching and learning that is intended to be broadly communicated. For this purpose, a number of books are available that walk beginners through the steps needed to intentionally develop projects that would be publishable as the scholarship of teaching and learning (McKinney, 2007; Savory et al., 2007). Finally, for those who are interested in learning more about chemical education research, the ACS recently published a symposium series volume devoted to the topic (Bunce & Cole, 2008a).

As interest in the scholarship of teaching and learning has grown, the number of experts in this area on campuses and at professional meetings has also increased. Many colleges and universities have teaching centers with staff trained to support faculty in improving their teaching. Tapping into the knowledge and experience offered by these individuals *before* implementing an assessment project is a powerful way to increase the likelihood that data collected will specifically address desired pedagogical and professional outcomes. If there is a possibility that classroom assessment results may be communicated outside the campus at which the study was conducted, additional planning including human subjects' approval will likely be required. The National Institutes of Health (NIH) offers free online training and certificates in use of human subjects in research (National Institutes of Health, 2009). In addition, most colleges and universities have personnel who are trained to assist faculty in these matters.

Summary

As chemists, we are fortunate to be part of a broad community that sees the value of using evidence to improve student learning in chemistry higher education. We are teaching in a time when more and more resources exist to support those interested in using assessment to improve student learning. Assessment can be a powerful window into the minds of our students. When taken from this perspective, well thought-out assessment projects that meet the needs of specific learning environments may be the most effective way for teachers to also continue to be life-long learners.

Acknowledgements

I sincerely thank my colleague and collaborator, Dr. Vicky Minderhout, of Seattle University for her intellectual and practical contributions to the work described in this chapter. Dr. Minderhout was involved in writing or revising nearly all of the POGIL biochem activities and participated in all described assessment efforts. Finally, my deepest gratitude to the hundreds of students and faculty members who have participated in assessment projects aimed at improving POGIL biochem activities.

REFERENCES

American Chemical Society Legislative and Government Affairs. (2010, April 26). *Scholarship.* Retrieved from http://www.divched.org/DivCHED/statement.pdf

Anderson, T. R. (2007a). Bridging the educational research-teaching practice gap. The importance of bridging the gap between science education research and its application in biochemistry teaching and learning: Barriers and strategies. *Biochemistry and Molecular Biology Education, 35*(6), 465–470.

Anderson, T. R. (2007b). Bridging the educational research-teaching practice gap. The power of assessment. *Biochemistry and Molecular Biology Education, 35*(6), 471–477.

Angelo, T. A., & Cross, P. K. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning*. Berkshire, England: Open University Press.

Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education, 63*(10), 873–878.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn*. Washington, DC: National Academy Press.

Bunce, D. M., & Cole, R. S. (2008a). *The nuts and bolts of chemical education*. Washington, DC: American Chemical Society.

Bunce, D. M., & Cole, R. S. (2008b). Using this book to find answers to chemical education questions. In D. M. Bunce & R. S. Cole (Eds.), *Nuts and bolts of chemical education research*. Washington, DC: American Chemical Society.

Cole, R. S., & Bauer, C. F. (2008). Assessing POGIL implementations. In R. S. Moog & J. N. Spencer (Eds.), *Process-oriented guided inquiry learning* (pp. 211–223). Washington, DC: American Chemical Society.

Dick, W., Carey, L., & Carey, J. (2005). *The systematic design of instruction* (6th ed.). Boston: Pearson.

Goffman, E. (1959). *The presentation of self in everyday life*. Woodstock, NY: Overlook Press.

Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York: W. H. Freeman and Company.

Hanson, D., & Moog, R. (2007). Process-oriented guided-inquiry learning. In S. W. Beyerlein, C. Holmes, & D. K. Apple (Eds.), *Faculty guidebook* (4th ed., pp. 387–390). Lisle, IL: Pacific Crest.

Hanson, D., & Wolfskill, T. (2000). Process workshops—A new model for instruction. *Journal of Chemical Education, 77*(1), 120–130.

Loertscher, J., & Minderhout, V. (2009). *Foundations of biochemistry*. Lisle, IL: Pacific Crest.

McKinney, K. (2007). *The scholarship of teaching and learning*. Bolton, MA: Anker Publishing.

Mettetal, G. (2001). The what, why and how of classroom action research. *The Journal of Scholarship of Teaching and Learning, 2*(1), 7–13.

Minderhout, V. (2007). Creating a facilitation plan. In S. W. Beyerlein, C. Holmes, & D. K. Apple (Eds.), *Faculty Guidebook* (4th ed., pp. 359–362). Lisle, IL: Pacific Crest.

Minderhout, V., & Loertscher, J. (2007). Lecture-free biochemistry. *Biochemistry and Molecular Biology Education, 35*(3), 172–180.

National Institutes of Health. (2009). *Research involving human subjects.* Retrieved from http://grants.nih.gov/grants/policy/hs/index.htm

National Research Council. (2002). *Scientific research in education*. Committee on Scientific Principles for Education Research. R. J. Shavelso & L. Towne (Eds.), Center for Education, Division of Behavioral and Social Sciences Education. Washington, DC: National Academy Press.

Process Oriented Guided Inquiry Learning (POGIL). (2009). Retrieved from www.pogil.org

Savory, P., Burnett, A. N., & Goodburn, A. (2007). *Inquiry into the college classroom*. Bolton, MA: Anker Publishing.

Spencer, J. N. (1999). New directions in teaching chemistry: A philosophical and pedagogical basis. *Journal of Chemical Education, 76*, 566–569.

Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478.

Walvoord, B. E. (2004). *Assessment clear and simple*. San Francisco: Jossey-Bass.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

CHAPTER 8

ACTIVE LEARNING IN THE CHEMISTRY CLASSROOM AT THE U.S. NAVAL ACADEMY

Daniel W. O'Sullivan and Christine L. Copper
U.S. Naval Academy

Learner-centered approaches to education have been studied extensively. Numerous books and journal articles have addressed the use of active learning techniques in all educational levels and in academic disciplines ranging from science and engineering to the humanities (Johnson, Johnson, & Smith, 1991; Michael & Modell, 2003; Silberman, 1996). Furthermore, the term *active learning* is used in many contexts in the education literature but is difficult to narrowly define. Our use of this term refers to classroom or laboratory experiences during which students do things and think about what they are doing as they do them. References specifically related to college chemistry instructors' use of active learning techniques include approaches such as cooperative learning, collaborative learning, guided inquiry, peer-led team learning, and problem-based learning (Duch, Groh, & Allen, 2001; Gosser, Strozak, & Cracolice, 2006; Moog & Farrell, 2008; Pienta, Cooper, & Greenbowe, 2004, 2009). The idea behind using any of these active learning teaching strategies is to engage students with course material and with each other versus having them sit silently and listen as a faculty member lectures to them about chemistry topics.

At our university, every freshman (approximately 1,100/year) must pass two semesters of general chemistry. These courses are taught by approximately 25 instructors each semester as section enrollment is maintained at 20 students. Attendance is mandatory to the 150 minutes of class and 110 minutes of lab time each week, all of which are taught by professors versus teaching assistants. The course sequence is coordinated by two or three faculty members at any one time. They provide materials to be used in all sections of the course such as a common syllabus, textbook, electronic homework, laboratory experiments, classroom demonstrations, and three exams per semester including a common final exam. The faculty can present the material in any manner of their choosing, but they must cover specific topics in time for each of the common exams.

Our students have many demands on their time including a high credit load, military training and obligations, and daily athletic activity. These demands often cause them to come to class exhausted and distracted, which makes for a poor audience for a traditional lecture-style teaching approach. In 1999, our department undertook an initiative aimed at introducing more active learning strategies into our general chemistry curriculum to offer the best possible education to our students. Many of the strategies employed in our active learning initiative involved students working in pairs and/or groups and included items such as problem-solving worksheets, creative testing strategies including group testing, hands-on learning activities such as manipulating chemicals and molecular models in the classroom, "explain the demo" worksheets allowing the students to draw their own conclusions about phenomena they were observing, student presentations, and even friendly competitions to review course material. The details of our efforts and an initial assessment of their effectiveness have been reported elsewhere (Copper & O'Sullivan, 2003). In that study, it was found that active learning strategies produced a statistically significant improvement in student performance during the first semester of the four that were studied (academic years 1999–2001). Furthermore, and perhaps most importantly, it was found that active learning approaches had no adverse effect on student performance during any of the semesters studied.

During the time frame of our initial study, the faculty members who were part of the active learning initiative were scheduled to teach in the only two classrooms that had tables at which students could work in groups. The remaining sections of the course were taught in rooms that had immovable furniture arranged in rows of stadium seating typical of a lecture hall. In 2004, we moved into a newly renovated building, and all of the classrooms have movable furniture, which allows for rows of seating or grouping of student seats into tables of four. To encourage faculty to continue to promote an active learning classroom environment, it was decided that the furniture in two of the new classrooms would be positioned such that there were five tables of four students, a group work arrangement (GRP), and the remainder of the rooms would be situated such that students were in rows that faced the front of the room, a traditional classroom setting (TRAD). At the time this decision was made, there was no plan in place to assess the influence of the furniture arrangement on student performance. Furthermore, there was no reason that faculty in the "GRP rooms" or the "TRAD rooms" could not quickly change the furniture arrangement to the other style. However, an informal survey of the faculty indicated that classroom furniture arrangement was rarely changed. This difference in the physical layout of our classrooms became the initial independent variable influencing student performance for this work, because it was an obvious difference in educational setting. However, since it was not guaranteed that the GRP rooms were being used only for active learning and the TRAD rooms only for lecturing, we also examined the performance of students receiving instruction from professors who self-characterized their teaching style as belonging to one of three types:

- LEC: Majority lecture, less than 10% of the classroom time spent in active or group learning activities.
- HYB: Some lecture, between 10% and 50% of the classroom time spent in active or group learning activities.
- ACT: Minimal lecture, greater than 50% of the classroom time spent in active or group learning activities.

Methods

Assessment Instruments

The initial experimental approach in this work was to compare the performance of the students in GRP rooms ($n \approx 200$ per semester) to the TRAD room students ($n \approx 800$ per semester), which act as the control group in the room configuration portion of this study. Presumably, the TRAD group of students experienced a more traditional, primarily lecture-based general chemistry course, whereas those students in the GRP rooms would experience a more active class environment. However the instructors teaching in the TRAD rooms were not required to lecture exclusively. To avoid adversely affecting the students' education in the TRAD rooms, faculty instructing those sections could conduct their classes in any way that they desired. Occasionally, these faculty members employed active-learning-type activities in their classes, although most of the class time was lecture-based.

Additionally, the faculty members were asked to self-assess the teaching environment irrespective of the classroom design. The faculty identified their lecture style as one of the three types above, and the student outcomes were compared for the students experiencing the two most different teaching styles (LEC and ACT). During each semester studied, approximately 400 students had a LEC instructor while about 100 students per semester had ACT instructors.

Grades calculated at 16 weeks into the semester, common final exam grades, and course grades were used as the dependent variables in the study to compare student performance in the GRP room sections relative to the students experiencing a more traditional general chemistry instruction environment, the TRAD students. These same three measures were used to examine student performance based on the instructor type that the faculty self-assessed. The 16-week grade is a measure of all the graded work in the course (such as quizzes, exams, lab reports and homework assignments) except the final exam. The final exams were given simultaneously to all students enrolled in the courses. These two items are believed to be measures of individual understanding of the course material. A number of faculty members have observed that many students do not perform well on the final exam as a result of the belief that the students' course letter grade is largely determined prior to the final. Consequently, both the 16-week grade and the performance on the final exam were examined. The course grade was also used; however, this measure is potentially more subjective as it is dependent on instructor input rather than just individual student effort.

Verbal and math scores on the SAT were used to determine if there was a significant difference in initial ability for the various populations. There have been a number of studies that have shown a strong direct relationship between math SAT scores and performance in introductory chemistry courses (Andrews & Andrews, 1979; Bunce & Hutchinson, 1993; Glover, Kolb, & Taylor, 1991; Spencer, 1996).

Results

Classroom Setting

The verbal and math SAT, 16-week grade, final exam scores, and course grades of students in the two types of rooms, GRP and TRAD, each semester are presented in Table 1. Since enrollment in the different sections was determined by the registrar, we used the verbal and math SAT scores to determine if there was a significant difference in the initial abilities of the students in the two populations. It is apparent from the verbal and math SAT score averages that the two populations are very similar. In order to assess whether the performance or initial abilities in the two populations of students were statistically different, a two-tailed unpaired *t*-test was used. This test was used to determine the confidence level at which two means were statistically different when the number of observations determining the means is different for each group. The number of students in each group was different each semester (see Table 1). A statistical comparison of means between each group for the SAT scores is presented in the SAT column of Table 2. A multiway analysis of variance (ANOVA) was performed with the math SAT, verbal SAT, and room configuration as factors

Table 1

Performance Metrics for Students Experiencing GRP and TRAD Classroom Settings

		Number of Students	SAT		16-week		Final Exam		Course	
		n	verbal	math	x	sd	x	sd	x	sd
Fall 2005	**GRP**	256	661	635	2.30	0.95	1.82	1.12	2.22	1.01
	TRAD	798	669	647	2.33	0.97	1.97	1.15	2.27	0.99
Spring 2005	**GRP**	382	671	643	2.14	1.00	2.08	1.15	2.24	0.96
	TRAD	628	669	646	2.16	0.99	2.13	1.10	2.24	0.96
Fall 2006	**GRP**	344	673	645	2.58	0.94	2.18	1.20	2.53	0.96
	TRAD	695	657	637	2.30	0.97	1.85	1.15	2.26	0.92
Spring 2006	**GRP**	247	666	644	2.28	0.93	2.18	1.06	2.28	0.91
	TRAD	760	664	639	2.17	0.98	1.91	1.14	2.19	1.00
Fall 2007	**GRP**	236	653	632	2.41	0.93	2.01	1.13	2.31	0.99
	TRAD	804	654	630	2.48	0.97	2.16	1.08	2.45	1.00
Spring 2007	**GRP**	254	645	622	2.21	0.98	1.78	1.14	2.19	0.96
	TRAD	758	660	635	2.30	1.01	2.09	1.17	2.31	1.01
Fall 2008	**GRP**	264	647	626	2.41	1.01	2.04	1.19	2.35	1.04
	TRAD	752	652	630	2.43	0.94	2.06	1.14	2.38	0.98
Spring 2008	**GRP**	249	654	636	2.30	0.85	2.05	1.17	2.31	0.92
	TRAD	732	652	629	2.25	0.94	1.97	1.12	2.26	0.94
Fall 2009	**GRP**	257	655	634	2.82	0.86	2.11	1.16	2.64	0.93
	TRAD	820	655	634	2.49	0.95	1.97	1.10	2.40	0.99
Spring 2009	**GRP**	254	649	626	2.52	0.93	2.16	1.11	2.49	0.91
	TRAD	792	657	636	2.25	0.99	2.04	1.15	2.28	0.95

for each performance metric, and the results are presented in Table 2. As analytical chemists, we generally express the difference between two means from a *t*-test as a confidence level. If the *t*-statistic value exceeds 1.17, 1.29, or 1.65, the differences are significant at the 75%, 90%, or 95% confidence levels, respectively. A positive sign of the *t*-statistic value would indicate that the GRP room students' performance exceeded the TRAD room students' performance on the metric. Conversely, if the *t*-statistic has a negative sign, then the TRAD room students' performance exceeded the GRP room students' performance on the metric. In the behavioral sciences, a confidence level (alpha) of 95% is used to indicate a statistically significant difference, $p < 0.05$. If the confidence-level criterion is relaxed to 75%, the corresponding p value would be < 0.25.

Over the five-year period of this study, the performance of more than 5,000 students in the two introductory chemistry courses was examined. The 16-week, final

Table 2

Statistical Data Comparison Results for GRP and TRAD Classroom Students

	ANOVA				*t*-test	
	Data type	**16-week**	**Final**	**Course**	**Math SAT**	**Verbal SAT**
Fall 2005	p[1]	0.42	0.34	0.76	0.03	0.08
	t[2]	-0.40	-1.86	-0.67	-2.17	-1.78
Spring 2005	p	0.51	0.25	0.93	0.44	0.60
	t	-0.41	-0.64	0.07	-0.77	0.52
Fall 2006	p	0.01	0.01	0.04	0.14	0.00
	t	4.46	4.30	4.01	1.48	3.37
Spring 2006	p	0.62	0.02	0.84	0.42	0.71
	t	1.58	3.25	1.29	0.81	0.38
Fall 2007	p	0.38	0.03	0.06	0.72	0.79
	t	-1.00	-1.76	-1.81	0.35	-0.26
Spring 2007	p	0.97	0.03	0.72	0.02	0.00
	t	-1.17	-3.64	-1.72	-2.32	-3.16
Fall 2008	p	0.64	0.70	0.87	0.35	0.28
	t	-0.26	-0.20	-0.45	-0.94	-1.09
Spring 2008	p	0.48	0.29	0.44	0.20	0.75
	t	0.66	0.94	0.74	1.29	0.31
Fall 2009	p	0.00	0.06	0.00	0.99	0.96
	t	4.96	2.13	3.50	0.02	-0.04
Spring 2009	p	0.00	0.04	0.00	0.08	0.13
	t	3.75	1.53	3.10	-1.73	-1.53

[1]The *p*-values in the ANOVA columns are based on a multiway ANOVA computation; the *p*-values in the *t*-test columns are from a simple two-tailed student's *t*-test computation.
[2] The values for *t* are all from a simple two-tailed student's *t*-test computation.

exam, and course grades of these students in the two room configurations are shown in Figure 1. For 3 of the 10 semesters, the course grades of the students experiencing a GRP classroom setting exceeded that of the TRAD room (control group) students by 0.24 units on a 4.0 scale (see Table 1). The enhanced performance is significant at the 95% confidence level for each of these semesters. There are no semesters in which the TRAD room students' course grades exceed those of the GRP room students on the course-grade instrument at the 95% confidence level (see Table 2). There are a number of semesters during the study that the performance measured with the course-grade metric is nearly identical for students experiencing each type of classroom setting and where the differences are not statistically significant at the 95% confidence level (see Table 2).

The semesters with statistically significant differences in the course-grade metric also exhibit similar differences in other metrics such as the 16-week grade, the final exam grade or the verbal SAT score (see Table 2). The multiway ANOVA analysis for each metric, 16-week grade, final exam grade, and course grade, showed that math SAT and verbal SAT scores were statistically significant factors for each group every semester (*p*-values not shown). The *p*-value in Table 2 for each metric shown in the ANOVA section represents whether the room configuration was a significant factor for the metric. The room configuration was a significant factor for the 16-week grade metric in 3 of the 10 semesters, and in 5 of 10 semesters for the final exam grade metric, and in 3 of 10 semesters for the course grade.

In the Fall of 2005 and the Spring of 2007, the GRP room students' average math SAT score is significantly below the TRAD group's, yet the course performance of these students is not below that of the TRAD group. There are only two instances in which the TRAD students outperformed the GRP room students, which were in the Fall and Spring of 2007 when the TRAD group outperformed the GRP room students on the final exam. Verbal and math SAT scores for the TRAD group in the Spring of 2007 are higher than the GRP students, indicating the TRAD group of students was advantaged from the outset. These results indicate that room configuration was a significant factor influencing student performance on a number of metrics independent of factors such as SAT performance, which are known to influence student outcomes in general chemistry.

It was found that students studying general chemistry in a GRP classroom setting conducive to active learning educational approaches performed well on individual student assessment metrics and in the course as a whole. There is no evidence that this type of classroom environment adversely impacted the students' proficiency in chemistry. Since not every instructor assigned to a GRP room performed group work, and instructors assigned to TRAD rooms did not necessarily lecture all the time, we examined the performance of students based on the instructors' self-assessment of the teaching environment in their classes regardless of the desk configuration in the rooms.

Self-Assessed Instructor Type

The instructors were asked to categorize their lecture style. The three choices involved a self-evaluation of the relative balance of class time spent in a lecture format relative to class time spent in an active learning situation. The greatest contrast in

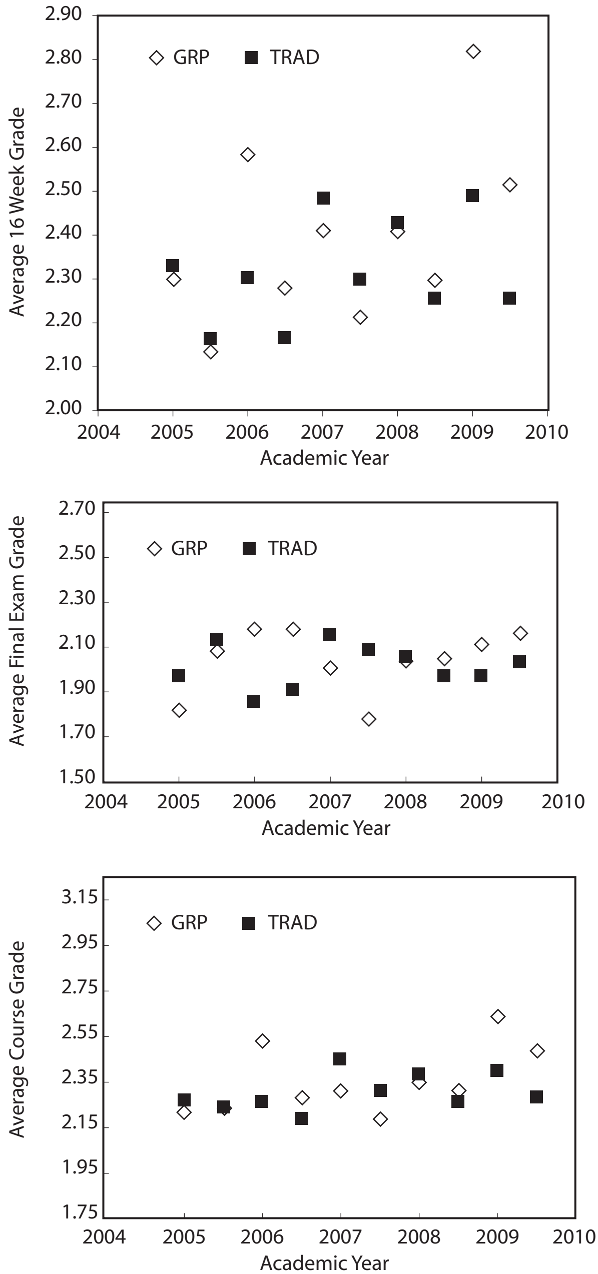


Figure 1. Performance of students with classes in the GRP rooms compared to students who had classes in TRAD rooms from Fall 2005 to Spring 2009. Top panel shows the average 16-week grades, center panel shows the average final exam grade, and the bottom panel shows the average course grade.

lecture environment was between the self-assessment of the lecture environment as either LEC, primarily a lecture class, and ACT, where greater than 50% of the class time consisted of active learning activities—thus minimal lecture. The performance of the students experiencing these two different types of lecture environments was examined using the same metrics used in the room configuration study. Since the number of faculty teaching the courses each semester changes as do their teaching approaches, the number of students in the two study groups changed from semester to semester, as is shown in Table 3.

