Assessment of Chemistry

THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

1435 E. Piedmont Drive, Suite 211
Tallahassee, FL 32308

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.

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.

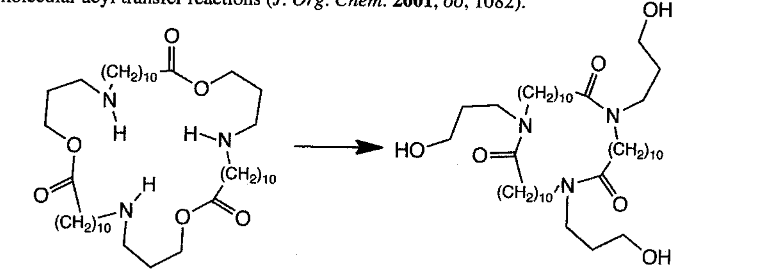
To support the testing implied by Figures 1 and 2, we have implemented the following practices:

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.

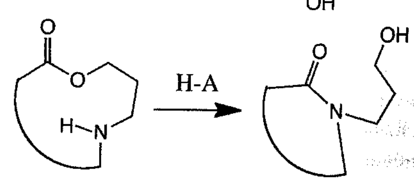
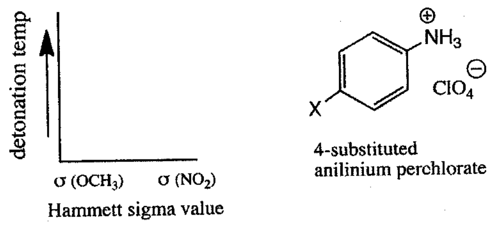


Figure 1. An example of a literature-based examination problem.

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.

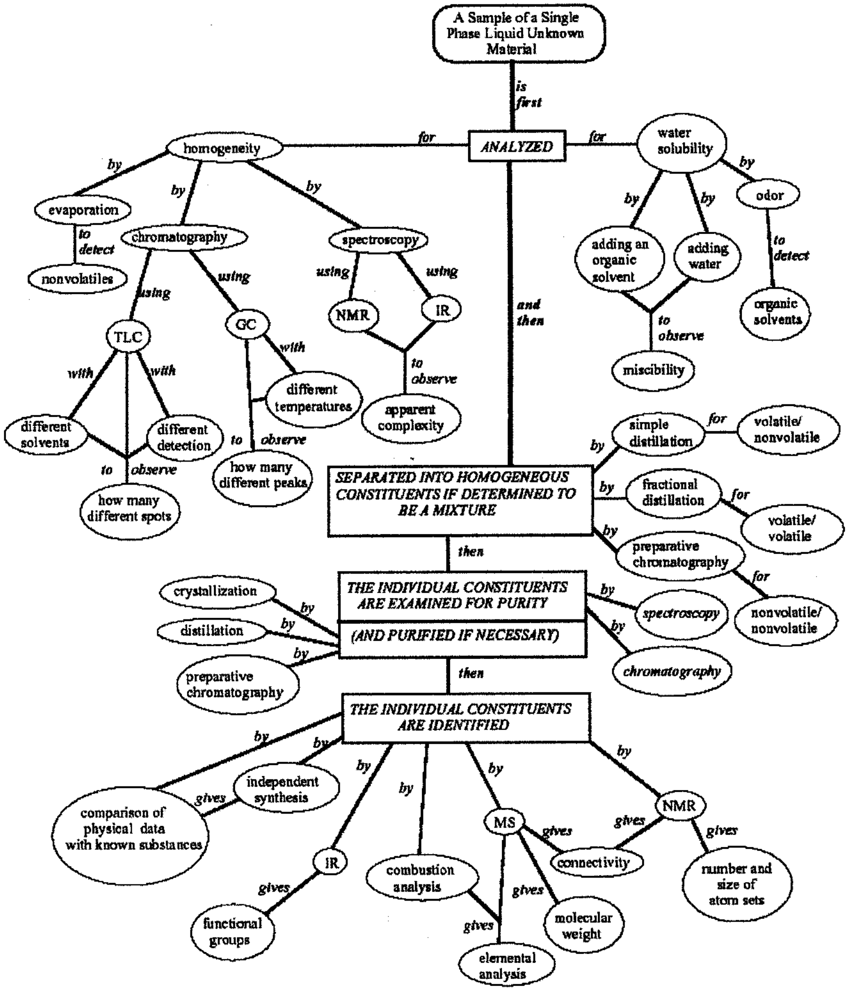


Figure 3. Concept map compiled from 5 experts.

