Investigating Classroom Myths through Research on Teaching and Learning

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

ACS SYMPOSIUM SERIES **1074**

Investigating Classroom Myths through Research on Teaching and Learning

Diane M. Bunce, Editor
Department of Chemistry
The Catholic University of America
Washington, DC

Sponsored by the
ACS Division of Chemical Education

American Chemical Society, Washington, DC

Distributed in print by Oxford University Press, Inc.

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.fw001

Library of Congress Cataloging-in-Publication Data

Library of Congress Cataloging-in-Publication Data

Investigating classroom myths through research on teaching and learning / Diane M. Bunce, editor ; sponsored by the ACS Division of Chemical Education.
p. cm. -- (ACS symposium series ; 1074)
Includes bibliographical references and index.
ISBN 978-0-8412-2659-3 (alk. paper)
1. Chemistry--Study and teaching--Research--Congresses. I. Bunce, Diane M. II. American Chemical Society. Division of Chemical Education.
QD40.I65 2011
540.71--dc23
2011035065

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48n1984.

Copyright © 2011 American Chemical Society

Distributed in print by Oxford University Press, Inc.

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of $40.25 plus $0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

Editor's Biography

Diane M. Bunce

Diane M. Bunce is a chemistry professor at The Catholic University of America in Washington, DC, where she has taught since 1985. She is the co-editor of *Nuts and Bolts of Chemical Education Research*, ACS Symposium Series 976, and the Associate Editor of the *Journal of Chemical Education*, Chemical Education Research Section. Diane served as an original author on all three of the American Chemical Society's curriculum projects: *ChemCom: Chemistry in the Community*, *Chemistry in Context*, and *Chemistry*. Her research focuses on how students learn chemistry and the mismatch between the way we teach and how the brain operates.

Chapter 1

Myths: What Are They and How Can We Challenge Them?

Diane M. Bunce*

Chemistry Department, The Catholic University of America,
Washington, DC 20064
***E-mail: Bunce@cua.edu**

Myths are cultural artifacts that result from our beliefs of how teaching and learning take place and interact with each other. Myths easily become "accepted truths" because they provide simple answers that are consistent with our beliefs about the teaching and learning processes. In order to challenge myths, we must be aware of the beliefs we hold and strive to provide targeted data to challenge these simple answers to multi-faceted questions. Myths can only be dispelled with convincing data and experiments that confront myths must therefore have clear research questions, suitable methodologies, and appropriate data collection and analysis. Such research must also have a firm theoretical framework so that the research results can be used to address our overarching beliefs about teaching and learning.

Introduction

Comments are often made by teachers, administrators and students alike regarding what constitutes good teaching, what problems exist in specific courses, and why learning is difficult in these courses. These comments are repeated so often that we as a community start to believe them, regardless of whether there are data available to support such beliefs. This can be the origin of unchallenged, but commonly accepted, myths. The myths seem to provide believable answers to our questions. These answers are satisfying because they are consistent with what we believe to be true about the processes of teaching and learning.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

So what are our beliefs about teaching and learning? What paradigm are they based on? To illustrate the connection between our beliefs and how these belief systems influence the myths we find satisfying, let's consider two examples. These examples deal with how learning and teaching interact and the teacher's role in the classroom.

If we believe that a teacher's role as expert is to provide knowledge to students, then the myth that may seem most satisfying to explain why some students don't learn in our courses is that the students are not trying hard enough. On the other hand, if we know that they are working hard, then a secondary explanation might be that they were not prepared well enough to handle the rigors of the course we are teaching. Our belief in how teaching should take place shapes our answers to the question of why students are not learning.

If, however, we believe that students must integrate the knowledge they hold in long term memory with the new knowledge to which they are being exposed, then we might not believe that lecture or a teacher's presentation of knowledge is the most effective way to teach. In this situation, the answer to the question of why some students are not learning might not be that students are not trying or are not prepared. Instead, a more satisfying answer may be that the instructional method (presentation of knowledge by an expert) doesn't match how learning takes place.

In both of these examples, the problem is the same—some students are not learning. The answer or myth we accept to explain this dilemma is based upon our belief that either the students are not trying or the instructional method is not adequate. Both myths are equally unsupported by data. They are myths because they are cultural artifacts arising from our belief system. They are not answers based upon research into the problem. One of the dangers of myths is that they provide easy answers that may or may not be accurate. If we accept these myths because they are consistent with our beliefs about teaching and learning, we may not see the need to investigate the reasons for the problem and thus miss an opportunity to supply data to address the question and help us formulate an objective and proven explanation of the problem.

Myths are dangerous not only because they provide easy answers to problems in teaching and learning but also because they can become lodged in our collective psyche as simplistic answers to complicated problems. If we are not skeptical of these easy answers, then our research may be designed incorrectly because it may be based on a false foundation. This could result in incorrect questions being asked and inaccurate conclusions being reached.

What If?