Typically, there were about 100 students in the ACT group and about 400 students in the LEC group in any given semester (see Table 3). The balance of the approximately 1,100 students in a general chemistry course each semester experienced an instructional environment self-assessed as HYB (10–50% active or group learning activities). Both a multiway ANOVA analysis and *t*-tests were used to discern statistically significant differences between groups. For the *t*-test calculations, we focused on a comparison between LEC and ACT lecture environments, since these represent the greatest difference in the lecture experience for the students. However, the number of students experiencing an ACT-lecture environment is relatively small, thus adversely impacting a robust statistical evaluation of the differences.

The 16-week, final exam, and course grades for each group in each semester are plotted in Figure 2. Comparison of the student performance sorted on teaching style relative to sorting by classroom furniture arrangement (Figure 1) show a greater consistency of better performance on these metrics with ACT classroom environments (Figure 2). In 9 of the 10 semesters examined, the ACT group performance is better than the LEC group for the 16-week grade metric. In 5 of the 10 semesters, ACT performance exceeds the LEC performance on the final exam, and in 7 of 10 semesters, the course grades of the ACT students exceed those of the LEC students. With respect to course grades, the average difference between the groups is +0.09 grade units on a 4.0 scale. Considering only the semesters where the ACT students' performance was better than the LEC group, the average impact on the grade point average is +0.15 units, as shown in Table 3.

Of the metrics examined, there are only three instances where the performance difference between these two groups of students is statistically significant ($p < 0.05$), and all three involve the 16-week grade metric. The performance of the students experiencing an ACT lecture environment exceeded that of the LEC students in the Spring of 2005 and 2007 and the Fall of 2009 at the 95% confidence level, *t*-statistic > 1.65 (see Table 4). If the confidence level is relaxed to 90% (*t*-statistic > 1.29), the ACT group performance is better on the 16-week grade metric in 6 of 10 semesters, on the final exam in 1 of 10 semesters, and in the course grade in 3 of 10 semesters (see Table 4). As in the room configuration study, a multiway ANOVA analysis was performed for each metric with math and verbal SAT scores and teaching type as factors. Verbal and math SAT scores were statistically significant factors for all metrics in all semesters ($p < 0.05$, data not shown). Teaching type was found to be a significant factor for the 16-week grade in 6 of 10 semesters, and for the final

Table 3

Performance Metrics for the Students Experiencing LEC and ACT Instructors

		Number of Students	SAT		16-week		Final Exam		Course	
		n	verbal	math	x	sd	x	sd	x	sd
Fall 2005	LEC	444	671	647	2.27	0.97	2.00	1.10	2.25	0.99
	ACT	95	664	639	2.48	0.93	2.01	1.09	2.39	0.98
Spring 2005	LEC	464	666	641	2.12	1.07	2.11	1.07	2.23	0.94
	ACT	39	680	647	2.49	0.85	2.46	1.14	2.51	0.88
Fall 2006	LEC	526	662	642	2.34	0.99	1.97	1.17	2.33	1.03
	ACT	115	661	625	2.48	1.04	2.05	1.24	2.43	1.09
Spring 2006	LEC	534	668	643	2.13	1.02	2.07	1.11	2.22	1.00
	ACT	112	658	626	2.42	0.89	2.16	1.10	2.41	0.94
Fall 2007	LEC	431	650	628	2.39	1.00	2.14	1.12	2.41	1.00
	ACT	137	655	635	2.56	0.90	2.18	1.11	2.49	0.92
Spring 2007	LEC	404	657	633	2.28	1.04	2.07	1.22	2.34	1.04
	ACT	100	651	625	2.27	0.95	1.91	1.04	2.24	0.90
Fall 2008	LEC	331	653	628	2.35	0.98	1.96	1.14	2.33	1.01
	ACT	123	630	619	2.36	1.02	1.92	1.10	2.23	1.01
Spring 2008	LEC	386	655	635	2.18	0.97	2.06	1.11	2.26	0.97
	ACT	91	649	625	2.23	0.83	2.03	1.17	2.30	0.84
Fall 2009	LEC	473	657	634	2.50	0.96	2.00	1.09	2.42	1.00
	ACT	40	643	622	2.83	0.94	2.13	1.09	2.68	0.97
Spring 2009	LEC	545	661	635	2.27	1.02	2.12	1.14	2.34	0.98
	ACT	78	657	637	2.37	0.85	2.05	1.10	2.32	0.86

exam in 3 of 10 semesters, but for only one semester for the course grade metric. In only three instances did the LEC group exceed the ACT group, and only one of those was a performance metric. The LEC group exceeded the ACT group on the final exam metric in the Fall of 2008. The other two instances where the LEC group exceeded the ACT group were for the average math SAT in Fall of 2006 and the average verbal SAT in the Fall of 2008. In these semesters, the LEC pool of students was better positioned going into the course than the ACT pool, yet the performance of the ACT pool of students in both of these semesters is indistinguishable from the LEC pool on all but one course metric. These results indicate a positive impact of an active learning environment on student performance in the general chemistry course experience when the active learning component is a significant portion (at least 50%) of the classroom time. There are no instances of an adverse outcome on any student performance metric, including metrics which are predominantly based on an individual's capability, such as the final exam grade.

To determine if the active learning classroom environment affected one pool of students over another, the course performance metrics for each instructor type were examined after sorting the students based on their math SAT scores. Math SAT scores

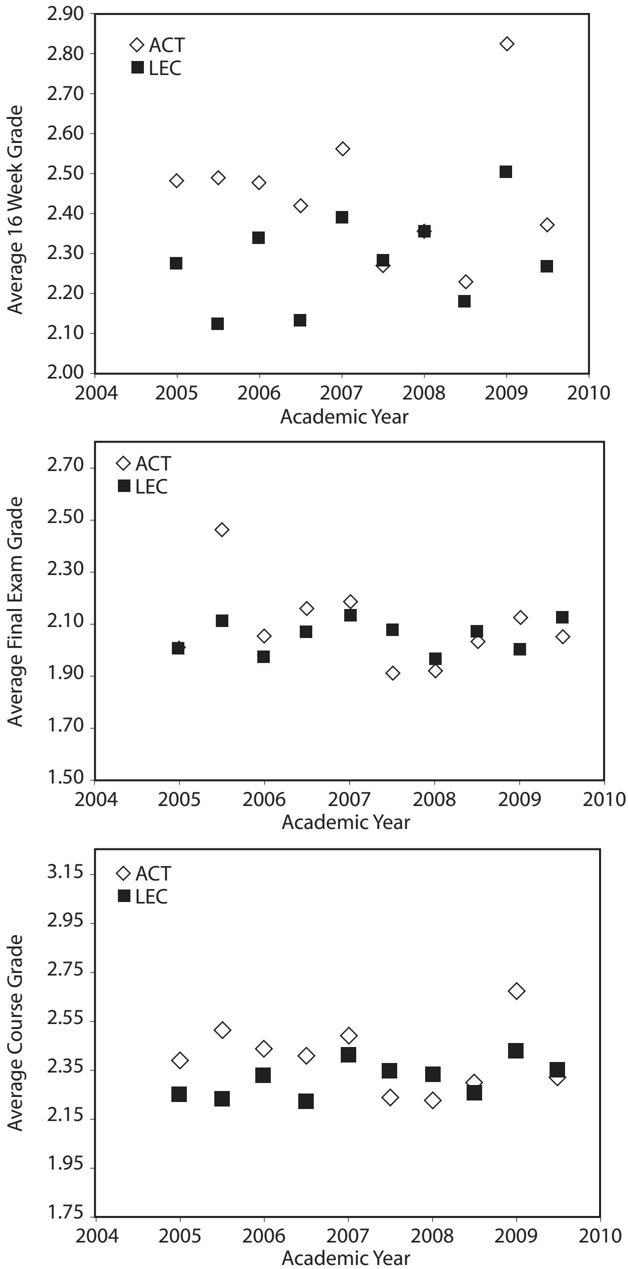


Figure 2. Performance of students with LEC instructors compared to students who had classes with ACT instructors from Fall 2005 to Spring 2009. Top panel shows the average 16-week grades, center panel shows the average final exam grade, and the bottom panel shows the average course grade.

were used since there have been several studies indicating that a student's math SAT score can be a predictor for performance in general chemistry (Andrews & Andrews, 1979; Bunce & Hutchinson, 1993; Glover et al., 1991; Spencer, 1996). Students in each type were sorted into seven bins by math SAT score starting at scores < 500. All subsequent bins were 50 units in math SAT score wide increasing by 50 with each bin up to the last bin at 800. The number of students in each bin for each type of instructor is shown in Figure 3. The distribution of students as a function of math SAT is similar in both groups, but the number of students is, of course, much larger in the LEC group. The greatest deviations in relative percentage of students with a given math SAT score bin occurs with more LEC students in the 650 to 700 math SAT bin and a slightly higher relative number of students in the ACT pool with math SAT scores in the 550 to 600 math SAT bin.

Table 4
Statistical Comparison of Metric for the LEC and ACT Lecture Environment Students

	ANOVA				*t*-test	
	Data type	**16-week**	**Final**	**Course**	**Math SAT**	**Verbal SAT**
Fall 2005	p[1]	0.01	0.13	0.12	0.34	0.39
	t[2]	1.92	-0.05	-1.25	0.96	0.87
Spring 2005	p	0.02	0.04	0.07	0.62	0.20
	t	2.29	1.96	1.80	0.49	1.27
Fall 2006	p	0.08	0.34	0.39	0.04	0.90
	t	1.38	0.65	1.01	-2.04	-0.13
Spring 2006	p	0.01	0.00	0.03	0.05	0.17
	t	2.76	0.78	1.86	-1.95	-1.38
Fall 2007	p	0.33	0.09	0.25	0.36	0.45
	t	1.80	0.42	0.81	0.92	0.75
Spring 2007	p	0.97	0.36	0.32	0.32	0.41
	t	-0.09	-1.24	-0.92	-1.00	-0.82
Fall 2008	p	0.13	0.02	0.23	0.22	0.00
	t	0.04	-0.33	-0.95	-1.24	-3.14
Spring 2008	p	0.01	0.40	0.41	0.21	0.42
	t	0.47	-0.24	0.36	-1.26	-0.82
Fall 2009	p	0.00	0.44	0.06	0.38	0.24
	t	2.06	0.70	1.52	-0.87	-1.18
Spring 2009	p	0.04	0.91	0.54	0.92	0.67
	t	0.84	-0.52	-0.18	0.10	-0.43

[1]The *p*-values in the ANOVA columns are based on a multiway ANOVA computation, the *p*-values in the *t*-test columns are from a simple two-tailed student's *t*-test computation.
[2]The values for *t* are all from a simple two-tailed student's *t*-test computation.

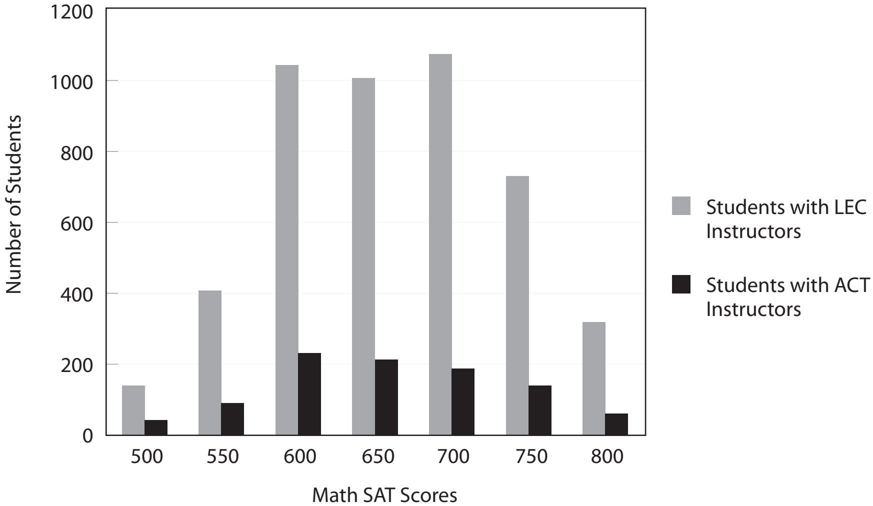


Figure 3. Histogram of the number of students in each math SAT bin for different instructor types.

Figure 4 shows that the percentage of students achieving an A for a course grade increased with increasing math SAT score, and there was very little difference between the LEC and ACT groups' performance. A similar trend is observed for the percentage of students receiving a course grade of B. The percentage of students increased from around 15% B in the lower math SAT score bins to over 30% with math SAT scores over 700. In nearly all of the math SAT score bins, a greater percentage of students from the ACT lecture environment achieved a grade of B. The most pronounced difference occurred for the less-than-500 math SAT bin, where 23.8% of the students in the ACT environment received a B compared to 14.5% in the LEC environment. The trends observed for D and F grades have the opposite slope of those observed for A and B grades for both lecture types. The percentage receiving F grades are nearly identical in both groups, particularly when one considers that the number of students receiving failing grades is quite small. The differences are not robust. For the ACT cohort with math SAT scores below 500, only 6 students received an F, 14% of the group. For LEC students in this math SAT range, 31 students received an F, 22.5% of the group. Although the percentage difference is relatively large, the number of students in the cohort is quite small. Students receiving Cs in both lecture-type environments exhibit similar distributions. A lower percentage of students with high math SAT and low math SAT scores received grades of C in the ACT environment. On the upper end, this is largely a result of more students receiving Bs. However on the lower end, it appears more students experiencing the ACT environment with math SATs in the 550 bin and less-than-500 bin may have received lower grades than the comparable group experiencing the LEC environment.

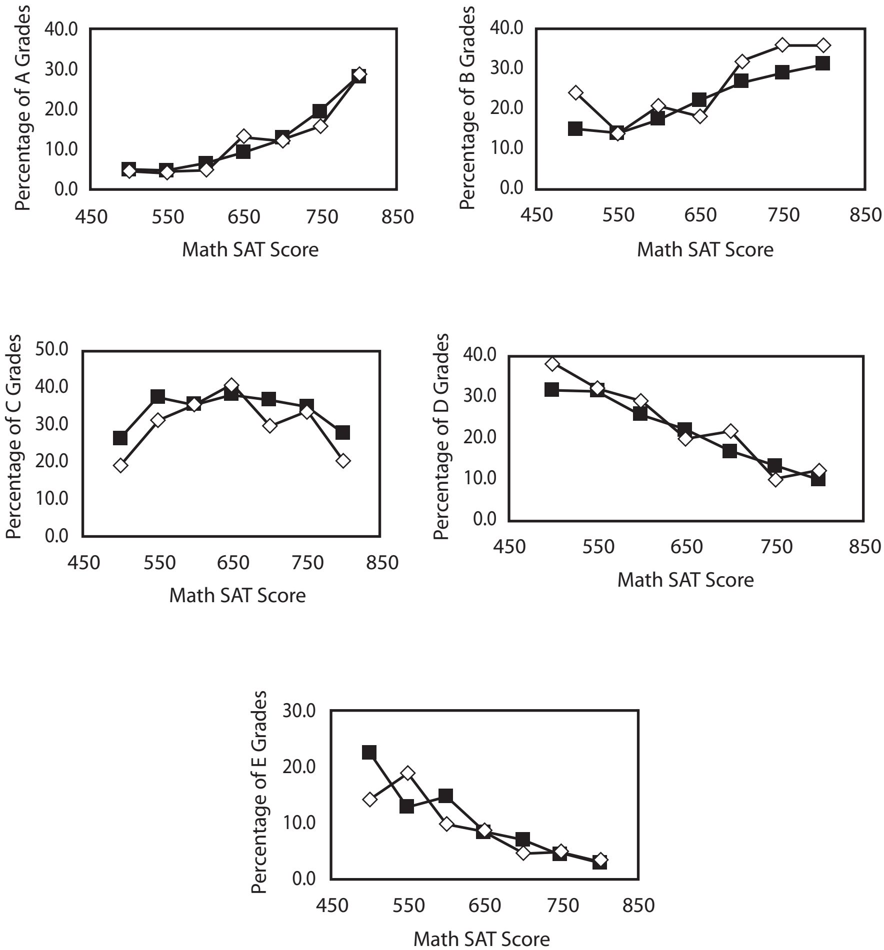


Figure 4. Math SAT scores and performance on the final exam with LEC (dark squares) and ACT (open diamonds) instructors.

Conclusions

In this study, the general chemistry performance of over 5,000 students over five years was examined based on two different criteria. First, classroom geometry, and hence the ability for the instructor to easily implement group work and active learning exercises was considered. Secondly, student performance was examined based on the instructor self-assessment of the classroom environment as primarily lecture, < 10% of the class time involved in active learning activities, or as an active learning experience, where > 50% of the classroom time involved students "doing" rather than listening.

In the classroom geometry assessment for 3 of the 10 semesters, the course grades earned by students experiencing a GRP classroom setting exceeded those of the TRAD students by 0.24 units on a 4.0 scale. The difference in course grade is significant at the 95% confidence level for each of these semesters. In this study, there were no semesters in which the TRAD students' course grades exceeded the GRP students on the course grade instrument, and that is statistically significant.

Examining the performance of students based on the teaching style they experienced demonstrated a positive outcome for students experiencing an active learning lecture environment in general chemistry. Comparison of the student performance sorted on teaching style relative to sorting by classroom furniture arrangement demonstrated consistently larger differences between groups measured using a number of metrics including the course grades, final exam grade, and the grade at 16 weeks. In 9 of the 10 semesters examined, the ACT group performance was better than the LEC group for the 16-week grade metric. In 5 of the 10 semesters, ACT performance exceeded the LEC performance on the final exam, and in 7 of 10 semesters the course grades of the ACT students exceeded those of the LEC students. Considering only the semesters where the ACT students' performance was better than the LEC group's, the average impact on the grade-point average is +0.15 units. This study provides some evidence of a positive impact of an active learning environment on student performance in general chemistry, when the active learning component is a significant portion (at least 50%) of the classroom time. There are no instances of an adverse outcome on any student performance metric, including metrics which are predominantly based on an individual's capability, such as the final exam grade.

In short, with a large pool of students and a diverse professional instructor pool, the outcomes are predominantly positive for students experiencing an active learning general chemistry course. Concerns expressed by faculty regarding the adverse impact on student performance due to less material coverage as a result of the time required to implement active learning activities in the classroom are not supported by these data.

The work reported herein focused on the relationship between the physical layout of the classroom (or the teaching style of the instructor) and student performance on graded work. One important variable that was not addressed in this work is that of teacher and student perceptions of the classroom environment. Future efforts to assess our active learning initiative will include the use of the College and University Classroom Environment Inventory (CUCEI) survey, which allows one to gather data in areas including instructor efforts to relate to students and to try new teaching techniques like active learning and student perceptions of their interactions with the instructor and each other in order to learn the course material. Details about the CUCEI and results from its use are described in a text by a leading classroom environment researcher (Fraser, 1986).

Lessons Learned

A unique characteristic that our department has when compared to many others is the sheer number of faculty members (approximately 40), all of whom spend a significant part of their career teaching the general chemistry course. The assessment of room configuration and teaching type presented in this chapter was possible as a result of a large number of our faculty members' willingness to utilize active learning in their classrooms. The introduction of active learning approaches in the general chemistry curriculum began a decade ago with a small group of faculty in the department. Having an initial group of about 10 faculty work to develop and test active learning materials for both the classroom and laboratory settings allowed for a division of labor and a compilation of ideas. We believe this cooperative effort was essential, as it allowed faculty, even in the first semester of the project, to use many active learning items in their own classes without the pressure of designing them all alone. It also created a situation in which enough of the students taking general chemistry were experiencing an active learning environment to allow for statistically legitimate assessment of the efforts and a large enough population of students in the "experimental group" such that they did not even realize that an experiment was taking place. If students realize that their situation is much different than that of other students, they may be skeptical of it. Furthermore, undertaking an initiative such as the one described in this work without a robust assessment possibility will make it hard to determine the success or failure of the effort.