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

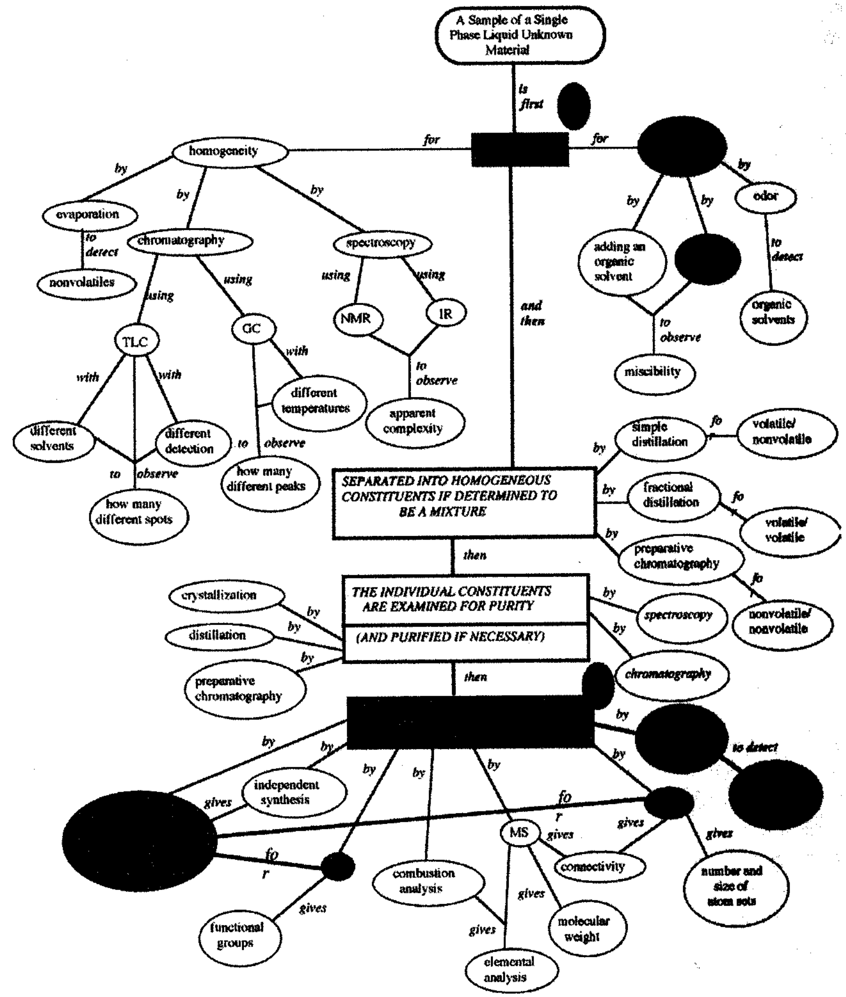


Figure 4. Analysis of a student response (traditional class; from a group of 19).

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

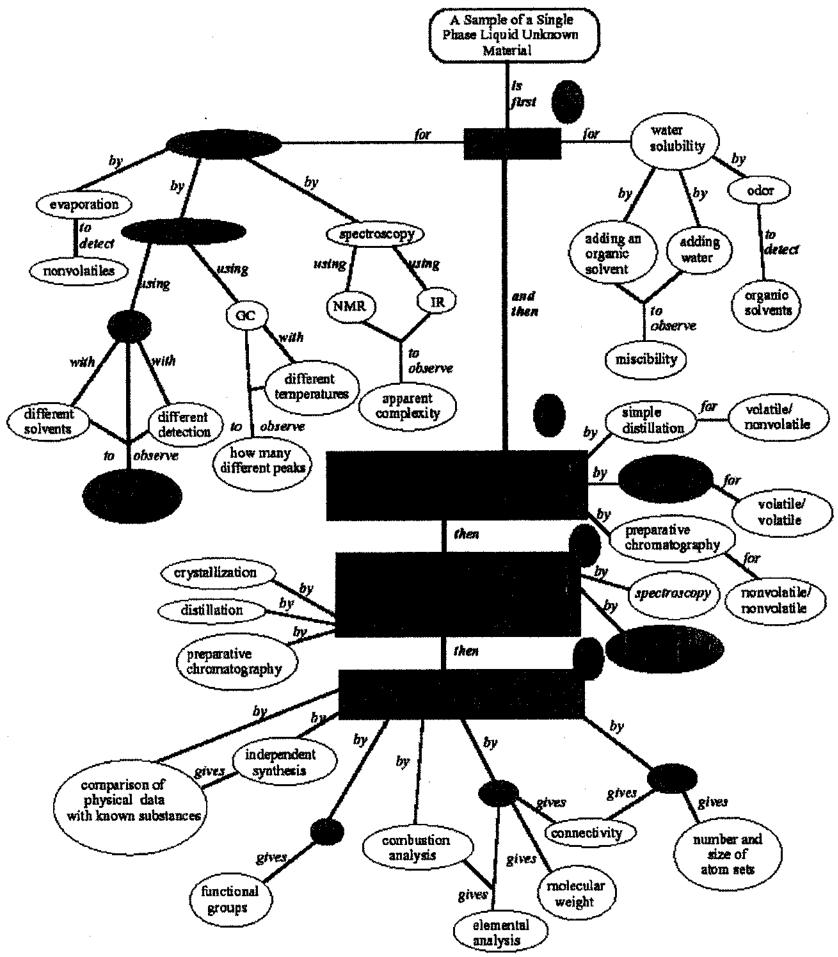


Figure 5. Analysis of a student response (Structure/Reactivity; from a group of 22).

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?"