What if we decided to challenge some well established myths? We would need data. This data should come from research that was tightly designed and executed. Such research, like all research, should be skeptical of easy answers to complicated questions. The research methodology should account for important variables in teaching and learning, and control those that cannot be measured directly. The data analysis should be a deep evaluation of the variables themselves, and would need to go beyond reporting of percentages and calculations of simple t-tests.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Investigating such research questions might require unique methodologies that include more than the usual pre/post achievement measures and survey instruments or use these instruments in new ways. Tools used in such experiments would need to precisely measure the variables of interest and may involve novel approaches or techniques to accomplish this effectively. And lastly, an adequate amount of data should be collected to allow accurate conclusions to be drawn. Lack of sufficient data could result in a false interpretation thus creating a new myth.

Myth-challenging research may need to start small, investigating myths that directly impact our classroom experience. Subsequent research could move to more global questions related to how learning takes place in general. By examining what we currently believe, and challenging ourselves to ask questions that confront these beliefs, we could develop research questions that have the potential to substantially increase our understanding of teaching and learning. The end result could be data that challenge the myths, causing us to confront our beliefs of how people learn and the role of teaching in that process. There is also a place for educational and cognitive psychological theories in this research. Without theories to help explain and provide the framework for such research, the resulting investigations might result in a "million points of light" about teaching and learning that do not coalesce to help us understand the bigger picture. Theory-based research helps build on what was learned in previous research and extends our multi-faceted understanding of the teaching and learning interaction in a more documented fashion. A more in-depth discussion of the importance of using a theoretical framework to guide research is found in the Nuts and Bolts of Chemical Education Research symposium volume (*1*). Results from individual experiments that might not make sense on their own can be tied to the theoretical framework and start to fill in the puzzle of our understanding

Purpose of This Book

The purpose of this book is to provide examples of research that address a wide array of cultural beliefs about teaching and learning manifested as easy answers (myths) to common problems. The myths range from how long students pay attention in class to which chemistry topics are more difficult to learn than others. The methods employed in these investigations include familiar tools such as surveys and achievement measures as well as newer electronic means of collecting data such as the use of clickers. It is not always necessary to formulate complicated experiments to investigate myths. Sometimes it involves asking the right question at the right time of the right person. However not all myths are incorrect. The book highlights some of these investigative approaches, both conventional and unconventional.

How To Use This Book

This book can be used in several ways. First, the book can be used to provide data to challenge some commonly believed myths about teaching and learning. Some of the research presented proves, disproves or increases our understanding of commonly held myths. Second, the chapters of this book can be seen as examples of chemical education research collected in one place so that the breadth and depth of such investigations can be compared. Some investigations look at myths in detail while others focus more broadly on several interacting variables. Both types of research add to our understanding of different kinds of myths. Third, the book showcases a range of research questions and methodologies used to investigate them. This range emphasizes the importance of matching the methodologies used to the research questions asked. Lastly, the book offers examples of research that applies the chemical education research principles presented in the previous *Nuts and Bolts of Chemical Education Research* (*2*) symposium series volume, including the importance of constructing a good research question, choosing an appropriate theoretical framework and drawing conclusions based upon the data collected.

The book is organized into three sections that include the following:

Table 1. Book Organization

Section I	Chapters 2-5	Research dealing with myths or beliefs that teachers experience in the courses they teach.
Section II	Chapters 6-9	Research that investigates more global myths that deal with types of courses, longitudinal change, or larger institutional units.
Section III	Chapters 10-12	Research addressing the use of new statistical or research methodologies and tools to investigate familiar questions

The end result of using this book is to see how carefully crafted research questions and well designed studies can add to our understanding of how learning takes place and the role of teaching in that process.

References

1. Abraham, M. R. Importance of a Theoretical Framework for Research. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2008; pp 47−66.
2. Bunce, D. M.,Cole, R. S., Eds. *Nuts and Bolts of Chemical Education Research*; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2008.

Chapter 2

How Long Do Students Retain Knowledge after Taking a General Chemistry Test?

Diane M. Bunce*,[1] and Jessica R. VandenPlas[2]

[1]Chemistry Department, The Catholic University of America, Washington, DC 20064
[2]Department of Chemistry, Grand Valley State University, Allendale, MI 49401
***E-mail: Bunce@cua.edu**

Teachers often report that students forget chemistry soon after a test is completed. Is this true? Is it equally true for different general chemistry courses? This chapter describes how an experiment was designed to explore this question and the research done to address it. The methodological decisions, results, and interpretation of the results are presented here. The authors describe both the research done to address the question and the decisions made that shaped the research but did not necessarily appear in the final manuscript. The results include the fact that not all students experience a decay in knowledge following a test and when such a decay does occur, it happens within the first 48 hours and remains unchanged for up to 2 weeks. Some possible explanations of why some students experience a decay in knowledge might be due to the fact that they have had fewer opportunities to test their knowledge in between tests or did not have a comprehensive final exam scheduled for their course. In addition they may have been taught using a spiral curriculum but as novices, they may not have have understood the implicit spiral nature of their curricula. Lack of student motivation is another possible explanation.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Introduction