We also learned that once the positive experiences of the professors in the initial active learning working group became known to others in the department, they wanted to join the group. This allowed for faculty to rotate into and out of the group, thus more ideas came to the group and more faculty implemented active learning in their classes. Furthermore, although we have not done a complete analysis of the student comments on course and instructor evaluation forms, we can report that many of our students have positively commented on the active learning activities when they have evaluated the class and/or their instructor.

REFERENCES

Andrews, M. H., & Andrews, L. (1979). First-year chemistry grades and SAT math scores. *Journal of Chemical Education, 56,* 231–232.

Bunce, D. M., & Hutchinson, K. D. (1993). The use of GALT as a predictor of academic success in college chemistry. *Journal of Chemical Education, 70*, 183–187.

Copper, C. L., & O'Sullivan, D. W. (2003). Evaluating active learning. *Journal of College Science Teaching, 32*, 448–452.

Duch, B. J., Groh, S. E., & Allen, D. E. (2001). *The power of problem-based learning*. Sterling, VA: Stylus Publishing.

Fraser, B. J. (1986). *Classroom environment*. London, England: Croom Helm.

Glover, D., Kolb, D., & Taylor, M. (1991). Another option for chemistry dropouts. *Journal of Chemical Education, 68*, 762–763.

Gosser, D. K., Strozak, V. S., & Cracolice, M. S. (2006). *Peer-led team learning: General Chemistry* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Active learning: Cooperation in the college classroom.* Edina, MN: Interaction Book Company.

Michael, J. A., & Modell, H. I. (2003). *Active learning in secondary and college science classrooms.* Mahwah, NJ: Lawrence Erlbaum Associates.

Moog, R. S., & Farrell, J. J. (2008). *Chemistry: A guided inquiry* (4th ed.). New York: John Wiley and Sons.

Pienta, N. J., Cooper, M. M., & Greenbowe, T. J. (2004). *Chemists guide to effective teaching* (Vol. 1). Upper Saddle River, NJ: Prentice Hall.

Pienta, N. J., Cooper, M. M., & Greenbowe, T. J. (2009). *Chemists' guide to effective teaching* (Vol. 2). Upper Saddle River, NJ: Prentice Hall.

Silberman, M. (1996). *Active learning: 101 strategies to teach any subject*. Needham Heights, MA: Allyn and Bacon.

Spencer, H. E. (1996). Mathematical SAT scores and college chemistry grades. *Journal of Chemical Education, 73,* 1150–1153.

CHAPTER 9

FROM COURSE REDESIGN TO CURRICULAR REVIEW: ASSESSMENT IN CHEMISTRY AT THE UNIVERSITY OF IOWA

Norbert J. Pienta
University of Iowa

Introduction

The desire to document or measure the knowledge, skills, and attitudes of chemistry students can be motivated by different events or needs. Two such circumstances and their outcomes in chemistry at the University of Iowa were a general chemistry course-sequence redesign and the evaluation of the departmental curriculum in anticipation of reaccreditation. A description of the evaluative process for each could focus on a theory-driven plan, starting with first principles. The current knowledge about assessment is sufficiently advanced that the projects described here could have started a priori; instead, the approach is more phenomenological.

The course redesign and curricular review provide two assessment case studies. The first example documents the process by which a traditional, large-enrollment introductory-chemistry sequence underwent changes, implemented to address student dissatisfaction, unacceptable success levels, and demands from other programs that used these courses to fulfill their degree requirements. Demonstrating success required both qualitative and quantitative measures, the latter apropos to the discerning scrutiny of a faculty group made up of scientists. The outcome of the redesign was measureable, sustained, and transformative—student satisfaction and success increased as did the approval of constituencies who required the courses.

The assessment plan for chemistry's undergraduate curriculum was motivated by an institutional reaccreditation, potential changes required for degree accreditation by chemistry's professional organization, and the turnover of a substantial number of faculty in the department. Each of those factors provided different timelines, motivation, and expectations. Originally skeptical of the need for a curricular assessment, the faculty eventually accepted its desirability. The success in the course redesign aided in the buy-in of the latter venture, and the faculty ultimately produced an exemplary model.

Case 1: Assessment of the General Chemistry Course Redesign

Background: Traditional Courses and Common Problems

An instructional model, common in chemistry but also appearing frequently in other science disciplines, involves large-enrollment lecture courses coupled with discussion or recitation sessions and a hands-on laboratory experience. Lectures are convened in large auditoria, recitation or discussion takes place in small sections run by teaching assistants (TAs) or in even smaller settings in a help or resource center, and laboratory experiments are also performed under the watchful eyes of TAs. Chemistry at the University of Iowa is no different—large lectures (300–400 students) for three hours per week and discussion sections of 20–24 for one hour per week, utilizing graduate TAs. Until the redesign in 2002, this defined each of two lecture courses, Principles of Chemistry I–II for which students earned three semester hours (s.h.). The old laboratory was a separate two-s.h. course with one hour of lab lecture and a three-hour laboratory each week. Old Principles II was the pre- or corequisite for the laboratory. The old system is summarized in Table 1, while the new courses are outlined in Table 2. The total number of credit hours and contact hours were maintained in the transition.

Table 1

General Chemistry Courses, Pre-2002

course	*old Principles I*	*old Principles I*	*old Principles Lab*
sem hrs	3	3	2
sessions	3 @ 50 min (lecture) 1 @ 50 min (discussion)	3 @ 50 min (lecture) 1 @ 50 min (discussion)	1 @ 50 min (lab lecture) 1 @ 170 min (lab)

Table 2

General Chemistry Courses, 2002 and After

course	*new Principles I*	*new Principles II*
sem hrs	4	4
sessions	3 @ 50 min (lecture) 1 @ 50 min (discussion) 1 @ 90 min (lab session I) or 1 @ 150 min (lab session II)	3 @ 50 min (lecture) 1 @ 50 min (discussion) 1 @ 90 min (lab session I) or 1 @ 150 min (lab session II)

Some instructors (or entire general chemistry programs) use grades or average GPA in a course to demonstrate rigor. Many have heard about (or even experienced) the well-intentioned pronouncement during the first lecture, "Look at the person sitting to your left and to your right. One of them will not be here next semester." The intention is that the threat or thought of failure in the class would motivate the students to a higher level of success. Perhaps this conduct has its origins in behaviorism and the ideas of behavioral conditioning that accompanied that psychological and learning theory (Merriam & Cunningham, 1989). That is not to say that standards and avoiding grade inflation are bad. But student perceptions of the instructors' attitudes become part of the tradition of a course. Such was the case at Iowa, and by the middle 1990s the Iowa Principles of Chemistry sequence, particularly Principles I, had achieved the status of a "weedout" course. This special student designation certainly meant that the course was very challenging, but additionally it suggested that these courses were used to reduce enrollments and eliminate students. Somehow, the perception was that faculty were encouraged or perhaps even rewarded to lower enrollments. In some years prior to the redesign, the rate of D, F, and W grades was as high as 35% in the Principles I course, and this certainly contributed to the reputation. The grade of "W" is a withdrawal from the course after the first few weeks but before the end, and "DFW rate" appears in a later section of this chapter as one measure of student success.

Chemistry departments teach several courses that are required by science, engineering, and preprofessional majors in addition to courses that fulfill general education science requirements for business, humanities, and social science majors. All of these programs are faced with trying to balance the prerequisite and background needs of their majors with the courses specific to their own discipline. In some instances, the needs of academic programs are also driven by professional certification or their own programmatic accreditation. Such is the case for engineering, pharmacy, medicine, dentistry, and nursing, all of which require chemistry courses. Engineering reaccreditation also played a role in the chemistry redesign. In 1997, ABET, formerly known as the Accreditation Board for Engineering and Technology, adopted Engineering Criteria 2000 or EC2000, a plan that focuses on what is learned rather than what is taught (ABET, 2009). In response, the Iowa College of Engineering sought to change its chemistry requirements, both in terms of the number of required courses and also the course content and pedagogy. The cycle for a Pharmacy review was not far behind engineering, and action was required.

Formulating a Plan: Practical Needs vs. Model Pedagogy

The current enrollment at the University of Iowa (ca. 30,000 total students, of which about 21,000 are undergraduates) has remained steady over the last decade, the period discussed here. Most undergraduates originate from Iowa (ca. 60%) or surrounding states (25%). Until recently, the university offered admission to State of Iowa students in the top half of their high school class. Of the students currently enrolled in high school in Iowa, about 65% take a chemistry course; the same group will graduate with an average of three mathematics and three science courses (Iowa Department of Education, 2008). The university's College of Liberal Arts and Sciences, where a large majority of students first enroll, requires three mathematics and three science courses (University of Iowa, 2008). Of the about 4,400 students entering the University of Iowa each fall, about 1,450 enroll in Principles of Chemistry I. Given the historic performance of Iowa high school students on standardized testing and on college admissions tests, student failure rates of 35% in the old Principles of Chemistry I course was problematic.

The management of student expectations and grades may be a sufficient reason to consider a course redesign but was only part of the Iowa plan. Although practical for some administrative reasons, a separate laboratory course did not seem to have much pedagogical value. Integrating the laboratory and lecture portions of the Principles sequence would provide the opportunity to replace confirmation exercises with guided-inquiry experiments, would enable most engineering majors to take a single four-s.h. course with lecture and lab combined, and would support other best practices in the integrated courses. The widespread use and general ineffectiveness of "cookbook" laboratory experiments has been documented (Abraham et al., 1997). In their study, a survey of introductory courses at a large number of institutions, Abraham and coworkers confirm a prevalence of experiments aimed at learning facts rather than promoting critical thinking. Subsequent to that study, alternative approaches begin to show improved learning. Thus, recent reviews of chemistry learning in the laboratory environment list several pedagogical interventions and research studies that support the changes Iowa sought to undertake (Nakhleh, Polles, & Malina, 2003; Pienta & Amend, 2005).

Another practical matter with redesign has to do with the number of instructors, teaching assistants, and rooms, particularly the laboratories. The redesign could not increase the number of any of those. In fact, funding from the Pew-funded Program in Course Redesign conducted by the National Center for Academic Transformation (2009) specifically sought to effect cost savings through the increased use of technology. At least some economy was found in changes within the lecture portion, where teaching assistants that served as graders were replaced by using "electronic" homework—web-delivered systems that assigned, graded, and managed class assignments. At first, the integrated laboratories required for the first semester, Principles I, appeared to be a deal-breaker because of the larger number of students enrolled in that course and the attrition that normally occurred in the traditional system. In other words, in the pre-redesign system, the laboratory course was taken during or after the second lecture course, and by that time there were fewer students than at the beginning of the entire sequence. The redesign pedagogy and infrastructure that was devised (*vide infra*) took care of that potential problem.

The Redesign

Some of the details of the course redesign appear in Table 2—the credit hours and the types of class meetings. The major organizational change involved making the laboratory into an integral part of the course—coupling it with the lecture portion. In the final plan, the laboratory consisted of two sessions that would meet in two consecutive weeks. The first, called the "case-study" session, met for 90 minutes in a classroom. It includes instruction and activities performed in anticipation of the session held the following week in which the students meet for two and a half hours in a laboratory.

Some might hesitate to schedule such a large amount of preparation. However, the case-study section allows a didactic introduction, coverage of safety and techniques, and activities that prepare the students for the wet lab experience. Having both reading and written assignments in preparation for both types of sessions means that students are better informed and far more organized. The name of the case-study session comes from a contextual problem or scenario that is associated with it. For example, the laboratory on thermodynamics involves the energy needs of a runner during the case-study session and measurement of the caloric content of food items in the following week as the laboratory experiments. Some of the concerns about traditional laboratories include context, preparation, and sufficient time for discussion, reflection, and questions (Nakhleh et al., 2003; Pienta & Amend, 2005). The Iowa redesign contains all of these elements.

Several organizational aspects made the new plan conservative of resources. Case-study sessions are conducted by an instructor, but all of the lab sections that meet at a specific time attend the same session. In other words, two to four lab sections of 20–24 students each attend the same case-study session. Depending on enrollments, five to seven identical case study sessions are held in one week, all taught by the same instructor. Furthermore, to efficiently use the laboratories, the students in the course are in groups of sections whose activity is staggered by one week. Group 1 meets in case study 1 in week 1 and wet lab 1 in week 2; Group 2 meets in case study 1 in week 2 and in wet lab 1 in week 3. That way, all facilities are used every week. Teaching assistants convene the wet labs, conducting sections over two consecutive weeks and thereby making them better prepared and more effective.

The schedules for content in the lecture class and the laboratory portion are coordinated. Student testing on the lab portion consists of questions on the lecture exam and the written assignments in the lab, summing up to 25% of course grade. The departmental tradition of giving evening common exams for the entire course was continued. Additional pedagogical components also were implemented in the discussions and lecture. Changes were made to the rooms for discussion—tablet-arm chairs were replaced with small tables and chairs, facilitating group work on worksheets or on web-based simulations, animations, or assignments made possible by a laptop available for each pair of students. The lecture portion involved concept tests, peer instruction, and, at times, personal response devices to manage them. Mazur has demonstrated the effectiveness of peer-instruction in large-lecture physics courses using these interventions (Mazur, 1997).

Measuring Outcomes

After two years of planning and preparation of materials, particularly for the laboratory, the new courses were introduced in Fall 2002 with Principles I and in Spring 2003 with Principles II. Outcomes assessment of the students in the Fall 2002 class was conducted (a) by comparing pre- and posttest scores on the Iowa Chemistry Diagnostic Exam; (b) from scores on a question set from an American Chemical Society (ACS) standardized exam; (c) by comparing results on common final exam questions from a previous fall offering of the old first-semester course; and (d) by using the DFW rates from a time period before and after the 2002 implementation. Besides the quantitative measures, a series of focus groups about course components conducted before, during, and after the redesign confirmed what we observed from the numerical data and enabled us to fine-tune some of the changes.

Iowa Chemistry Diagnostic Exam as a pre- and posttest. A chemistry diagnostic exam was developed to advise students about entering a one-semester "prep" chemistry course versus the redesigned course (Pienta, 2003). The pretest was administered the first week of classes using the course management system (i.e., WebCT) and again in the last week of classes as a for-credit homework assignment. The pretest group (N = 754) showed a mean score of 18.5 ± 4.6 while the posttest yielded a mean score of 24.7 ± 3.9. The posttest group consisted of 690 students, and 641 of them completed both the pre- and posttests. This resulted in a mean difference score of 6.2, which was significantly different ($t = 34.9$, $p < 0.0001$). Student achievement on the constituent questions is summarized in Table 3. That table lists the percent correct for the redesign students in the first (pre-) and last (post-) week of classes. For comparison, the performance of students at the beginning of the second semester course (4:014) is included (Fall 2002, N = 267, average correct = 19.2 ± 5.5). The students in the redesign course made considerable gains in learning. Increases were observed for every question in going from the pre- to the posttest. In 29 out of 30 questions (all except the second question), the redesign class outperformed the students who took the traditional course previously. The second question is not covered in the redesigned first-semester course but would have been for the population that was taking the old second-semester course.

American Chemical Society standardized exam. A set of 25 questions from an exam of the ACS Examinations Institute (First Term General Chemistry) was administered to the new Principles I students on the last day of classes. The time allotted did not allow use of the entire exam, which comes in two forms, blue and gray, for which questions and foils are scrambled. For this assessment, 666 out of 784 students (352 blue, 314 gray) participated and received extra credit for their effort. Students thought that the partial credit that they received would be proportional to their performance on those questions. Table 4 reports the percent correct in the Iowa redesign group compared to normative data supplied by the ACS Examinations Institute. The last column in Table 4 is the difference between the redesign class and the norm value. The redesign group shows uniform performance above the level of the comparison group with gains in 30 out of 32 questions. The Iowa blue mean difference was 6.8 ($t = 5.06$, $p < 0.0001$) while the Iowa gray difference was 8.5 ($t = 5.08$, $p < 0.0001$).

Common examination questions. A set of final examination questions was selected from a traditional offering of Principles I for use in the redesigned course (4:011) in Fall 2002. Thus, the 20 questions, distributed over all 11 of 13 chapters that form the new curriculum and that were common to both courses, are summarized in Table 5. Question types could be <u>A</u>lgorithmic (i.e., calculational) or <u>C</u>onceptual. The use of different textbooks and instructors and the corresponding language and wording were different in the two years—focus groups identified three questions (# 5, 7, and 10) in this category. For these 20 questions, the original (old Principles I) average score was 12.6 out of 20, while that from the Fall 2002 offering was 12.5 ± 3.4 (out of 20). That these scores are identical suggests no decrease in the rigor of the 2002 course but also no apparent improvement in the lecture material from the added laboratory experience.

DFW rate. The course redesign implementation in Fall 2002 also marked a change in student success rate. The fall and spring enrollments represent somewhat different constituencies. The fall classes are populated primarily with students in their first semester at the university. The spring class can contain a substantial percentage

Table 3

Comparison of Student Scores on Iowa Chemistry Diagnostic Exam

Question (Results are % correct)	*type**	*4:011* pre*	*4:011* post*	*4:014* pre*
acid/base: order of acidity of common solns given pH	C	58	75	65
acid/base: pOH and pH calculation	A	36	51	66
atomic structure: conclusions from classic experiment	C	74	92	54
atomic structure: # of elementary atomic particles	A	74	93	79
concentration: conversion between defined units	A	47	68	45
concentration: representations using spheres	C	66	78	55
electronegativity: listing order	A	56	93	81
electronegativity: predicting bond types	C	50	89	66
heat of combustion: heat content of fuels from data	C	63	79	50
heat of combustion: heat of combustion calculation	A	64	87	48
ideal gas: PV = nRT calculation	A	34	73	55
ideal gas: sealed cylinder with piston	C	76	84	72
Lewis structure: matching formula with structure	C	44	74	43
limiting reagent: % yield given starting material	A	38	70	49
limiting reagent: stoichiometric relationships	C	35	72	55
mole: molarity of ions from a salt	C	15	43	27
mole: percent of ion in a salt	A	49	68	47
periodic chart: find data on chart	A	93	98	96
periodic chart: identifying groups	C	74	94	84
periodicity: electron configuration	A	73	97	90
periodicity: matching outer shell or valence electrons	C	87	99	89
reactions: grams of product calculation	A	45	71	42
reactions: reactants and products as sphere	C	67	90	65
states of matter: characteristics	C	54	73	51
stoichiometry: coefficients to balance a reaction	A	85	95	84
stoichiometry: sphere represent reactions	C	94	97	84
structure: formulas from sphere representations	C	76	87	72
structure: molecular geometry	A	50	90	69
unit conversion: converting units	A	80	79	71
unit conversion: missing conversion factor	C	71	92	66

*type: A = algorithmic or calculational question; C = conceptual question. 4:011 = new Principles of Chemistry I; 4:014 = old Principles of Chemistry II.

Table 4

Comparison of Redesigned Course with ACS Standard Exams

	Iowa blue % correct	*Norm blue* % correct	*Iowa gray* % correct	*Norm gray* % correct	*Iowa* % gain
1			37.4	30.8	6.6
2			66.3	54.7	11.6
3			73.0	67.7	5.3
4	77.8	64.8			13.0
5	89.5	78.4			11.1
6	81.8	66.7			15.1
7	76.2	69.0			7.2
8	90.6	77.3	81.5	69.5	12.6
9	59.2	51.5	47.6	44.0	5.6
10	68.3	67.7	75.0	64.7	5.4
11	53.4	43.6	55.6	37.1	14.2
12	33.7	27.6	34.2	28.8	5.8
13	45.5	43.2	43.3	37.4	4.1
14			76.8	57.8	19.0
15			66.4	66.4	0.0
16	67.5	65.8	64.2	56.3	4.8
17	87.1	66.0	81.9	58.8	22.1
18	72.0	67.1	66.6	68.7	1.4
19	80.3	70.9	72.5	69.9	6.0
20	76.9	68.0	66.7	71.5	2.0
21	64.5	57.0			7.5
22	45.8	55.5	40.6	50.4	-9.8
23	55.7	50.0	48.6	39.4	7.5
24	54.7	35.5	52.1	31.3	20.0
25	81.4	79.1	86.4	73.8	7.4
26	29.3	33.3			-4.0
27	60.8	55.3	58.0	55.0	4.3
28	66.9	63.2	69.5	59.5	6.8
29	33.4	29.5	37.7	30.0	5.8
30	81.5	78.2			3.3
31			79.2	54.5	24.7
32			42.8	32.5	10.3

Table 5

Common Final Exam Questions from Old and New Principles I

		type	*old Prin I % correct*	*Fall 2002 % correct*	*Fall 2002 net gain*
1	electron configuration	A	88	93	+5
2	moles in balanced equation	A	68	59	-9
3	significant figures	A	32	36	+4
4	equilibrium: LeChatelier	C	35	37	+2
5	quantum numbers	A	71	47	-24
6	atomic structure	A	85	89	+4
7	periodicity: nuclear charge	C	55	61	+6
8	ionic electron configuration	A	86	87	+1
9	atomic radius	A	81	71	-10
10	hybridization	C	61	35	-26
11	acid/base titration	A	23	15	-8
12	molecular geometry	C	52	66	+14
13	partial pressure of gases	A	58	38	-20
14	kinetic molecular theory	C	78	91	+13
15	covalent bonding	C	60	48	-12
16	ideal gas law	C	69	65	-4
17	bonding: octet rule	C	49	61	+12
18	redox reactions	C	47	41	-6
19	periodicity: electronegativity	C	74	89	+15
20	thermochemistry	A	83	86	+3

(> 30%) of students who started in the one-semester prep chemistry class. (The students would place into the one-semester course in the fall based on their score on the Iowa Chemistry Diagnostic Exam.) The first two rows in Table 6 represent data from before the redesign. The last two rows come from the redesign year and from offerings since then. The comparison includes additional variables, including the instructors. Typically, the fall class has two to three lecture instructors while the spring has one to two instructors, depending on the enrollment. Instructors will typically teach a course three times before cycling into another assignment; the data in Table 6 represent many different combinations of faculty. A decrease of 10% in the DFW rate accounts for about 50 students in the spring and almost 100 in the fall, who are now succeeding in the course. Because the data from the common final examination questions (*vide supra*) suggest that the level or difficulty did not change significantly from year-to-year and among different instructors, the outcomes are interpreted as positive results.