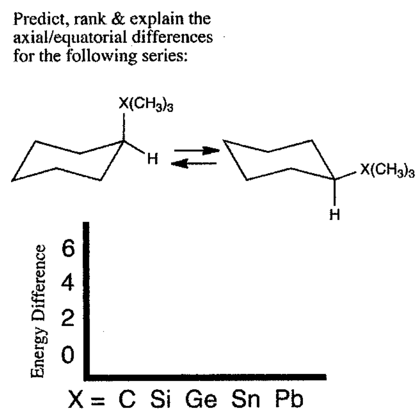


Figure 6. Page 1 of the counterintuitive task.

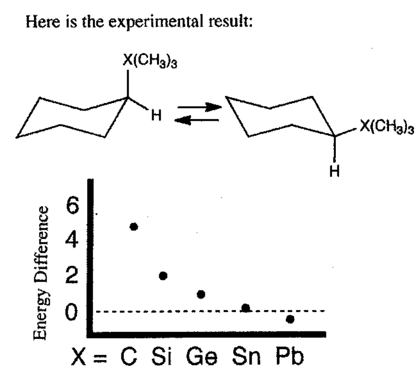


Figure 7. Page 2 of the counterintuitive task.

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.

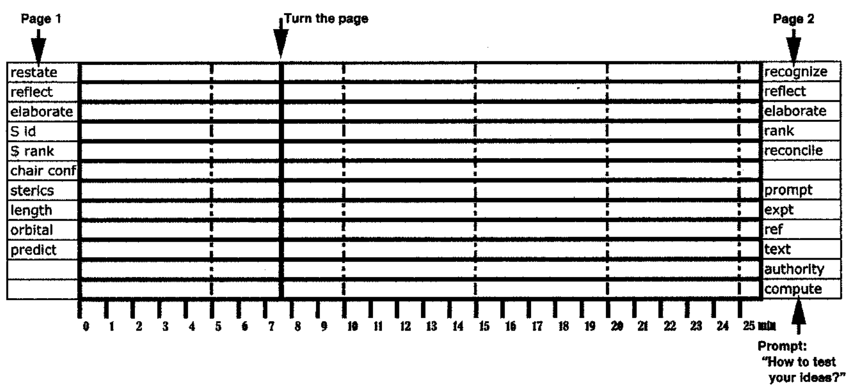


Figure 8. Template for coding the counterintuitive task interviews.

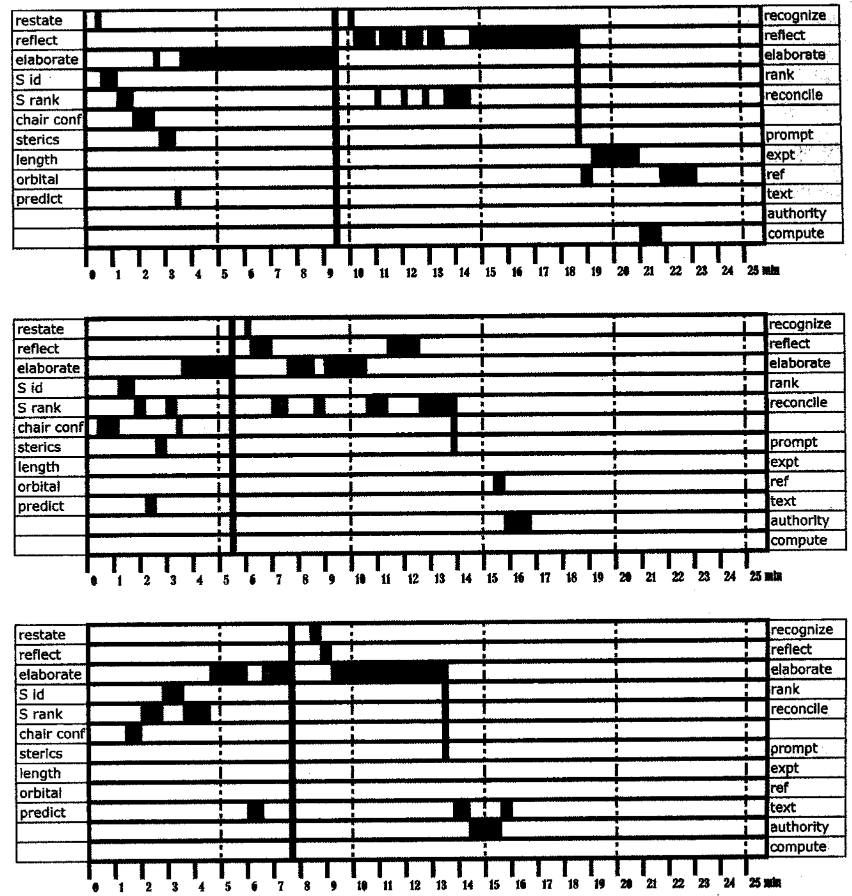


Figure 9. Representative coded interviews (top: 9T, a midcareer graduate student; middle: 9M, a student from the SSG program; bottom: 9B, a student who did not take the SSG option).

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. (1980). *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.