Table 6

Rates of D, F, and W Grades Pre- and Post-Redesign of Principles I

	average enrollment	*%D*	*%F*	*%W*	*%DFW*	*average GPA*
1994–2002 spring	516	10.9	5.6	9.9	26.3	2.31
1994–2001 fall	766	9.8	5.8	9.6	25.1	2.35
2003–2009 spring	472	6.8	3.4	6.6	16.8	2.50
2002–2009 fall	884	6.1	3.5	6.0	15.7	2.58

In the intervening years since the redesign, and, in fact, even within a few years of it, student attitudes appeared to change. Before the redesign, students complained to their respective programs, and their advisors forwarded that information to us. After the redesign, concerns from all of these programs diminished or disappeared. Several groups on campus supported tutor services for the old Principles students; after the redesign, they stopped the practice or greatly reduced the numbers, i.e., their services were no longer required as they had been previously. Current or former students no longer passed on to entering students the impression that these were "weedout" courses. Engineering was satisfied using Principles I as the sole course to satisfy the ABET requirements because of some changes in content, the addition of the laboratory, and institution of problem-solving strategies and inquiry into the course components. The courses are still considered challenging and difficult to some. That Principles I is offered in the first semester of many students' matriculation means that some withdrawals correspond to changes in major, often accompanying the discovery of an entirely new interest or aptitude. In retrospect, the assessment of the redesign would have benefitted from longitudinal data using examinations or instruments over the entire 15-year period. Generating assessment data, even in the absence of planned changes, creates a valuable resource, one that can be used to confirm the value of the *status quo* or as the justification for change.

Potential Chemistry Assessment Tools

Since the Iowa general chemistry redesign, the chemical education research community has designed and validated a series of instruments and, in some cases, cross-validated their use with ones from other educational areas (Holme et al., 2010). These examinations or instruments are useful because they have been tested, their use is supported by learning theories, and appropriate statistics are available. In some cases, these data include normative values from national use, including a variety of institutions and student demographics. Most of these tools would benefit either course or curricular changes in chemistry. They probe content knowledge, learning abilities, or attitudes. This is exactly the type of longitudinal data one should have for courses or programs. A brief summary and source of additional information is provided for each.

Standardized exams from the ACS Examinations Institute. The ACS Exams Institute produces standardized examinations at every level of chemistry instruction, creating new versions on a regular basis (American Chemical Society, 2009). For general chemistry, the collection is quite large, covering one or two semesters of instruction and including a variety of perspectives (e.g., brief, paired questions, conceptual). Two undergraduate placement exams are available: the California Chemistry Diagnostic Test (1997 and 2006 versions) and the Toledo Examination (1998 and 2009). The DUCK or Diagnostic of Chemistry Knowledge is a comprehensive exam intended to measure content and problem solving learned over an entire undergraduate career. The DUCK exam is a series of scenarios with accompanying questions. Another new examination, testing learning in the laboratory, is under development as are other assessments.

Group Assessment of Logical Thinking (GALT). The GALT test classifies students according to their logical thinking as concrete, transitional, or formal thinkers (Roadrangka, Yeany, & Padilla, 1982). An application of GALT in a chemistry context (i.e., predicting student success) has been reported (Bunce & Hutchinson, 1993).

CHEMX: Instrument to assess students' cognitive expectations. CHEMX is a survey instrument that measures an aspect of knowledge about learning known as cognitive expectations (Grove & Bretz, 2007). Student attitudes about learning can be compared longitudinally or in comparison to other groups including experts like graduate students and faculty.

MCAI: Metacognitive Activities Inventory. Metacognitive activity can produce substantial improvements in problem solving and learning in chemistry. An instrument to evaluate students' metacognitive ability in solving chemistry problems has been designed and validated (Cooper & Sandi-Urena, 2009) and an application of its use demonstrated (Cooper, Sandi-Urena, & Stevens, 2008).

IMMEX. This web-based system uses a probabilistic approach (i.e., self-organizing artificial neural networks) for studying how students approach complex qualitative chemistry problems (Stevens, Soller, Cooper, & Sprang, 2004). The tool and its analyses have been applied in various interventions including the value of group work (Cooper, Cox, Nammouz, Case, & Stevens, 2008).

ASCI: Attitude toward chemistry. ASCI is 20-item semantic differential assessment instrument for measuring student attitudes toward the subject of chemistry, including scales on interest and utility, anxiety, intellectual accessibility, emotional satisfaction, and fear (Bauer, 2008).

Case 2: The Chemistry Curriculum Assessment

Background

A curricular assessment has different needs and opportunities compared with the course redesign. Whereas the assessment goals for a science course include content knowledge, problem solving, and critical thinking, evaluation of a curriculum must include the vertical integration of knowledge and skills. How does one successfully determine whether an educational program has produced a working version of chemist or scientist? Such a checklist would contain content knowledge, technical skills including manipulative ones, the ability to use the scientific method to design and complete experiments, and the capability to communicate effectively. Different components of the Iowa chemistry curriculum assessment arose from the three major circumstances listed in the introduction—university reaccreditation, professional organization requirements, and the composition of the department's faculty. Each provided unique needs and opportunities that will be discussed.

The Role of University Reaccreditation

The University of Iowa has been a member of and accredited by the North Central Association of Colleges and Schools (NCA). Every 10 years, the university undergoes review by the NCA's Higher Learning Commission (HLC). In preparation for the most recent 2008 reaccreditation site visit, the university conducted a self-study, beginning in earnest several years before. The process involved a steering committee (of which the author was a member) and several subcommittees to manage the data collection and interpretation (University of Iowa, 2008).

The need for information and the desire to engage departments and programs in their own formative self-evaluation gave rise to two approaches to curricular assessment. The Iowa Center for Teaching sponsored a weeklong workshop on assessment of courses and curricula using its own staff and a national expert. Two chemistry faculty (and 14 additional colleagues from other departments) participated and learned about the value of the process and the features and techniques of evaluation. The impact of a single group would be small, but the intention of conducting these workshops multiple times was preempted by the decision to require all programs and majors to devise an assessment plan. A national assessment expert was selected to conduct workshops and to serve as a consultant by visiting individual departments. These sessions were held over the span of several days, and additional help sessions were scheduled, employing the Center for Teaching and other appropriate experts on campus.

Departmental responses ranged from skepticism to outrage. To a large extent, the negative responses expressed concern that any kind of assessment would be used by the administration to make decisions, especially ones about resources. In many instances, departments recognized that each program was being asked to devise a plan so specific to *their* needs that those plans would be unique and make comparison unlikely. A series of bi-weekly brown-bag lunches over most of one semester provided suggestions, feedback, and advice to departmental representatives. Most majors and programs accepted the need and made steady progress. A few individuals came to the help session looking to rally their colleagues in protest but without much success. It was interesting to note that those faculty who had completed the summer workshop a year earlier, almost to a person, became the leaders in their departments to spearhead or complete the project. Clearly, the voluntary program had been successful, but the accelerated timeline preempted the model by which additional assessment leaders emerged. With only one or two exceptions, all programs on campus submitted a plan. The exceptions struggled with reaching consensus within their department about what to do and what might work.

Creating Consensus

The ACS's Committee on Professional Training maintains a set of guidelines, which it uses to evaluate chemistry programs across the United States. Periodically reviewed and updated, the latest set of guidelines appeared in 2008 (American Chemical Society, 2008). The University of Iowa offers an ACS-accredited bachelor's degree and submitted its five-year renewal in 2009 (American Chemical Society, 2010). Departmental discussions in anticipation of the accreditation renewal came from the changes that appeared in the new guidelines—outcomes based on performance rather than on individual courses, recognition that content organization no longer requires traditional subdisciplines (i.e., the analytical, biological, inorganic, organic, and physical areas) but could be interdisciplinary, and adoption of a list of ancillary knowledge. Besides the formal courses and laboratory experiences, the ACS advocates mastery of additional skills related to problem solving, chemical literature, safety, communication, teams, and ethics. In addition, the ACS expects that the periodic reports on which reaccreditation is based contain data and information collected via rational and well-conceived assessment plans. The plans must also demonstrate the means to continuously update them.

In the past, ACS approval was an outcome. However, ACS accreditation is now becoming a process. University accreditation organizations are also promoting the same message—asking their constituencies to devise ways to conduct analyses and make improvements on a regular basis with a frequency closer to every year rather than every five or 10 years. As a result, faculty members are more likely to be confronting the issue. And some guidance is being provided. For example, the ACS program-approval documents include the examples of a senior capstone experience, team projects, and student presentations as means to integrate various experiences, skills, and assessment. Furthermore, undergraduate research is touted as "a highly effective means for imparting, integrating and assessing..." the overall package (American Chemical Society, 2008).

Over the last several years, a turnover of faculty members in the Iowa chemistry department has been accompanied by a new interest in looking at the curriculum and courses with a greater vision for the future than concern about the past. New colleagues are likely to have come from institutions that have their own assessment plans or at least components. It was just such a faculty colleague that spearheaded the group that devised the Iowa chemistry assessment plan.

Generating an Assessment Plan for Chemistry

Like their colleagues, chemistry faculty members were originally skeptical of the curricular assessment plans required by the Office of the Provost. They saw their own opinions rather than external data as a basis for evaluation, and they were happy with what they saw. At a departmental meeting, one colleague stated that the curriculum already was successful. His measure was that our undergraduates, including ones that worked with him personally, were accepted into and succeeded in very good graduate programs. The conversation seemed to focus on "them" evaluating "us." With some help, the conversation eventually turned to asking about evidence and data to support claims, a perspective one might expect from scientists. What evidence do we have that our graduates succeed in graduate programs? What percentage of our graduates took this path? To what schools? If some aspect of the curriculum appeared successful, was it required of all students? For example, should every senior major write a thesis?

A committee was created and, through a formative process that involved several faculty meetings, a plan was formulated. The provost's original request for assessment plans from individual majors and programs was initiated by identifying some examples from within the university and on websites of equivalent institutions. Finance, Psychology, and Chemistry were selected as exemplary plans from the University of Iowa. Those departments had organized their faculty quickly, built consensus, and created successful strategies.

The University of Iowa Chemistry Assessment Plan contains features that have historical origins in the department, match ACS review criteria and suggestions, arose from the university accreditation, or were contributed from the personal experience of colleagues. Measuring content knowledge is relatively simple, especially because the department has used ACS standardized examinations to measure the competency of entering graduate students. (ACS Exams and content knowledge were measured in the course redesign also.) Course syllabi and examinations are objective means to describe the students' experience for all formal courses. The affective side of an education and measuring problem-solving skills is more challenging. But students who participate in undergraduate research make independent decisions and describe them as part of theses or during poster presentations that are already part of the department's traditions. Tracking student production of these materials, the student's reflections about their experience, and some faculty evaluators' impressions of both seemed like an appropriate and achievable goal.

A summary of the chemistry assessment appears in Figure 1, and the complete document is archived on the university reaccreditation site (University of Iowa, 2007). The "Our Values" preamble outlines the departmental philosophy and strategic goals with respect to undergraduate education. The "Credo" lists the skills and knowledge—the four areas each contain several subdivisions. The role and effectiveness of current teaching methods is addressed. The assessment tools define what will be used, who will be responsible for administering it, what the instruments or tests will measure, and how all parts of the credo are accommodated. Finally, a syllabus checklist is provided with the expectation that each component be given a score based on the level at which it is found in a particular course.

Iowa Chemistry Assessment Summary

- Our Values
- Credo
  - Knowledge and understanding
    - Nomenclature, syntax, and symbology
    - Mathematical models and quantitative theories
    - Microscopic, macroscopic, and symbolic descriptions
    - Content knowledge (organic, inorganic, analytical, physical)
    - Content sub-disciplines (e.g., environmental, chemical education)
    - Basic laboratory skills
  - Skills to acquire, analyze, and process data
    - Acquisition through database, library, journal searches
    - Read literature critically
    - Solve problems using the scientific method
    - Express thoughts clearly through writing and presentations
  - Independent and creative thought
    - Assess facts
    - Postulate hypotheses
    - Plan experiments
    - Interpret results
  - Knowledge of the profession in society
    - Mechanics of conferences, presentations, journal publication
    - Jobs and roles of chemists
    - Current topics in chemistry and their societal context
    - Ethics in science
- Current Teaching Methods
- Assessment Tools
  - Collecting information with specialized exams (majors)
    - Standardized ACS exams
  - Collecting information from existing curricula (majors and non-majors)
    - Final poster presentation or final report/paper
    - Language added to each syllabus indicating parts of credo addressed
  - Collecting information from a single class common to all majors
    - Capstone course in chemistry
- Syllabus Checklist
  - Scientific writing
  - Reading and interpreting literature
  - Chemical knowledge base
  - Laboratory skills and safety
  - Developing testable theories/experimental planning
  - Utilizing mathematical models

Figure 1. Summary of the Iowa Chemistry Assessment.

Summary

The first case, or example, outlines a redesign of a large introductory course sequence. It shows how different measures and evaluative data can measure learning gains and success rates between the old and new courses. The project was undertaken to make instructional improvements and not as a research study. In the latter case, one would collect information about content knowledge, learning abilities, and attitudes before, during, and after the changes. In the Iowa redesign, the goal was to improve instruction, and the results show learning gains (from comparisons of pre- and posttest data), efficacy of the overall redesign (from the rates of DFW grades), and the maintenance of course standards or rigor (from the common final exam questions before or after the redesign). In order to judge the value of each component (e.g., integration of lab and lecture, case study vs. wet lab sessions, changes to the nature of discussion section) a stepwise or incremental approach would have been necessary, and additional data would have been required. Nonetheless, based on the evaluative data, the redesign was declared a success from the institutional and departmental perspectives, as well as that of the funding agencies.

The second example recounts the experiences in developing a chemistry curricular assessment. It combines the circumstance arising from a university reaccreditation, the periodic review of the bachelor's degree by the ACS, and interest and experiences of the current group of faculty members. Each of those contributed specific needs and the resulting qualitative or quantitative measures that are used as part of the overall plan. Although the assessment plan is considered a completed activity, it is subject to updating and improvement as the collected data and interpretations warrant, as is good practice for all assessment plans.

Acknowledgements

The author is grateful to the team of faculty who constituted the "General Chemistry Brown-Bag Lunch Group," the unofficial course redesign committee. In particular, Profs. Sonja Franklin, Jason Telford, Russell Larsen, Sarah Larsen, Dwight Tardy, and Dr. Lynn Cannon (Director of Undergraduate Labs) devised and wrote the original case studies and created the laboratory experiments or modified old ones to the new model. Financial support is acknowledged from the Pew Program in Course Redesign and the National Science Foundation (NSF-DUE-00-88072) for supporting the redesign and components including the Iowa Chemistry Diagnostic Exam. Additional financial support from the Department of Chemistry, the College of Liberal Arts and Sciences, Dean Linda Maxson, the Office of the Provost, Associate Provost Lola Lopes, and the University Facilities group for the remodeling and outfitting of discussion classrooms is acknowledged. The efforts of Drs. Russell Larsen, Mohamed el Mazaawi, and Amy Strathman, implementing the new laboratory components, were remarkable and are appreciated. We are grateful to departmental chairs Darrell Eyman, Daniel Quinn, and David Weimer for their support. Thanks again to Sonja Franklin for her leadership on the departmental assessment plan. Finally, we thank the students of the Principles of Chemistry sequence.

REFERENCES

ABET. (2009). *Engineering Criteria 2000 (EC2000).* Retrieved from http://www.abet.org/index.shtml

Abraham, M. R., Cracolice M. S., Graves, A. P., Aldhamash, A. H., Kihega, J. G., Palma Gal, J. G., & Varghese, V. (1997). The nature and state of general chemistry laboratory courses offered by colleges and universities in the United States. *Journal of Chemical Education*, *74*, 591–594.

American Chemical Society (ACS). (2008). *Undergraduate professional education in chemistry: ACS guidelines and evaluation procedures for bachelor's degree programs.* Committee on professional training guidelines. Retrieved from http://portal.acs.org/portal/PublicWebSite/about/governance/committees/training/acsapproved/degreeprogram/WPCP_008491

American Chemical Society (ACS). (2009). Division of Chemical Education Examinations Institute. Retrieved from http://chemexams.chem.iastate.edu/

American Chemical Society (ACS). (2010). *Approval for bachelor's programs.* Retrieved from http://www.acs.org

Bauer, C. F. (2008). Attitude toward chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of Chemical Education, 85*, 1440–1445.

Bunce, D. M., & Hutchinson, K. D. (1993). The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry (SYM). *Journal of Chemical Education*, *70*, 183–187.

Cooper, M. M., Cox, C. T., Nammouz, M., Case, E., & Stevens, R. (2008). An assessment of the effect of collaborative groups on students' problem-solving strategies. *Journal of Chemical Education*, *85*, 866–872.

Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, *86*, 240–245.

Cooper, M. M., Sandi-Urena, S., & Stevens, R. (2008). Reliable multi-method assessment of metacognition use in chemistry problem solving. *Chemistry Education Research and Practice*, *9*, 18–24.

Grove, N., & Bretz, S. L. (2007). CHEMX: An instrument to assess students' cognitive expectations for learning chemistry. *Journal of Chemical Education*, *84*, 1524–1529.

Holme, T., Bretz, S. L., Cooper, M., Lewis, J., Paek, P., Pienta, N., . . . Towns, M. H. (2010). Enhancing the role of assessment in curriculum reform in chemistry. *Chemistry Education Research and Practice*, *11*, 92–97.

Iowa Department of Education. (2008). *The Annual Condition of Education Report.* Retrieved from http://www.iowa.gov/educate/index.php

Mazur, E. (1997). *Peer instruction: A user's manual.* Upper Saddle River, NJ: Prentice-Hall.

Merriam, S. B., & Cunningham, P. M. (Eds.). (1989). *Handbook of adult and continuing education*. San Francisco: Jossey-Bass.

Nakhleh, M. B., Polles, J. & Malina, E. (2003). Learning chemistry in a laboratory environment. In J. K. Gilbert, O. De Jong, R. Justi, D. F. Treagust, & J. H. Van Driel (Eds.), *Chemical education: Towards research-based practice*. Dordrecht, Netherlands: Kluwer.

National Center for Academic Transformation. (2009). *Program in course redesign*. Retrieved from http://www.thencat.org/PCR.htm

Pienta, N. J. (2003). A placement examination and mathematics tutorial for general chemistry. *Journal of Chemical Education*, *80*, 1244–1246.

Pienta, N. J. & Amend, J. (2005). Electronic data collection during laboratory activities. In N. J. Pienta, T. L. Greenbowe, & M. M. Cooper (Eds.), *Chemist's Guide to Effective Teaching*. New York: Prentice-Hall.

Roadrangka, V., Yeany, R., & Padilla, M. (1982, January). *The construction of a Group Assessment of Logical Thinking (GALT).* Paper presented at the meeting of the National Association of Research in Science Teaching (NARST) annual conference, Athens, GA.

Stevens R., Soller, A., Cooper, M., & Sprang, M. (2004). Modeling the development of problem solving skills in chemistry with a web-based tutor. *Lecture Notes in Computer Science*, *3220*, 580–591.

University of Iowa. (2007). *Undergraduate assessment by the Department of Chemistry Spring 2007.* Retrieved from http://www.uiowa.edu/reaccreditation/outcomes/chemistry.pdf

University of Iowa. (2008). *Fulfilling the promise: Reaccreditation 2008.* Retrieved from http://www.reaccreditation2008.uiowa.edu/

CHAPTER 10

EVALUATING UNDERGRADUATE CHEMISTRY REFORM: CHALLENGES, OPPORTUNITIES, AND DIRECTIONS AT MIAMI UNIVERSITY OF OHIO

Jane Butler Kahle, Kathryn Scantlebury, Sarah Beth Woodruff, and Yue Li
Miami University of Ohio

Abstract

This chapter discusses the evaluation of two large-scale projects to reform first- and second-year chemistry courses. The projects, funded by the National Science Foundation (NSF) under its Undergraduate Research Center (URC) program,[1] focused on providing students with authentic research experiences in undergraduate chemistry courses with the goal of increasing both the number and the diversity of undergraduates electing to continue to study chemistry. Two projects, which have characteristics representative of most of the URC projects, and their evaluations are discussed in this chapter. Both projects chose to change the nature of their introductory courses through modules that included cutting-edge research and real-life applications of chemistry. Both involved multiple partners across a variety of institutions (research universities, two- and four-year colleges and universities, public and private institutions). The partnerships provided unique challenges (availability of equipment on some campuses, institutional support, timely reporting) as well as opportunities (replication of modules, cross-institutional research).

The evaluations involved collecting data from courses or laboratory sections that included one or more modules and from courses or sections that did not. Although the evaluation of each project involved different procedures and instruments, commonalities were found and generalities emerged that may indicate directions for other large-scale, undergraduate chemistry projects as well as their evaluations.

Findings suggest future directions for improving undergraduate chemistry courses as well as for evaluating large-scale, multisite projects. First, the research modules significantly affected the understanding of the nature of science for women students as well as for students who planned on professional careers (e.g., medicine, dentistry). Second, women students in module, compared to nonmodule, sections were more actively engaged in learning chemistry. Third, Non-White students reported more interest in chemistry/science and had more positive perceptions of learning chemistry through the lab, compared to White students. Fourth, the availability of equipment, particularly at two-year institutions, was a challenge, suggesting that strategies

[1] The program's name was later changed to Undergraduate Research Collaboratives.

for sharing equipment across campuses must be addressed. Fifth, there are more advantages than disadvantages to evaluating multisite projects. Last, longitudinal studies are needed to assess if active involvement with chemistry research increases the number of students interested in pursuing chemistry majors and careers.

Background

The Undergraduate Research Center (URC) program was primarily an initiative of the NSF's Division of Chemistry with strong support from the Divisions of Undergraduate Education, Human Resource Development, and Research, Evaluation, and Communication (National Science Foundation, 2003). The program was directed at developing new models and partnerships that would provide first- and second-year college students with research experiences in chemistry. In 2003, NSF sponsored a workshop to develop the first Request for Proposals for the URCs (National Science Foundation, 2005). Proposals were requested that focused on the academic year, provided research experiences in chemistry for large numbers of students, and developed new models and partnerships between two- and four-year institutions that were scalable and sustainable. Further, students were to be engaged in current research projects that used modern tools and methods. A primary goal of the URCs was to provide undergraduate students with opportunities to develop an understanding of how scientists produce chemical knowledge through research. The alignment of URCs with other divisions at NSF, namely Undergraduate Education and Human Resource Development, meant that there also was an expectation that projects would influence curriculum reform in chemistry and improve student attitudes toward science, potentially increasing the number of students from groups underrepresented in the sciences who selected advanced science courses and careers. The program announcement yielded 141 proposals: 53 for full grants and 58 for planning grants.[2] Seven hundred colleges and universities were involved. One proposal received full funding, and 20 planning grants were supported in the initial solicitation. Eventually, five URC projects were funded. The funded projects had several commonalities: they (a) built new partnerships and/or models; (b) focused on chemistry research; (c) expanded opportunities for undergraduates to become involved with chemistry research; (d) demonstrated partner equity; (e) were sustainable after funding ended; (f) included rigorous, external evaluation plans; and (g) focused on expanding research capacity and infrastructure through promoting excellence in undergraduate education (Kuczkowski, 2005).

In general, within programs, the NSF approach to funding projects reflects a fundamental belief that a project should have the flexibility to determine its own priorities. Such flexibility results in great variability among funded projects and across multiple sites within the same project. While variability between and within projects is a concern to evaluators in making claims regarding project outcomes, Gullickson

---

[2] The disparity in number is due to the receipt of 30 linked collaborative proposals that were not reviewed separately.

and Hanssen (2006) suggest that conducting cross-site and cross-project evaluation is a viable way to address overarching questions. The synthesis of the evaluations of URC projects, discussed in this chapter, provides both challenges (multiple sites) and opportunities (possible comparisons across sites and projects).

Because of the unique opportunity to review and/or evaluate several projects under the same program umbrella, Miami University of Ohio's Evaluation and Assessment Center for Mathematics and Science Education (E & A Center) began to identify and examine challenges and opportunities that were found across projects or that were unique to one project. In this chapter, the characteristics of two representative URC projects and their evaluations are described. In some cases, details and/or findings from projects have been combined to ensure anonymity; however, the projects and their evaluations are typical of those funded by the URC program. Those sections are followed by a discussion of issues in large-scale evaluation as well as sections discussing the challenges, opportunities, and possible directions for large-scale projects and their evaluations.

Descriptions

First, two projects that typify the URC program are described. Next, the evaluations of those two projects are compared and contrasted.

Project Descriptions

Characteristics representative of one or more of the five URCs are used to describe the two typical projects discussed here: Research Experiences in Chemistry Education (RECE) and Undergraduate Chemical Education Center (UCEC). Both projects were based at large, research universities, and each had multiple partners. RECE's partners were all within one state, and the partnership built upon an established consortium of chemistry department chairpersons. Eventually, 14 different institutions participated in RECE, 13 of which were public institutions. One institution was a community college; the others ranged from midsized, selective universities to large, research universities. On the other hand, UCEC had partners in three noncontiguous states. Its partners included two community colleges, one of which is a historically black college; research universities as well as a former teacher's college were among its partners. Again, the partners represented a variety of higher education institutions.

One project (RECE) had a centralized leadership model. The chairperson of the chemistry department of the host institution was the project's principal investigator (PI). In addition, he organized and hosted the chemistry consortium for the state prior to the URC. He had both clout within the institution and credibility across the state. On the other hand, the leadership in the other project (UCEC) involved faculty from several campuses (and states) as coprincipal investigators. Although its PI, a chemistry faculty member at the host institution, put together a working consortium across states and a wide variety of institutions, the consortium did not have a history of collaboration. Responsibility for that project's implementation was shared across several institutions.

RECE's reform of undergraduate chemistry involved three courses, general chemistry, organic chemistry, and analytical chemistry, while UCEC's efforts involved two courses, general chemistry and organic chemistry. On campuses with multiple sections of the same course, the reform materials were used in some sections and traditional materials in others. On campuses with only one section of a specific chemistry course, the reform materials were used, and findings were compared with findings from traditional sections on other campuses. All findings in each project were analyzed by whether a student had studied a module (reform) or not (traditional). Due to small sample sizes and limited resources, findings were not analyzed by specific institution. This decision produced a challenge for project personnel, because each partner was interested in how successful the implementation was on its own campus.

Both projects involved undergraduate students who studied a module (i.e., were in a reform course or section of a course) and those who did not study a module (i.e., were in a traditional course or section of a course). The number of students who participated and who returned questionnaires varied across the years in both projects as shown in Table 1.

From the evaluation's point of view, a continuing challenge was the uneven numbers of students participating at two very different types of institutions. Both host

Table 1

Number of Undergraduate Student Questionnaires Received Across Years for UCEC and RECE

Year/Semester for UCEC	**UCEC Participant***	**UCEC Non-Participant***	**Total**
Spring 2006	188	210	398
Fall 2006–Spring 2007	479	821	1300
Fall 2007–Spring 2008	181	348	529
Fall 2008–Spring 2009	352	622	974
Total	**1200**	**2001**	**3201**
Year/Semester for RECE	**RECE Participant**	**RECE Non-Participant**	**Total**
Spring–Summer 2006	409	484	893
Spring–Fall 2007	1599	577	2176
Winter/Spring–Fall 2008	2067	37	2104
Total	**4075**	**1098**	**5173**

**The numbers for UCEC Participants and Non-Participants were calculated after matching pre- and postresponses for each semester; RECE did not collect preparticipation responses.*

institutions, as stated, were large, research institutions; and, particularly in general chemistry, the number of students involved at those two institutions was very large. Conversely, the number of students participating at the community colleges was very small. Further, the ratio of participant to nonparticipant students varied in the two projects, as shown in Table 1. These issues were a challenge for the evaluation, which used a variety of statistical approaches to compensate.

In both projects, faculty at the host and partner institutions were involved in developing and implementing the modules. Two different models were used: (a) the adaptation and refinement of existing materials to include authentic chemistry problems and real-life applications of chemistry; and (b) the development of new curricular materials, which often involved cross-institutional teams of chemists. The self-contained modules lasted from 3 to 4 weeks in institutions that were on the quarter system and slightly longer at those with semesters.

In the single-state project, the modules were clustered into two research themes. All needed laboratory equipment and materials were specified; and, in the case of expensive equipment, arrangements were made to borrow equipment and instruments from the host institution. Indeed, the support of the host institution and its initiation of loaned equipment and instruments were critical to the success of the project and suggest an important direction for other programs and projects directed at changing the nature of undergraduate science courses.

Although fewer modules were developed and implemented by the multistate project, they, too, focused on authentic chemistry problems and included real-life applications. One issue that arose early in that project was the lack of equipment and instruments at the partner community colleges. Because the project spanned several states, it was not possible to institute the same type of resource sharing that was feasible in the single-state project. This challenge may have limited the type of module developed and may have affected the extent of implementation of the project. As shown in Table 1, fewer students were involved with the reform materials in the multistate project.

In both projects, the modules involved active student participation to solve new, as opposed to "cookbook," chemistry problems. Students collected, shared, and analyzed data before drawing conclusions. Different findings were discussed, and procedures were replicated as needed. In many instances, students studied chemistry problems in the field, and data, collected for the same module, were shared across campuses. The curriculum development presented both challenges and opportunities for the faculty developers and, later, for the faculty implementers. For example, many faculty were supported during the summer to turn their research into undergraduate modules. And, during implementation, support was available to include undergraduates in research groups.

Evaluation Descriptions

The evaluations of all URC projects involved a mix of quantitative and qualitative methods. The design of the evaluations, described here, differed by project but yielded similar comparisons. For example, one project used a post-only design and compared responses from participants (e.g., those in a module section or course) and nonparticipants (e.g., those not in a module section or course). The other, however, used pre/post project data collection, and analyses included pre/post comparisons and two-way ANOVA comparisons by pre/post and by participant/nonparticipant. And, typical of the URCs, neither project allowed the evaluation to use any standardized tests to collect student achievement data. Project personnel were convinced that existing tests would not adequately assess learning with the reform materials, and neither had the resources to develop, test, and validate assessments for the modules. Further, it was questionable whether or not faculty at the partner campuses, who were implementing the modules, would agree to any assessment that they had not developed. The lack of achievement data to support other evaluation findings was an unmet challenge.

The original research design for the evaluation of UCEC was developed prior to the involvement of the E & A Center and relied heavily on qualitative data, collected initially by the evaluators and later by project staff. Those data consisted of interviews with instructors in the module courses, peer leaders, module developers, and undergraduate chemistry students. It also used a questionnaire, developed primarily by project personnel. An unexpected challenge arose because students were identified by their student ID numbers; however, project personnel had not foreseen that there could be duplications of student ID numbers across institutions. This situation prevented the evaluation from matching student pre- and postresponses as well as responses across the years, and lowered the number of questionnaires that could be used in the evaluation.

The host institution in the centrally organized project (RECE) was able to allocate resources to the collection and scanning of all evaluation data. Data collection in the dispersed leadership model (UCEC), however, was delegated to each partner, who was responsible for forwarding the data to the E & A Center. Partners, who were dependent upon the host institution for funding, were diligent in collecting and sending data to the host institution for RECE. However, partners in UCEC, who received no financial support from the E & A Center, often were late in sending raw data or simply forgot to do so. In order to address this challenge, the project switched from paper questionnaires, which were completed in class, to online data collection midway through funding. However, this change introduced a new challenge as many students did not complete the online questionnaire or did not respond to every item. From the point of view of the evaluators, the RECE strategy was efficient and resulted in higher response rates and more reliable data.

UCEC used an *Undergraduate Student Questionnaire* that had six subscales: "Interest in Chemistry/Science," "Real Life and Science," "Authentic Scientific Lab Practices," "Perceptions of Learning Through Lab," "Belief in Chemistry Knowledge," and "Collaborative Learning in Courses." All but the last subscale had acceptable reliabilities (all Cronbach alphas were over 0.80). Responses to all items were on a Likert-type scale that ranged from *strongly disagree* to *strongly agree*.[3] Item Response

[3] Responses were *strongly disagree*, *disagree*, *barely disagree*, *barely agree*, *agree*, and *strongly agree*.

Theory-Rasch Model was used to change raw scores into a Rasch score for each of the six subscales. The Rasch Model was chosen because it allows comparisons across cohorts and institutions. All statistical analyses of student data (from both two- and four-year institutions) compared Rasch mean subscale scores of students in the reform sections and courses with those of students in the traditional sections/courses. Repeated measure analyses of Rasch subscale mean scores were done, and data were compared preparticipation and postparticipation for students in reform as well as those in traditional courses or sections.

RECE's *Undergraduate Student Questionnaire* was adapted from the E & A Center's valid and reliable *Student Questionnaire*. Three of a possible five subscales were used: "What Instructors Do," "What Students Do," and "Views of Science/Chemistry." All subscales had acceptable reliabilities with Cronbach alphas ranging from 0.67 to 0.91. Responses on the first two subscales ranged from *almost never* to *very often*,[4] while responses on the "Views of Science/Chemistry" subscale ranged from *strongly disagree* to *strongly agree*.[5] All of the statistical analyses for the RECE *Undergraduate Student Questionnaire* also used Rasch mean scores.[6] Independent sample *t*-tests were used to assess any differences between the mean responses of students who studied a module and of those who did not for each course. Only postmodule/nonmodule data were collected.

Issues in Large-Scale Evaluations

Recently, NSF has vigorously promoted evaluations as critical to ensure project and program credibility (Gullickson & Hanssen, 2006). Two important reasons for engaging in evaluation, noted in the *2002 User-Friendly Handbook for Project Evaluation* (National Science Foundation, 2002), are "to provide information to improve the project; and to provide new insights or information that were not anticipated" (p. 3). Katzenmeyer and Lawrenz (2006) suggest that many NSF-funded programs and projects have not been evaluated well or that the evaluations have such a low priority (e.g., underfunded) that the agency has limited evidence of impacts and effects. Further, Kumar and Altschuld (2008) emphasize the need for comprehensive evaluations of large-scale projects that might collectively lead to policy recommendations.

Logically, it might be assumed that projects designed and implemented to address the same NSF programmatic goals would be similar and would be internally consistent. Yet, aspects of context, scale, and the potential for variability can undermine even the best-intentioned efforts at uniformity. For example, contextual

[4] Responses were *almost never*, *seldom*, *sometimes*, *often*, and *very often*.

[5] Responses were *strongly disagree*, *disagree*, *undecided*, *agree*, and *strongly agree*.

[6] The order of responses differed among the subscales, switching from negative to positive and from positive to negative. In addition, negatively worded items on all subscales were reverse coded. The numerical value assigned to any response category was not shown on any instrument.

factors create limitations that must be accepted in any evaluation, but the evaluation of multiple sites in each of the projects discussed exacerbates this challenge. Variations in site characteristics may interact with participant characteristics to produce outcomes highly specific to the particular site (Guskey, 2000). However, multisite evaluations may identify contextual conditions under which large-scale interventions are most effective (Woodruff, Zorn, Noga, & Seabrook, 2009). From an evaluation's viewpoint, it is necessary to understand that contextual issues, such as policy, resources, and environment, cannot be strictly controlled. As described above, contextual issues differed in the two typical URC projects, which limited the extent to which findings could be generalized.

Although useful for project evaluation and for the synthesis of project-level evaluation data for program improvement, requiring specific objectives or outcomes for projects might be perceived as the "evaluation tail wagging the NSF dog" (Katzenmeyer & Lawrenz, 2006, p. 11). Thus, projects that appear to be similar, such as the two described here, may produce different results for subtle and unanticipated reasons. This situation poses a challenge, as traditionally evaluators have sought to isolate key features that contribute to project outcomes and/or to determine elements of effective project implementation. Conversely, the opportunity to evaluate more than one large-scale project could collectively lead to major policy implications, if findings are synthesized for use in program improvement. Although the evaluations of the two typical URC projects, discussed in this chapter, used similar approaches, differences between and within each project required flexibility in their evaluations. Each project's evaluation procedures and instruments differed; however, there were commonalities across the projects to reform undergraduate chemistry. Indeed, directions emerged that can inform evaluations of other large-scale reform efforts. Table 2 presents the *challenges, opportunities,* and *directions* that the evaluation identified in the two typical projects. They are discussed in the sections that follow.

Challenges

Several challenges already have been noted in the project and evaluation descriptions above. Yet, two overriding challenges have not been discussed. First, all URC projects were based in chemistry departments, and all had research chemists as their principal investigators. The evaluations were grounded in social science theory, while the project personnel understood research in a scientific field. The differences between both the methods of conducting the research and feasible outcomes had to be continually discussed and accommodated. For example, the discrepancy between scientific research designs and what is possible when humans are the subjects of the research may cause confusion and misunderstandings. This challenge is found across NSF projects as routinely scientists, mathematicians, and engineers are the principal investigators even for education projects. Routine evaluation issues, such as not modifying instruments during the evaluation, using identification numbers that are not likely to be duplicated, and ensuring that any instrument is reliable for the population studied, were challenges in these evaluations.

Table 2

Challenges, Opportunities, and Directions

Challenges	1. Accommodating scientific research expectations with social science methodology
	2. Establishing trust and collegiality between project and evaluation personnel
	3. Collecting achievement data and analyzing data across multiple sites
	4. Enhancing host institution support
	5. Assessing any increase in numbers and/or diversity of students opting for scientific careers in a five-year timeframe
Opportunities	1. Conducting cross-institutional research
	2. Evaluating reform curriculum in different situations and with diverse groups of undergraduates
	3. Providing venues for students to present research in the single-state model
	4. Involving undergraduates, who had experienced the modules, as peer mentors or peer leaders
Directions	1. Challenges presented by multiple sites are less important than the opportunities afforded
	2. Critical role of the host institution
	3. Importance of an established and functioning consortium for multisite projects
	4. Involvement of evaluators early in the life of a project, preferably during the proposal stage

Second, although NSF usually requires external evaluations, often the evaluator is not involved in developing the proposal. Rather, at the time of funding, and at NSF's insistence, one is contacted. Therefore, there may be the challenge of establishing trust and collegiality between project and evaluation personnel. Project personnel need to understand that consistency is important if findings are to be compared across the years of funding. Likewise, evaluators must realize that certain changes or modifications are necessary as a project progresses. Collaboration can result in reliable data and a richer evaluation—one that also is useful in suggesting modifications and/or changes during project implementation.

A challenge common to the evaluations of the projects described in this chapter was the collection and comparison of data across multiple sites. For example, both projects struggled to assist smaller institutions, particularly two-year institutions, with Institutional Review Board (IRB) issues. In one case, the host institution actually assisted a community college in establishing its IRB. This challenge was not anticipated by either the host institutions or by the evaluators.

Another unanticipated challenge was the degree of host institution support. In the evaluations discussed, institutional support affected both the reliability and validity of the data collected. In addition, as mentioned, the issue of single-state or multiple-state project was a challenge in at least two ways: sharing of equipment for the modules and collection of data.

Last, the challenge of assessing whether or not the URC projects and program resulted in increasing the numbers and diversity of students opting for scientific careers could not be met in the evaluation timeline provided. If funding agencies are serious about long-term goals, often stated in Requests for Proposals, evaluations must continue beyond the period of project funding. One project addressed this challenge during its no-cost extension period. Other projects have sought private support in order to extend the evaluation period.

Opportunities

Although multiple and diverse partners created many challenges for the URC projects, they also provided opportunities for the projects; that is, there were opportunities to replicate the reform curriculum and opportunities for cross-institutional research. One example involved using findings from one project to enhance the evaluation of the other as well as to generalize across the projects. For example, a goal of the URC program was to increase and diversify the number of students continuing in chemistry courses and careers. Therefore, both evaluations collected and analyzed student demographic data. However, in addition to items concerning race/ethnicity, gender, age, year in college, one project requested demographic information about student undergraduate major (or anticipated major) and career goals. Differences were found between the responses of students identifying themselves as *attending graduate school in the sciences* or *pursuing a professional degree*. There were strong indications that the modules were instrumental in changing the perception of chemistry for students who intended to pursue professional careers. These differences were especially notable in general chemistry courses, the introductory course that may positively or negatively affect a student's choice to continue to study chemistry and, perhaps, his/her career choice. A unique opportunity evolved when similar demographic categories were added to the other project's questionnaire so that the findings could be refuted or confirmed.

Because both of the projects discussed included multiple sites, both had opportunities to test the reform materials in different situations and with diverse groups of undergraduates. Across the years, the curriculum materials in both were revised and improved, and in both, the modules have become a permanent part of the curriculum for courses for which they were developed. This opportunity helps to ensure that the gains of the URC projects are sustained.

Further, the single-state (or institution) model, compared with the cross-state model, more easily accommodated venues for students to present their research, another goal of the URC program. For example, RECE annually held statewide seminars where undergraduates presented their research to project personnel, faculty, and students. From those meetings, student research was identified for presentation at regional meetings of the American Chemical Society.

Another opportunity that evolved as the projects progressed was the involvement of undergraduates—who had previously experienced the modules—as peer mentors or leaders in reform sections or courses. These opportunities were instrumental in changing student commitment to research science or to solidifying that career goal.

Directions

As described in this chapter, individual project and collective scale provided both challenges and opportunities for their evaluations. Each project was implemented at a large number of sites, impacting a significant number of participants. Although the grand scale of these evaluations, independently and collectively, provided the opportunity to merge data directly or via meta-analysis to create a more comprehensive evaluation, the extent to which those steps were reasonable and advisable was an issue. As noted by Kumar and Altschuld (2008), "Comprehensive approaches to evaluation of science initiatives are lacking; instead, disparate studies are used to draw evaluative conclusions" (p. 606). While not a program evaluation per se, the parallel evaluations discussed here suggest evidence, findings, and recommendations that may inform program changes and/or result in policy recommendations. If the same or similar variables predict outcomes effectively across projects, Gullickson and Hanssen (2006) suggest that the findings may provide information regarding fundamental practices for future projects.

Therefore, what directions may be gleaned from the typical projects described and from their evaluations? First, the challenges presented by multiple sites are less important than the opportunities they afford, particularly opportunities to revise and improve curriculum materials and to confirm or refute evaluation findings. Second, the role of the host institution is critical and needs to be addressed in each proposal for funding. Third, an established and functioning consortium can more easily initiate and sustain a complex, multisite project than a newly established consortium. Fourth, in order to establish mutual trust and collaboration and to ensure a high quality evaluation, evaluators should be involved early in the life of the project, preferably during the proposal stage.

In summary, evaluating one or more projects within a program provides a value-added to each individual evaluation. One can cross-check findings and confirm or gather additional data. In evaluating the typical projects described in this paper, evidence of change by women students as well as by students planning on professional careers in one project was verified by findings in the other one. Similarly, recommendations for midproject adjustments were informed by findings in both projects. That is, if something was working well in one, but not in the other (e.g., data collection), a recommendation for evidence-based change was made. In addition, shared evaluation findings can provide directions for existing as well as future large-scale projects. Data gathered from two, multisite projects that included four states lent strength to both the validity of the findings and to their generalizability to undergraduate chemistry education. This result was particularly important in addressing the overall vision and goal of the NSF Undergraduate Research Collaborative program.

REFERENCES

Gullickson, A. R., & Hanssen, C. E. (2006). Local evaluation in multisite STEM programs: Relating evaluation use and program results. *New Directions for Evaluation, 109,* 87–103.

Guskey, T. R. (2000). *Evaluating professional development.* Thousand Oaks, CA: Corwin Press.

Katzenmeyer, C., & Lawrenz, F. (2006). National Science Foundation perspectives on the nature of STEM program evaluation. *New Directions for Evaluation, 109,* 7–18.

Kuczkowski, R. (2005). Implementation of the first year of the NSF-URC program. In National Science Foundation, *Implementation of undergraduate research centers: A report on a National Science Foundation workshop* (pp. 14–17). Retrieved from http://www.scu.edu/cas/research/upload/URCworkshopfinalreport-3.pdf

Kumar, D. D., & Altschuld, J. W. (2008). The need for comprehensive evaluation in science education. *Review of Policy and Research,* 20(4), 603–615.

National Science Foundation. (2002). *The 2002 user-friendly handbook for project evaluation*. Arlington, VA: Author.

National Science Foundation. (2003). *NSF workshop—Exploring the concept of undergraduate research Centers: Workshop final report.* Retrieved from http://urc.arizona.edu/

National Science Foundation. (2005). *Implementation of undergraduate research centers: A report on a National Science Foundation workshop*. Retrieved from http://www.scu.edu/cas/research/upload/URCworkshopfinalreport-3.pdf

Woodruff, S. B., Zorn, D., Noga, J., & Seabrook, L. (2009, April). *State-sponsored professional development: Lessons learned through dialogue across evaluation, theory, research, and practice*. Panel presentation at the annual meeting of the American Educational Research Association, San Diego, CA.

CHAPTER 11

STRUCTURE AND REACTIVITY AT THE UNIVERSITY OF MICHIGAN

Brian P. Coppola
University of Michigan

In 1989, the University of Michigan Department of Chemistry broke rank with the vast majority of colleges and universities and eliminated the two-semester general chemistry course as the postsecondary introduction to the discipline. Instead, students with a reasonable background begin their college-level study with an organic chemistry course that we call *Structure and Reactivity*. From the start, the development of this course was based on sound pedagogical principles and contemporary instructional recommendations. Over the last 20 years, and through roughly 50,000 students, the department has not only continued to evolve the course in both content and method, but also carried out substantive research on student learning that has informed practice. In this chapter, I will trace the development of the course, and describe in detail three cases of alignment between our explicitly identified learning goals, our pedagogical approaches to achieving those goals, and the methods we used to assess our outcomes.

Introduction & History

In 1989, the Department of Chemistry at the University of Michigan restructured its undergraduate curriculum. Details about the origins and process of that change can be found in two publications in the *Journal of Chemical Education* (Coppola, Ege, & Lawton, 1997; Ege, Coppola, & Lawton, 1997). Briefly, we have approximately 2,800–3,200 students each fall term who intend to take an introductory chemistry class. Based on information from a placement examination, as well as from academic advisors, about 1,600–1,800 students (mostly in engineering programs) begin with a one-term course in general chemistry principles. The other 1,200–1,400 students take *Structure and Reactivity*, which any chemistry instructor would recognize as a one-year introductory course in organic chemistry. Around 55–60% of the enrollments in this latter course are first-term, first-year students, and they are the majority of the future physical and biological science majors, chemical engineers, and preprofessional (medicine, dentistry, veterinary) students.

In our view, incoming university students who have demonstrated a baseline degree of chemical literacy do not need another year of introductory physical "general" chemistry followed by a year of "sophomore organic" chemistry. Many important concepts typically taught in general chemistry arise during exploration of the structures and reactivity of organic compounds and the inorganic species that interact with them. From our experiences in teaching organic chemistry for sophomores, we already knew the answer to the questions posed below.

> Isn't it possible to teach bonding, VSEPR, polarity, physical properties, the periodic properties of elements, acidity and basicity, oxidation and reduction, energetics and kinetics using organic as well as inorganic structures? Oxygen, nitrogen, sulfur, phosphorus, silicon, boron, the halogens, and many transition metals are very much a part of "organic chemistry"? What is necessary is a context in which this rich chemistry can be explored in ways that revisit a few important themes throughout an entire year and which provides opportunity for students to practice these themes with increasing understanding and sophistication. Such a context is found in mechanistic organic chemistry because it is the area where a structural molecular approach and mechanistic rationalization of reactivity are most highly developed. (Ege et al., 1997, p. 74)

The first-term *Structure and Reactivity* class, divided into sections with 300–350 students each, meets three times per week in a large lecture hall. There are a number of formal and informal learning resources made available to those enrolled in the course. Smaller groups of 18–24 students meet with graduate student instructors for one-hour recitation sections. Nearly all of these students also are registered for the first-term laboratory course, which meets for four hours once a week. In addition to office hours and appointments, each faculty member also offers a two-hour open session once a week that we call "workshops," where the only ground rule is that authentic questions about the subject matter must be asked. In other words, "Can you do problem 24(b)?" is not a subject matter question, while "Can you explain how to evaluate conformational energy differences with Newman projections?" is. The Science Learning Center, a service unit of the College (http://www.lsa.umich.edu/slc) facilitates the formation of peer-led study groups in the majority of our introductory science courses, serving an estimated 50–75% of students in these classes.

There are three examinations (common to the entire course, a 60-minute exam given in a one and a half-hour evening period) and a comprehensive final examination (a one and a half-hour exam, also common to the entire course, given in a two-hour period). Students in the first-term course who wish to receive Honors credit (there are usually 120–140 of them) can do so by participating in the Structured Study Group program (Coppola, 2001b; Coppola, Daniels, & Pontrello, 2001; Varma-Nelson & Coppola, 2005), which is described below.

The second-term course is arranged much like the first. One difference, though, is that there is no recitation section. Instead, that hour of credit is shifted to the laboratory course, which meets for a formal hour of lecture in addition to the four-hour laboratory session. This lecture time is devoted primarily to instruction in spectroscopic identification using the appropriate chapters from the *Organic Chemistry* text. In 2005, we shifted the traditional organizer in the second-term class from organic synthesis to bio-organic chemistry, reflecting the evolution of the field as well as the interest of our students.

In 1994, we redefined what it meant to take these organic classes for Honors credit. In an effort to gather together the science-motivated students, we began offering Structured Study Groups (SSG)—a supplemental instruction option wherein students from any of the large lecture sections could elect to meet for an additional two hours per week, in groups of about 20, facilitated by an upper-level undergraduate leader. The pedagogical organization in the SSGs is based on studio instruction, where students have creative, divergent, and generative assignments each week that they bring to the session for peer review, critique, and self-editing (Coppola, in press). These science-oriented students have a choice during the second term. Those who are enrolled in the large sections may once again elect the SSG option. Alternatively, we do offer a separate class for students interested in pursuing a more research-oriented experience. About 100 students enroll in this course, which offers a laboratory integrated with the lecture, a series of term-long projects, and greater reliance on primary scientific information and experimental design. All the students in this section also meet for SSGs, which dramatically extends the nature of the course.

Pedagogical Features & Learning Objectives

Pedagogical features and learning objectives are linked because of the principle of alignment (Bransford, Brown, & Cocking, 1999; Porter, Smithson, Blank, & Zeidner, 2007), which, while commonly used in precollege settings to describe the link between tests and standards, can refer more broadly to the understudied link between learning goals and pedagogical methods. Indeed, there is still an unfortunate tendency to see pedagogical methods as neutral to the subject matter, as "magic bullets" that can improve learning catholically regardless of the context (Eberlein et al., 2008).

In this paper, I will present three cases. In each case, I will identify the learning objective and the pedagogy we decided to use in order to achieve that goal, and provide a brief summary of the supporting details and rationale. Then, I will move to the aligned assessment method that we used to understand how well we did or did not achieve our outcome. The first case relates to the overall strategy we use to introduce students to the discipline through the organic chemistry subject matter; the second relates to the change we made toward more authentic, research-based laboratories; and the third relates to how students in the Structured Study Group program develop a higher order learning skill, namely, reflective self-assessment.

Case One: An Introduction to the Discipline

Objective: Modernize the Introduction to Chemistry
Pedagogy: Use Mechanistic Organic Chemistry

In an essay titled "Organic Chemistry in the Introductory Course 2. The Advantages of Physical Organic Chemistry" (Coppola, 1997), we argued that a mechanistic approach to organic chemistry instruction was needed to move beyond the historically relevant functional group organization, because the field itself has done so:

> Traditionally, introductory organic chemistry has been presented from the perspective of synthetic transformations. A representative sampling of early twentieth century textbooks indicates a course where the laboratory played a prominent role, where issues of separation, isolation and identification by qualitative chemical testing schemes were integrated throughout the presentation. The functional group organization, first introduced by Conant

in 1928, was an effort to bring introductory organic chemistry instruction into line with the contemporary practice. The functional group approach was well established in research by the time of the 1928 publication date. In the preface, though, Conant is almost apologetic to instructors for the changes he introduced:

"The formal classification of compounds which is so valuable to the specialist may be barren to the uninitiated.... The author's experience...has led him to believe that the alcohols have certain advantages over the hydrocarbons as a point of departure..."

Conant helped move introductory organic chemistry instruction out of the nineteenth century just as the development of mechanistic organic chemistry began to advance rapidly. The notion of chemical structure was dramatically affected by the coupling of a general acceptance of the electronic structure of matter and the corresponding understanding of bonding. The first quarter of the twentieth century brought together progress in creating useful models for chemical bonding with a deeper structural understanding of the compounds of main group elements and their transformations. In the second quarter century, the application of physical chemistry to the problems of organic reactivity created a remarkably comprehensive and unifying conceptual framework. Understanding improved, the reliability of predicting new outcomes increased, and rational synthetic design emerged. (p. 1)

Like many advances in a discipline, the more sophisticated organizing principles are fewer in number than the less sophisticated version (that is, a few types of bonding changes supplant hundreds of transformations based on functional group identity). This is not to say that functional group identifications are not important or useful, but rather that they are subsumed under a set of unifying principles (higher organizers) used by practicing organic chemists. These organizers allow chemists to understand new and unfamiliar information by permitting them to formulate analogies.

Assessment: Literature-Based Examinations

Thinking about testing is often overlooked in discussions about assessment at the postsecondary level. And yet, examinations, probably more than anything else, transmit our learning agenda to our students; they are truly "a latent curriculum" (Tobias & Raphael, 1995). If examinations are not aligned with learning goals, then efforts to teach effectively are ignored by the learners for whom they are intended. One motivation for the change we made was that organic chemistry is structured so that state-of-the-art information from the primary literature can be presented to novice students on examinations. This assures us that we are true to the facts of science and not simply inventing trivial derivatives of classroom examples. We include the citation along with some contextualizing statements, which sends two messages to our students.

1. Memorizing the previous examples is not enough.
2. Understanding the subject matter of the introductory course lets you understand some of what chemists actually say about what they study.

The context of these problems has a great deal of intrinsic interest or relevancy because many examples come from medicinal and pharmaceutical chemistry or materials science. Our examination questions are like short case studies that can be explored by 1,200 introductory chemistry students. We reinforce the idea of multiple representations for the same phenomenon. Students might be asked to provide words, pictures, graphs, and numerical versions of the same idea. On nearly every exam, students suggest unanticipated but completely reasonable alternative solutions. These are important to note in class.

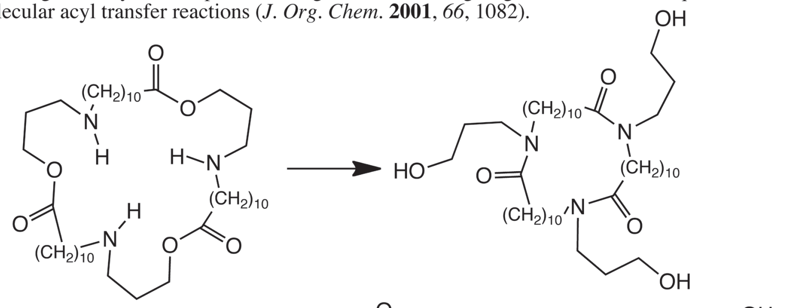
To support the testing implied by Figures 1 and 2, we have implemented the following practices:

1. *Make improvement count.*

In testing: because students develop their new skills at different rates, and because the course is truly cumulative each step along the way, we have devised ways to make improvement count. One simple but effective technique is increasing the point value of exams throughout the term without increasing the length of the exam. Our first exam is valued at 100 points, the second at 120 points, and the third at 140 points. It is worth more to do better later, so you do not have to be perfect at the outset, and practice has tangible value. It is likely that students overestimate the modest mathematical value of this scheme.

In assigning grades: we also gauge overall improvement in the class by arguing that there have been two independent measures of cumulative performance, namely, the average of the semester exams compared with the average on the final. We give the semester exams a flavor of formative assessment by considering that students whose final exam average is improved relative to their semester exams have arrived

The following macrocyclic compound undergoes an interesting ring contraction in a sequence of three intramolecular acyl transfer reactions (*J. Org. Chem.* **2001**, *66*, 1082).



Using the abbreviated structural drawings for the starting material and the product, shown on the right, provide a complete, stepwise mechanism for the acid-catalyzed version of this reaction. You may use H-A as your source of acid.

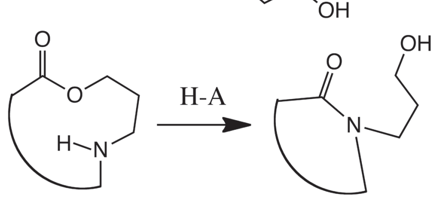
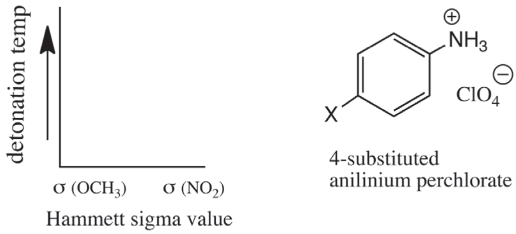


Figure 1. An example of a literature-based examination problem.

Mono-substituted anilinium perchlorates are known to detonate upon impact of the solid, or by heating (*J. Therm. Anal.* **1996**, *46*, 1751). A Hammett (physical organic) plot of the detonation temperature versus sigma values led the authors to conclude that proton transfer from the anilinium group is the rate determining step in the decomposition of these compounds. The more acidic the compound, the lower the detonation temperature.



Based on this information (a) Is the rho (ρ) value for this plot expected to be positive or negative; explain fully.

(b) Explain the relative acidity of the X=NO_2 compound to the X=H compound using words and structural formulas.

Figure 2. An example of a literature-based examination problem.

at their final numerical average through a different path than a person whose performance was flat (i.e., while getting E1=45%, E2=70%, E3=80%, and FE=90% gives, in our class, an overall average of 76%, this student has reached this point quite differently that a student who, for the sake of comparison, scored 76% on all 4 exams. These two students would get different grades assigned to them in our class).

2. *Use an absolute scale.*

Setting an absolute scale means more than saying 90–100% is an "A" grade. Our system depends on the fact that we give common examinations and fundamentally agree on course standards. These standards were determined empirically. By the third year of *Structure and Reactivity*, we had enough experience with offering the course and giving our examinations that we were able to set rough guidelines for performance based on the correlation of numerical values with the rich and informative student work presented to us on their papers. Such a system would not be easy with multiple-choice examinations. We have set our examination standards high, and we are comfortable with what achievement above (or below) certain levels tells us about student performance.

3. *Involve students in the process.*

We have used a technique that attempts to demystify the grading process for our undergraduate students. During the grading session for the first examination, I look for two problems with high variations in student responses. Before they are graded, I copy the student responses (four to six for each of two problems). I then combine these into a one-page, two-sided handout with all identifiers of the originators removed. During class the next day, and prior to posting the exam key, I use the first 25 minutes in an analysis of this handout. The total point values are still associated with the problems because they appear on the page. I direct the students to work in small groups, to consider the answers to these problems and to create a fair grading scale given the point values. This is, of course, exactly what the instructors have done prior to the grading session, and we are inviting my students to participate in an important part of the process. After 10 minutes, I call for the grading schemes and bring this discussion forward. The students invariably converge on the scheme that the instructors created the previous evening within a point or two. In the remaining class time, I give the final grading scheme for these two problems and direct the groups to actually assign scores, again, so that they can get a sense of the issues that we instructors face in looking at student work.

4. *Provide an extensive course pack of old exams (with no answers) and accompanying essays for effective use.*

Having old exams available for practice is not a revolutionary idea. It is fair for students to see representations of the style of examinations that will be quite different from their high school experience. There are two aspects of this practice that have been crucial for us. First, as described above, we use the primary literature as our principal source of examination questions. We quite deliberately select examples for students to elaborate on that do not match the examples from either the text or class. We want to communicate as clearly as possible to our students that we want them to learn how to extrapolate their understanding to new and unfamiliar examples.

We self-publish a course pack, available at our bookstores, that is about 175–200 pages long. A 20-page essay is included that gives an overview of what we have learned about student learning in this class (from our students, including through research studies), followed by four sections of about 40 pages each of representative pages from the four exams given over a five to six year period. In order to reinforce our belief in the value of developing teaching skills, we encourage our students to use the course pack as a way to catalyze conversations and discussions starting the first few weeks of class. This encouragement also comes by not providing a solutions manual. This makes our students very uncomfortable for a while, but we have them return to the essays and discuss this philosophy in class.

We issue a new edition of the course pack every year, replacing enough of the old problems so that students (and other organized student groups) that want to market their own solutions manuals are frustrated in their attempts.

Case Two: The Goals of a Laboratory Program

Objective: Understanding the Nature of Science
Pedagogy: Authentic Laboratory

As described in detail elsewhere (Coppola, 2010; Coppola, Gottfried, Gdula, Kiste, & Ockwig, 2006; Coppola & Lawton, 1995; Ege et al., 1997), we adopted a research-based orientation to our laboratory program. We took traditional technique-only exercises and re-imagined them as tasks with a comprehensible problem that contained a truly unknown feature. We recognized that an unknown in research did not need to be a large item—just authentically unknown.

For example, instead of presenting students with a compound (or even compounds) and a set of instructions for manipulating those compounds whose pedagogical end was only learning how to purify it and then collect chromatographic and spectroscopic data on it, we gave purpose to the gathering of data and posed a question that only the gathering and comparing of data, by the students, could answer. Into any given laboratory section of 24 students, we carry 30 or so vials of powdered, white, identical-looking solids. There are up to three vials of any given substance in any set, and the sets vary from lab room to lab room. Each vial is separately coded, and the code, only known to the personnel in the stockroom, is purposefully not revealed to any of the instructors. Individual students gather a cluster of experimental data (the exact cluster being determined by the class), in response to the single posed question: who else in class has the same substance that you do? The problem is comprehensible, it is authentic and uniquely driven only by the community of 24 students and the vials they have selected, and it cannot be solved unless and until the students devise ways to communicate their individuals results to each other, as a group, and inevitably struggle with important questions such as, "Is 150–151 degrees on my thermometer the same as 146–149 degrees on yours, given that the next highest melting group is in the 120s?" And the possible answers to that question, for instance, side-by-side analysis and/or mixed melting points, are exactly what any expert would need to do to answer that question.

We have introduced, by the second semester, some authentic research tasks. We have, for instance, distributed a recent research paper in which a certain chemical transformation is reported on a series of 10 substrates. If it looks as though it is the sort of procedure that could be carried out by large numbers of students in an undergraduate laboratory setting, then we will buy the reagents as well as a subset of the 10 reported substrates. In addition, we will buy a set of four to eight other substrates, not reported by the authors, but which one would reasonably predict ought to work under the same conditions. As a multiweek activity, we ask the students to (a) reproduce one of the literature examples, to be sure they have the skill set to do so, and then (b) select one of the new substrates and test it out. With hundreds of students focusing on a few new substrates, a statistical look at this new procedure emerges, and the students are truly carrying out new experiments in their introductory-level laboratory class.

Assessment: Performance-Based Task

In order to gauge the effectiveness of our new approach, when we introduced it, we collected data on how the skills of groups of students from the first *Structure and Reactivity* classes compared with those of students from the traditional sophomore organic laboratory course. During the three-year phase-in of the new program and phase-out of the old, both populations were in our department at the same time.

We used responses to a performance-based interview about an approach to solving a laboratory task. We conducted interviews with three groups of individuals. None of these groups knew of the study beforehand. The first group comprised randomly selected students from a section of the *Structure and Reactivity* course on a day during the last few weeks of class. These were first-year chemistry students. The second group comprised randomly selected students from a section of the traditional organic chemistry laboratory course during that same week. Although these latter students had had two full years of chemistry, they were the only legitimate comparison group because of their experience in organic chemistry. The third group was composed of five experts (two upper level graduate students and three faculty members, all organic chemists). We looked at how the two groups of student responses compared with the expert responses. The method of basing an analysis on concept maps had precedence and suited our purposes (Markham, Mintzes, & Jones, 1994; Wallace & Mintzes, 1990). The concept map (Figure 3) compiled from the responses of the five experts to the solution of the laboratory problem, described below, served as the basis for the comparison.

In the interview room, a small, capped vial containing about 5 mL of a clear, colorless liquid (dichloromethane) was placed next to a tape recorder. When the interview began, the subject was asked a version of the following query: "What stepwise procedure would you use to determine the nature of the material in this vial?" The interviewer challenged the responses in this think-aloud format by (a) questioning the significance of the suggestion ("What will you learn?"); and (b) offering that the suggestion led to a new problem, and asking how it might be resolved or reconciled ("That didn't work, what next?").

A feature of the solution to the problem compiled from the responses of the experts (Figure 3) is the sequence of four main components of an ordered process: (a) analysis, (b) separation, (c) purification, and (d) identification (hereafter referred to as the four "general concepts"). Appended to each of these are the more specific concepts and practices. There are a total of 47 entries on the expert's concept map. The students' interviews were transcribed, and the transcripts were used to identify which components of the expert concept map were present in the students' statements. Two representative student maps, one from each of the comparison groups, are shown as Figures 4 and 5. Three features from the student interviews were noted: (a) using a copy of the expert map as a template, the map entry was marked off when the student described the same feature. In all cases, the specific practice must have been mentioned in order for it to be marked off, while the more general concept ("analysis" "identification") might be inferred from the detailed description. (b) The original task also required description of a stepwise procedure. The chronological sequence of the general concepts used to describe the process, as suggested by the student, was also noted on the template. (c) When students suggested ideas not found on the expert map, these were mapped onto the template and counted separately.

One way to express the development of skills is the progression from novice to expert (Bowen, 1994; Bruer, 1993). Although true "expertise" is an amalgam of expert skills, appropriate and highly integrated prior knowledge and experience, as well as the knowledge of what skills and information are needed in a given situation, the students in the new first-year course appeared to hold a more "expert" conception of the task that they were assigned than the students from the traditional course.

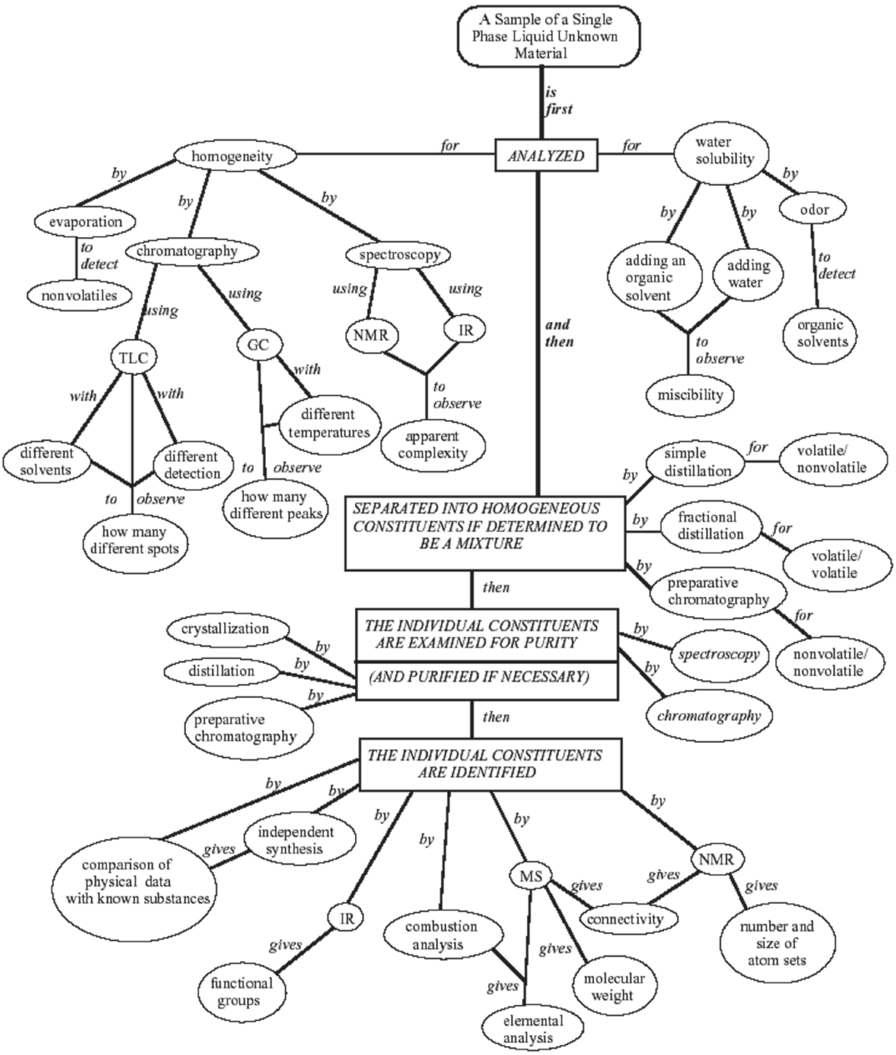


Figure 3. Concept map compiled from 5 experts.

1. *Chunking like the experts.*

Experts deal with complex tasks involving lots of declarative knowledge by chunking it and accessing it as needed (Gobet et al., 2001). Nearly all of the *Structure and Reactivity* students saw this as a complex task: (20/22) used three or four of the four general concepts, and the majority of them (17/22) used the expert procedural order. The students from traditional course were mainly focused on the identification aspect of the task. When they used an analysis step, they were all using

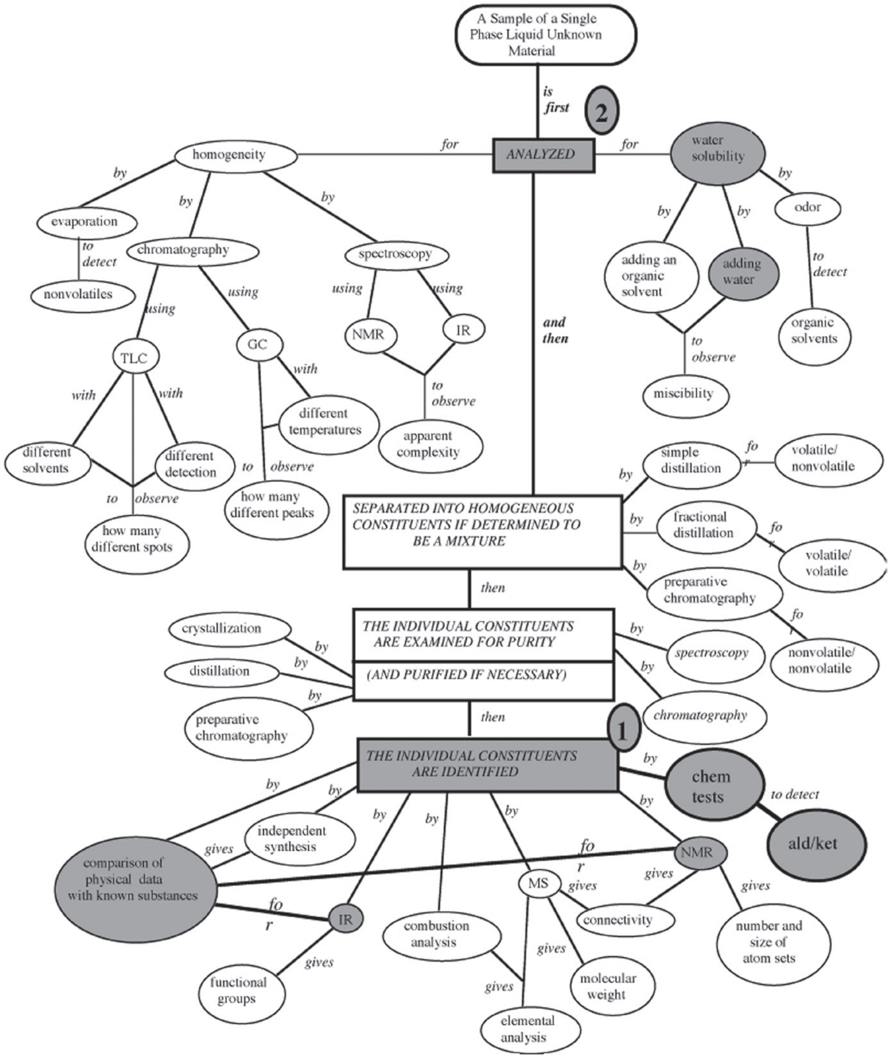


Figure 4. Analysis of a student response (traditional class; from a group of 19).

water solubility as structural evidence; *not one of these students explicitly considered the homogeneity of the sample* (left-hand branch on analysis concept). On the other hand, all of the *Structure and Reactivity* students who considered an analysis step (20/22) included an analysis of the homogeneity as part of their suggested solution.

2. *Having a repertoire of options.*

The average number of expert items that the *Structure and Reactivity* students matched was nearly three times greater than the matches demonstrated by the

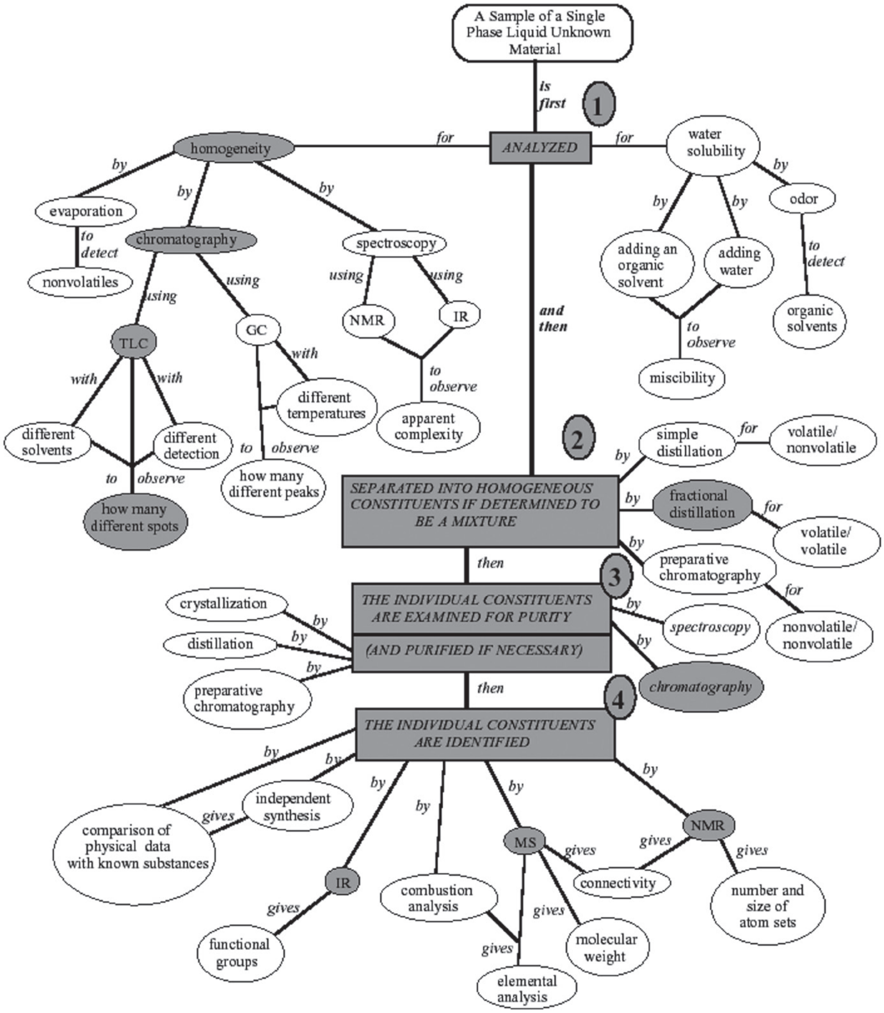


Figure 5. Analysis of a student response (Structure/Reactivity; from a group of 22).

students in the traditional course. In one study, traditional students matched 5.9 ($\pm$ 3.4), while the *Structure and Reactivity* students matched 16.1 ($\pm$ 3.2). In a separate, follow-up study the next year, traditional students matched 4.7 ($\pm$ 2.4), while the *Structure and Reactivity* students matched 13.5 ($\pm$ 4.3). Note that the noninstructor experts provided 17, 20, 25, and 29 entries, respectively.

3. *Today's answer, not yesterday's answer.*

One of the experiments in the traditional course was the qualitative identification of an unknown aldehyde or a ketone by chemical tests and the preparation of a solid derivative, a technique that, while still popular in the undergraduate teaching program, is not an experimental technique used in research since the late 1950s. Yet, the additional items suggested by 12/19 these students in the traditional class revolved around this theme, of course, because it was what they knew. Their answers were correct, and even thorough, in that context; but these are not the answers that any contemporary expert gives. The *Structure and Reactivity* students, who had routine access to FT-IR, GC, and FT-NMR data throughout the year, reflected their comfort with the instrumentation techniques by suggesting this kind of analysis as their primary strategy.

Case Three: Promoting Higher Order Learning Skills

Objective: Reflective Self-Assessment
Pedagogy: Structured Peer Review

Reflective self-assessment (Boud, 1995) is a high-level skill for learners that might be approximated by the ability to edit one's own work, to be able to look at it with critical eyes that are external to your own. We know this is an important skill that is challenging to develop. One vehicle for developing reflective self-assessment is through teaching, because you think differently about your knowledge when you anticipate the need to teach others compared to when you are aiming for private, personal knowledge (Coleman, Brown, & Rivkin, 1997).

The antecedent for this idea can be found in a strategy called "reciprocal teaching." Reciprocal teaching is an instructional strategy that was developed to improve reading comprehension in young (elementary and middle school) students (Brown & Palincsar, 1989; Palincsar, 1986; Palincsar & Brown, 1984; Palincsar & Klenk, 1991).

Palincsar (1986) describes reciprocal teaching as "an instructional activity that takes place in the form of a dialogue between teachers and students regarding segments of text. The dialogue is structured by the use of four strategies: summarizing, question generating, clarifying, and predicting. The teacher and students take turns assuming the role of teacher in leading this dialogue." In addition, "the purpose of reciprocal teaching is to facilitate a group effort between teacher and students as well as among students in the task of bringing meaning to the text" (p. 15).

Reciprocal teaching provides a menu of structured tasks that makes explicit the process used by good comprehenders (and good teachers). In a wide variety of carefully controlled studies, reading comprehension (making meaning from information) is improved by using reciprocal teaching.

In their research on college-level biology, Coleman et al. (1997) write: "Past research has shown positive effects on learning of both explanation and summarization. However, no study has examined the effects of explanation or summarization on a live audience. Also, there has not been a direct comparison of the two, and no research has been done on how explanation and summarization may cause different types of learning for the explainer and for the hearer" (p. 347). One of the conclusions they could draw was that students who read a text with the idea that they were to provide explanations to "their students" could respond more successfully to new questions about the reading (involving synthesis and extrapolation, so-called "far transfer problems") than students who read with the idea that they were to provide summaries to "their students." In their studies, they point to the pathway to developing Explanatory Knowledge: "Preparation to teach the contents of a text to another versus to understand it personally, may influence the mental representations that are created from text" (p. 347).

In designing the Structured Study Group assignments, we coupled notions of reciprocal teaching, explanatory knowledge, and peer review and critique in order to create an environment where the generation of a solution to a assigned task would be the beginning point—and not the typical end point—of thinking about a problem and its underlying lessons.

The SSG assignments typically involve generative activities in response to tasks that can diverge through personal creativity rather than converge onto a prescribed, concealed answer. In the very first SSG assignment, students pick a C_{10}-C_{13} molecule from a chemistry journal (after learning, in their session, how to decode line formulas, what journals are, where they are found, and what a proper citation format is) and are directed to construct (design and draw) five rational examples of molecules with the same formula. They then propose rankings for their created molecules based on 3 of 6 properties, including, for example, magnitude of dipole moment, boiling point, and solubility. They must also include written descriptions of their rationales.

At the beginning of the session, each student submits one copy of his or her work to the SSG leader, and the other copy is distributed to the class. One or two rounds of peer review follow. The reviewer does not correct the other student's paper, but rather answers a set of factual questions about the other's work: Does the molecule or reaction fit the prescribed criteria (yes or no?); is the format and information appropriate to the level of the class (yes or no?); is the citation formatted correctly (yes or no?). During this time, the discussion within the group is free-wheeling, and it is the time of greatest learning for the students. Although the only duty is to mark off a "yes" or "no," the first round of peer review can take up to an hour. Only when faced with reviewing the work of another, can students deal with issues that were either incorrectly understood or that simply did not occur to them. These students have a structured opportunity to make, recognize, and correct their errors before they get to an examination. After the reviewing is completed, the reviews and the unmarked papers are returned to the originator, and he or she has a chance to decide if any corrections are needed. This set of assignments and reviews are collected, and they form part of the basis for the leader's evaluation of the student's performance on that day.

Assessment: Performance-Based Task

Do students who experience weekly self-reflective assessment of their work develop the skills associated with that practice? To test this, we performed a study using an interview-based format. Three groups of subjects (a group of faculty and graduate student experts, and two groups of students) were presented with information based on which a prediction was solicited. The two student groups only differed, to the best of our ability to identify, in whether or not they participated in the SSG work.

We acknowledged that the student groups had a different class experience–we wanted to see if we could detect any difference empirically. Recall that all of the students shared most of the same experiences: all were a part of the same large lecture class for their formal course work, discussions, laboratory work, and so on. A subset of students also participated in the two hours of SSG and did the associated work. We used background demographics and academic performance in the course (using examination scores) in order to create an appropriate comparison group. The study was carried out one month after the end of the semester.

In our study, our subjects were presented with a two-page problem. On the first page, they encounter the series of trimethyl Group IV substituent groups and are asked to predict the order of relative energy difference between the two chair forms of the monosubstituted cyclohexane derivatives (Figure 6). The nature of the given information is such that the most likely prediction will be the opposite of the experimental results, and this incorrect prediction might well be anticipated to be given by both "A" students and "C" students. In the presence of an interviewer, the responses of the subjects were tape recorded while they described their thought processes. Once a prediction was made and the subjects completed their elaboration of it, the subjects were instructed to turn the page. After confronting the actual experimental results (Figure 7), the subjects were instructed to judge how the experimental results matched their prediction. The interviewer ended the interview by prompting the subject with the question "... and how would you test your ideas?"

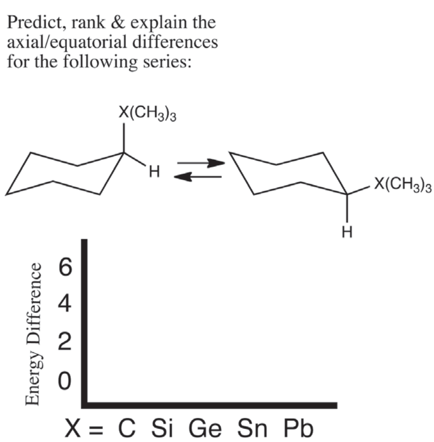


Figure 6. Page 1 of the counterintuitive task.

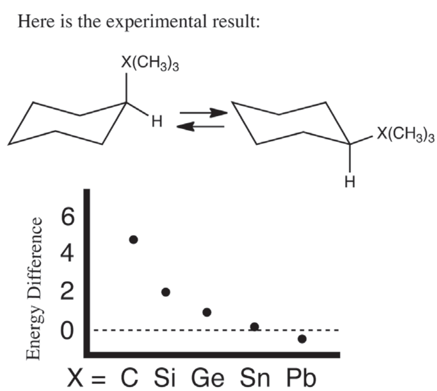


Figure 7. Page 2 of the counterintuitive task.

The prediction/evidence sequence presented in Figures 6 and 7 represents an example of a counterintuitive task (Alvermann & Hague, 1989).

Using the interviews themselves as the source of data, we applied Glaser's method of Grounded Theory Analysis (Glaser, 1992). We created categories for the activities in which the interview subjects engaged as they looked at each page of the problem (e.g., on page 1, "restate" means that the subject was restating the problem, and "S id" means that the subjects were identifying the substituent "X" groups). Similarly, we did this for the responses to the second page (e.g., "reflect" meant that the subject had identified a particular idea and was talking about it, "elaborate" meant that the subject was bringing in knowledge external to the evidence of the problem, and "reconcile" meant that the subject was trying to make the new information about the "X" groups from page 2 fit into their prediction from page 1). From this, we developed a timeline template (Figure 8) onto which we could then record the events that were happening in the student explanations as they started responding to page 1 and proceeded (Figure 9). We then coded the interviews according to what was being said at any given time, using a fully darkened mark if what was being said was correct as might be judged by a knowledgeable other, and a shaded mark if what was being said was incorrect.

Our expert group (N=6, 2 faculty and 4 midcareer graduate students, an example of the latter is shown in Figure 10T) demonstrated the following attributes: (a) all of them began by restating the problem; (b) all of them made a fairly early prediction after taking an inventory of the major factors related to the problem. This prediction was followed by a fairly extensive elaborative explanation; (c) except for the faculty member who was previously aware of the experimental results, the thought process used by the experts was cyclical: examination of an alternative model, rejection on the basis of a counter argument, and proposal of a new model; (d) upon prompting about how they would test their ideas, all of the experts relied on primary literature sources, the design of new experiments, and computational chemistry methods.

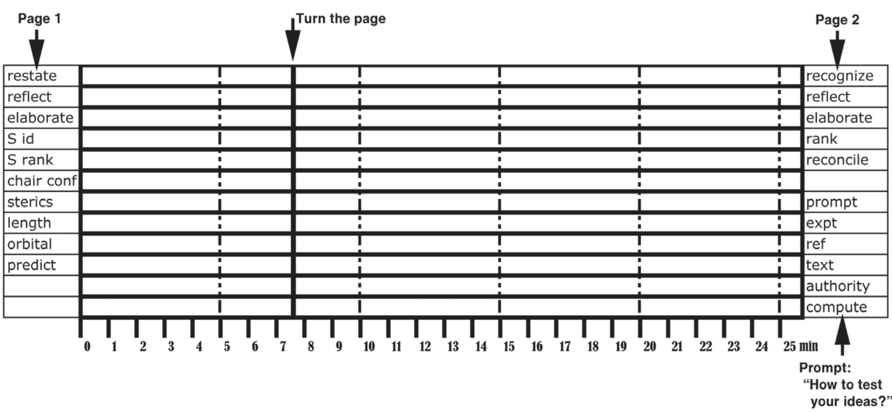


Figure 8. Template for coding the counterintuitive task interviews.

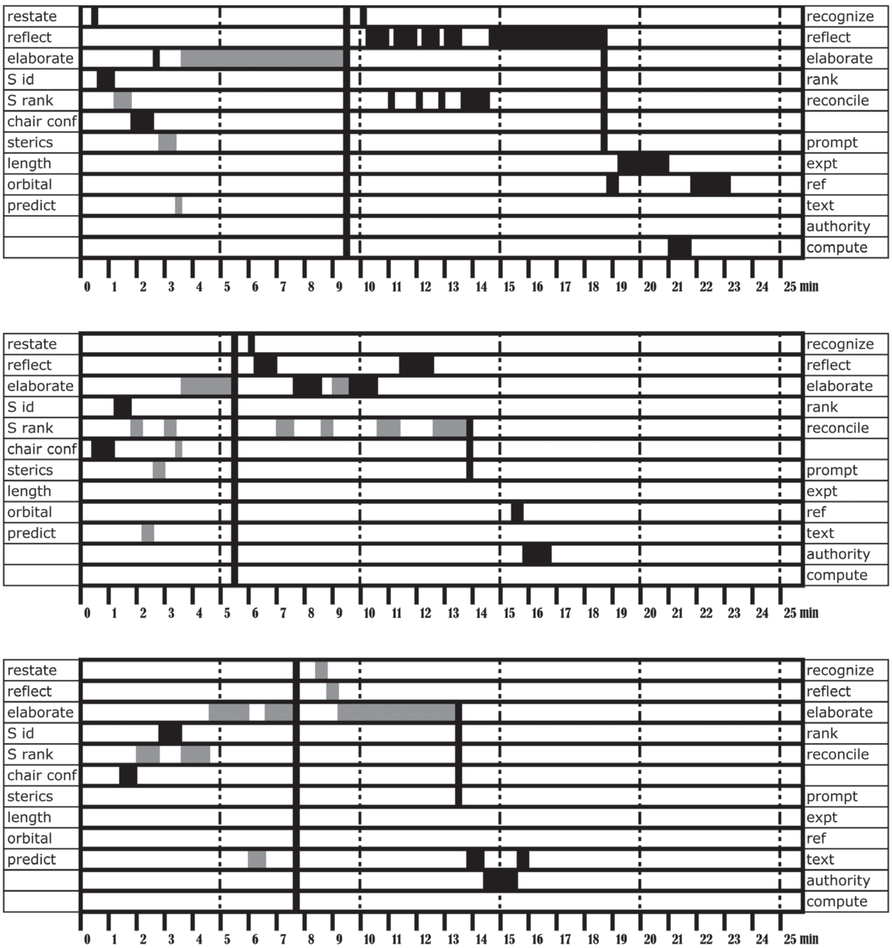


Figure 9. Representative coded interviews (top: 9T, a midcareer graduate student; middle: 9M, a student from the SSG program; bottom: 9B, a student who did not take the SSG option).

We interviewed 20 students from the SSG program and 20 students from the same class who did not opt for the SSG. While the grades of the SSG students reflected the distribution of the class as a whole, we intentionally only interviewed students with a "B+" grade or better.

We looked at these data in two ways. First we simply counted the incidents of expert behavior in the student subjects (Table 1). If a behavior was not observed in any of the student interviews, then we said the occurrence was "none"; if it was observed 1–6 times, then it was "few"; if 7–13 times, then it was "some"; if 14–19 times, it was “most”; and if all 20 times, then “all.” As can be seen in Table 1, the SSG students, more so than the non-SSG students, exhibited the characteristic behavior of the expert group.

Table 1

Comparison of Behavior Frequencies

used by all Experts	**SSG**	**non-SSG**
restating	none	none
early prediction	all	none
cyclical analysis	most	few
primary lit, new experiments	all	none
computational methods	some	none
not used by any Experts		
consult text, TA, Prof	some	all

In our second analysis, we explained the experiment to a group of six scientists and six nonscientists (faculty and students) and showed them the three most characteristic event recordings from the experts (i.e., Figure 9T and two others). We then gave them the 40 event recordings from the two student groups (20 SSG and 20 non-SSG), shuffled, and bereft of any identifiers. We asked these individuals to evaluate whether they thought the observed behavior, as evidenced by the pattern of the recorded event (i.e., Figure 10M, 10B, and 38 others), matched or did not match (a binary decision) the pattern of the expert set. At an average of 84% of the time, the SSG students’ patterns were matched with the experts, while only 10% of the non-SSG students were matched with this expert set—even though the non-SSG students were over-matched, based on exam performance, with respect to the SSG group.

This experiment suggests that the weekly assignments, wherein the SSG students brought generative assignments for peer review, critique, discussion and correction, developed in them a sense of reflective self-assessment on this content-based task that was more comparable to that of experts than the students who did not participate in SSGs. We have proposed that the key behavior seen in the recorded events (Figure 9T and 9M) is the ability to access a range of possible alternate explanations, test them out systematically, reject them when they lead to inconsistency, and then continue this cycle. The non-SSG students, in general, could identify a possible new explanations, but could not appropriately balance its implications against what was already known, thereby recognizing inconsistency and therefore did not show any ability to be able to reject the first thing they thought of (Figure 9B).

Conclusions & Implications

In implementing the *Structure and Reactivity* course sequence, we used a new curricular program in order to test our hypotheses about the higher level learning goals that we claimed were embedded in the mature subject matter of organic chemistry. Following an explicit notion of hypothesis testing, we were also part of the emergent national interest in developing and applying educational research methods to postsecondary classroom settings, which has only grown stronger over time under a number of guises: the Scholarship of Teaching and Learning (International Society for the Scholarship of Teaching and Learning [ISSOTL], 2010), Scientific Teaching (Handelsman et al., 2004), and Discipline-Based Education Research (DBER; National Science Foundation, 2010).

In this report, I have selected three from among a number of examples of assessments that we have carried out over the past 20 years. Let me reflect here on some of the themes that emerge from these cases.

Literature-Based Examinations

In building from the literature for constructing these exam problems, we are making the process of administering a 1,500-person test as true to an authentic disciplinary experience as possible: reading a journal article whose details are unfamiliar, but which can be understood by an application of general principles to the specific information. We have found that it takes 30–40 person-hours for a team of four faculty instructors to construct these examinations, plus the time contributed by three to five friendly collaborators who review and give feedback on drafts. Only a depth and breadth of subject matter mastery, combined with a consensus on the pedagogical design, allows us to share and critique openly as we converge on the final version of one of these tests.

We are trying to transmit to students as clearly as we can, including by the strategic inclusion of citations, that there are general concepts to be learned from the specifics in order to then apply them to new and unfamiliar situations. Well-designed examination questions avoid the "unfortunate coincidence," where getting a correct answer results from an incorrect pathway, or fuzzy logic (Davidson, Stickney, & Weil, 1980; Hoffmann & Coppola, 1996). If the correct answer can be produced, or selected, by simple decoding, pattern recognition, or memorization—without needing to follow a pathway in which the learner engages the underlying ideas—then two things happen: (a) getting the right answer for the wrong reason creates a sense of false confidence in the learner that productive learning is taking place (Baldwin, 1984); and (b) the learning that does occur is indistinguishable from nonsense (Gross-Glenn, Jallad, Novoa, Helgren-Lempesis, & Lubs, 1990; Redish & Smith, 2008).

Performance-Based Laboratory Task

Although it is tempting to see the study of laboratory skills as a direct comparison between an experimental group and a control group, it is not. Two different groups of students received different treatments, and so we expect differences in performance.

The question we wanted to answer required a point of reference: what is the external standard against which we can generate a value judgment about whether either of these groups of students was achieving the goal of learning about laboratory science?

My strategy, whenever possible, is to interrogate the discipline. In order to answer the question about whether either of these groups was learning chemistry, I first needed to ask what was chemistry's answer to the question. Thus, before carrying out the assessment task with the two student groups, we interviewed a group of graduate students and faculty members until we heard nothing else new in their replies, and we used an aggregate response from that group as our metric. This decision, to use the discipline itself as the point of reference, was not the only choice possible. We might have decided to ask our graduate and faculty respondents to answer the question as though they were undergraduate students in a traditional class, in which case their answers, and the resulting outcome, would have favored the other group.

We were able to learn, convincingly, that the undergraduate students in the new classes were solving the assigned task in a way that someone with much more experience in the discipline would answer it. Recently, a group at UC Berkeley has created a systematic way of measuring student performance against the perspectives of chemists, which they call a Perspectives model of assessment (Claesgens, Scalise, Wilson, & Stacy, 2008, 2009).

Performance-Based, Counter-Intuitive Task

As in the second case, there was no control versus treatment group, but rather two groups of students with a different set of experiences. Here, the group of students who participated in the Supplemental Instruction option did something extra.

It would be naïve to attribute any differences only to the structured instructional activities, however, because we know that the students who meet weekly in the Supplemental Instruction groups change a number of behaviors. Most importantly, they begin to associate with each other as a mutually supportive study group for much more than their assignments in this program. Yet, the group of interest, participating in activities in which they were critiquing the work of others in order to reflect on their own work, showed a pattern of thinking about the posed counterintuitive task that was unlike that of their peers and more like the pattern seen in more expert chemists.

In these three examples, I elected to emphasize the role that disciplinary expertise has played in developing, implementing, and understanding these assessments. There are other assessment strategies. We have carried out large-scale survey work using existing instruments (Zusho, Pintrich, & Coppola, 2003) as well as those we created for specific purposes (Kiste, Coppola, Lomont, Rothman, & Zhang, in press), and we have been the subjects of studies carried out by others. I have also attempted to illustrate the principle of alignment between our stated learning goals, our pedagogical approach to achieving those goals, and the assessment method that we used to evaluate our outcomes.

The broader implications from our experiences fall into a few categories. First, in addition to being discipline-centered assessments, all of the examples suggest that introductory science instruction can be anchored in active, contemporary ideas that represent the work of the science as the practicing scientists know it—in contrast with a common, fixed set of facts and procedures, calcified into the introductory program by whatever mechanisms operate to do so. The assessments do not point to how this might be done, however, which is a larger and more complex behavioral question about the use and reward of faculty time, and the collegial organizational structure of university departments.

Second, a type of traditional assessment, namely, an examination, was selected in order to emphasize that testing, more than anything else, transmits the goals and expectations that we have, as instructors. If, after all the classroom talk about critical thinking and reasoning as learning goals, students discover that memorization and pattern-recognition serve them, then the exam is not aligned with the goals; there is, at best, a hypocrisy that results from this misalignment.

Performance-based assessments provide rich and interesting information, but they are labor-intensive and difficult to implement on a large scale, and they require productive interdisciplinary collaboration between science and education. Improved test performance (getting, or selecting, the single, right answer), which is the ubiquitous method for evaluating instructional interventions, can produce compelling comparative data (Hake, 1998). The challenge for researchers in fixed-response methods of assessment is that the pathway is inferred: there is no direct evidence to differentiate deeper understanding of the subject from improved test-taking skills (Johnstone, 2003).

Third, education is not carried out in a neutral environment, nor is it a natural phenomenon, so studying teaching and learning have all the interlocking complexities of any social science experiment. Data and its analysis arise from assessments, but the result is tied strongly to the circumstance of the particular classroom, its instructor, its students, its institutional context, and so on. Data are not enough:

> Pedagogical innovation requires changes in faculty behavior, the most difficult change of all. It is the difference between knowing (intellectually) that a good diet and regular program of exercise are truly the right things to do and observing that the world has plenty of overweight, sedentary physicians who also smoke. Behavioral changes are more complex and difficult than just changing one's mind. (Coppola, 2001a, p. 70)

I would now add to this that pedagogical innovation also requires changes in student behavior, based on student expectations, and should have also been included in that passage.

Lastly, the centrality of the discipline is evident in these examples: in all three cases, the expertise of an organic chemist is needed in order to carry out the work. Yet, in order to implement the research on student learning reported here, being an organic chemist is not nearly enough. Interdisciplinary collaboration, which is so commonplace between chemists and their colleagues in physics, medicine, biological and life sciences, engineering, etc., as they take on complex research problems, is also the key to doing research in discipline-centered teaching and learning, a term I prefer to the others that are used.

A key feature in our work has been the open, productive collaboration between faculty members and students in science, science education, and related fields, on projects of mutual interest. Faculty colleagues in education and the learning sciences (psychology, educational psychology, anthropology, cognitive science, etc.) bring long-standing knowledge and traditions to design, carry out, and analyze the results from relevant experiments. But, they generally do not know the details of the physical sciences any more than a physical scientist knows about social science research, and there can be an unfortunate tendency for scientists to outsource the work to their colleagues ("do this and get back to me"), or, worse, to work in relative isolation reinventing naïve versions of what is already known how to do better. The last and perhaps most important implication from our work, then, is having the sort of institutional structures, including a broadly defined and supportive environment for interdisciplinary collaboration, that can bring researchers together to advance our understanding of postsecondary teaching and learning in the sciences.

Acknowledgements

The author expresses deep and sincere gratitude to the colleagues and collaborators who contributed to the work summarized in this manuscript, all of whom are co-authors on publications that can be found in the reference list. Additional thanks are extended to the thousands of students who have enrolled and excelled in the *Structure and Reactivity* sequence since its inception in September, 1989, and to the faculty colleagues who have been welcomed into the fold and have continued the work we started back then.

REFERENCES

Alvermann, D. E., & Hague, S. A. (1989). Comprehension of counterintuitive science text: Effects of prior knowledge and text structure. *Journal of Educational Research, 82*(4), 197–202.

Baldwin, B. A. (1984). The role of difficulty and discrimination in constructing multiple-choice examinations: With guidelines for practical application. *Journal of Accounting Education, 2*(1), 9–18.

Boud, D. (1995). *Enhancing learning through self-assessment*. New York: Routledge.

Bowen, C. W. (1994). Think-aloud method in chemistry. *Journal of Chemical Education, 71,* 184–190.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393–451). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bruer, J. T. (1993). The mind's journey from novice to expert. *American Educator, 17*(2), 6–15, 38–46.

Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2008). Assessing student understanding in and between courses in chemistry. *Assessment Update, 20*(5), 6–8.

Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry. *Science Education, 93*(1), 56–85.

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences, 6*(4), 347–365.

Coppola, B. P. (1997). Organic chemistry in the introductory course 2. The advantages of physical organic chemistry. *Chemical Educator, 2*(2): S1430-4171(97)02113-4. Retrieved from http://chemeducator.org

Coppola, B. P. (2001a). Full human presence. In A. G. Reinarz & E. R. White (Eds.), *Beyond teaching to mentoring* (New Directions in Teaching and Learning, No. 83, pp. 57–73). San Francisco: Jossey-Bass.

Coppola, B. P. (2001b). Structured study groups (SSGs): Using peer-facilitated instruction to develop self-assessment skills. In E. D. Siebert & W. J. McIntosh (Eds.), *College pathways to the science education standards* (pp. 63–65). Arlington, VA: NSTA Press.

Coppola, B. P. (2010). Laboratory instruction: Ensuring an active learning experience. In M. Svinicki & W. J. McKeachie (Eds.), *McKeachie's teaching tips* 13e (pp. 280–290). Belmont, CA: Wadsworth.

Coppola, B. P. (in press). Two cultures: Characterizing convergent and divergent assignments. *Journal of Advanced Academics*.

Coppola, B. P., Daniels, D. S., & Pontrello, J. K. (2001). Using structured study groups to create chemistry honors sections. In J. Miller, J. E. Groccia, & D. DiBiasio (Eds.), *Student assisted teaching and learning* (pp. 116–122). New York: Anker.

Coppola, B. P., Ege, S. N., & Lawton, R. G. (1997). The University of Michigan undergraduate chemistry curriculum. 2. Instructional strategies and assessment. *Journal of Chemical Education, 74,* 84–94.

Coppola, B. P., Gottfried, A. C., Gdula, R. L., Kiste, A. L., & Ockwig, N. W. (2006). The great Wakonse earthquake of 2003! A short, problem-based introduction to the titration concept. *Journal of Chemical Education, 83,* 600–603.

Coppola, B. P., & Lawton, R. G. (1995). "Who has the same substance that I have?" A blueprint for collaborative learning activities. *Journal of Chemical Education, 72,* 1120–1122.

Davidson, S., Stickney, C. P., & Weil, R. L. (l980). *Intermediate accounting concepts: Methods and uses* (pp. viii–ix). Fort Worth, TX: Dryden Press.

Eberlein, T., Kampmeier, J., Minderhout, V., Mood, R. S., Platt, T., Varma-Nelson, P., et al.. (2008). Pedagogies of engagement in science: A comparison of PBL, POGIL, and PLTL. *Biochemistry and Molecular Biology Education, 36,* 262–273.

Ege, S. N., Coppola, B. P., & Lawton, R. G. (1997). The University of Michigan undergraduate chemistry curriculum 1. Philosophy, curriculum, and the nature of change. *Journal of Chemical Education, 74,* 74–83.

Glaser, B. (1992). *Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5,* 236–243.

Gross-Glenn, K., Jallad, B., Novoa, L., Helgren-Lempesis, V., & Lubs, H. A. (1990). Nonsense passage reading as a diagnostic aid in the study of familial dyslexia. *Reading and Writing, 2*(2), 161–173.

Hake, R. (1998). A six-thousand-student survey. *American Journal of Physics, 66,* 64–74.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Scientific Teaching. *Science, 304,* 521–522.

Hoffmann, R., & Coppola, B. P. (1996). Some heretical thoughts on what our students are telling us. *Journal of College Science Teaching, 25,* 390–394.

International Society for the Scholarship of Teaching and Learning. (2010). Retrieved from http://www.issotl.org/

Johnstone, A. H. (2003). *Effective practice in objective assessment*. Hull, United Kingdom: LTSN Physical Sciences Centre.

Kiste, A. L., Coppola, B. P., Lomont, J. P., Rothman, E., & Zhang, L. (in press). Use and usefulness: Examining the ecology of the classroom. In D. Bunce (Ed.), *Debugging the myths about teaching and learning* (American Chemical Society Symposium Series). Washington, DC: ACS Publications.

Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). *Journal of Research in Science Teaching, 31,* 91–101.

National Science Foundation. (2010, February 4). *Strengthening undergraduate and graduate STEM education.* Testimony before the House Committee on Science and Technology Subcommittee on Research and Science Education. H.R., 111th Cong. Retrieved from http://www.nsf.gov/about/congress/111/jfm_stemeducation_100204.jsp

Palincsar, A. S. (1986). Reciprocal teaching. In A. S. Palincsar, D. S. Ogle, B. B. Jones, & E. G. Carr (Eds.), *Teaching reading as thinking*. Oak Brook, IL: North Central Regional Educational Laboratory.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175.

Palincsar, A. S., & Klenk, L. J. (1991). Dialogues promoting reading comprehension. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), *Teaching advanced skills to at-risk students* (pp. 112–130). San Francisco: Jossey Bass.

Porter, A. C., Smithson, A., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education, 20*(1), 27–51.

Redish, E. F., & Smith, K. A. (2008). Looking beyond content: Skill development for engineers. *Journal of Engineering Education, 97,* 295–307.

Tobias, S., & Raphael, J. (1995). In-class examinations in college science—New theory, new practice. *Journal of College Science Teaching, 24*(4), 240–244.

Varma-Nelson, P., & Coppola, B. P. (2005). Team learning. In N. Pienta, M. M. Cooper, & T. Greenbowe (Eds.), *Chemist's guide to effective teaching* (pp. 155–169). Saddle River, NJ: Pearson.

Wallace, J., & Mintzes, J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching, 27*(10), 1033–1052.

Zusho, A., Pintrich, P. R., & Coppola, B. P. (2003). Skill and will: The role of motivation and cognition in the learning of chemistry. *International Journal of Science Education, 25,* 1081–1094.

The *Assessment in the Disciplines* volumes provide assistance to faculty who have responsibility for assessing their academic programs and institutional researchers who report on student learning assessment activities. The discussions presented in this series will contribute to the development of assessment strategies to improve student learning across all sections of postsecondary education.

AIR
Association for Institutional Research

ISBN 978-1-882393-19-0