Both new and experienced teachers are often perplexed by what they see as students instantly forgetting knowledge that they previously answered correctly on recent tests. Experienced teachers can attribute this to students' poor test taking skills and use of surface knowledge learning techniques such as memorization as opposed to deep conceptual learning. Surface knowledge can be prone to knowledge decay within a short period of time because this knowledge is not well integrated in long term memory. Novice teachers may be more inclined to blame themselves for not being able to substantially affect student learning. It is quite possible that some of each of these is true—inadequate learning strategies on the part of the student and a mismatch between how we teach and the way students learn. But there are other variables that can affect the decay of knowledge. Among them are student motivation to truly learn the concepts; course structure that supports either deep or surface learning; student aptitude and the possible mismatch between student learning and the teaching level used in the course; or cognitive overload due to how the material is presented. In summary, student decay of knowledge may find its cause in either the student, the teaching or a combination of both. Before such causes can be fully investigated, it is important to document whether the belief that students quickly forget the knowledge they have previously demonstrated successfully on recent tests is truth or myth.

In some of our past unpublished research into learning, we have seen evidence of student decay of knowledge firsthand. In one study conducted in the early 1990's on the effect of achievement in high school chemistry, when an explicit problem solving approach was taught to a treatment group, the control group scored significantly lower on an unannounced quiz on the same material they had been tested on just 5 school days earlier. Students taught the explicit problem solving method did not demonstrate a significant drop in achievement on the unannounced quiz. In a current research project testing the effect of Process Orientated Guided Inquiry (POGIL), we interviewed students of experienced POGIL teachers at three schools around the country. During these one-on-one interviews, we showed students a question that had appeared on a chemistry test they had completed within one week prior to the interview. It was discouraging to see how many students were not able to answer the question during the interview when they had successfully answered the same question within the past week. Some students even admitted that they had understood the answer to the question at one point but couldn't recall during the interview either the answer or the logic they used to solve the question one week earlier. Even in our own teaching, it sometimes surprises us how often students don't seem to remember what they had just been tested on. Although this is unverified anecdotal information from small uncontrolled samples, the feeling persists that students "forget" the chemistry they once were successful in answering. It was this anecdotal information along with reports from the literature that led us to investigate whether students truly did "forget" or experience a decay in their chemistry knowledge within a short period of time following a test. We wanted to know if this is a misperception on the part of teachers (a myth in the teaching experience) or a real phenomenon that has its roots in either the cognition or sociology of student learning or the decisions that teachers make that shape their teaching.

What Does the Literature Say?

Decay of Knowledge

Wheeler et al. (*1*) studied the differential rates of forgetting for multiple study sessions versus multiple testing occasions. The results of this experiment show that although additional studying increases the amount of recall initially, after 48 hours, the amount of retention is equivalent whether a subject uses additional studying or testing situations to help retention. After one week, the amount of forgotten material was substantially less for those who experienced repeated tests. It is the time intervals used in Wheeler et al's study that are important for our study. Based upon the results of Wheeler et al's study, we chose a 48 hour window and the delayed window of 7 days to investigate both short and long term forgetting rates. Our goal was different from that of Wheeler et al. We were interested in knowing two things, namely, 1) Is there a decay in student knowledge? and 2) If so, at what time interval after a test does it appear? We were also interested in whether a documented decay of knowledge was stable over time. In other words, if it did exist, did the decay of knowledge remain the same over an extended period or did it change (worsen or improve) over a period of up to two weeks following a test?

Hockley (*2*) makes a distinction between the decay of knowledge that occurs with item information vs. associative information. Item information is defined as individual items and associate information is the connection between events. According to the psychologists these two types of information are encoded and stored differently and thus recalled differently as well. The results of Hockley's research show that the knowledge decay of associative information is much less than that of item information recognition. Karpicke and Roediger (*3*) further investigated the variables affecting forgetting by comparing the effects of multiple testing to multiple studying opportunities. The results show that of the two, multiple testing is the critical factor for promoting long-term recall and thus a smaller decay of knowledge. This result is explained in terms of the fact that the repeated retrieval from long term memory that is necessary during testing promotes retention more effectively than additional studying occasions that do not require retrieval from long term memory.

Many of the studies dealing with forgetting or decay of knowledge including Hockley's are conducted within a psychological laboratory situation and deal with students being taught and then being asked to recall a series of nonsense syllables, words, letters or numbers. Such psychological experimentation on the parameters of learning is necessary to learn what the brain is capable of. But laboratory testing of how learning (and forgetting) takes place is not enough. Classroom research in authentic learning environments is needed to apply these psychological findings within the multivariable reality of a classroom with real students and real teachers. This means that other variables that could affect the outcome might be unknown and/or uncontrolled. The challenge of authentic learning environment research is large but the potential benefits could offer some practical guidance for how teaching and learning can optimally be combined. Arzi et al. (*4*) attempted to address this problem in their long term retention study of science learning in a school setting. Their study looked at the amount of forgetting (decay of knowledge) that students experienced over a two year period. The researchers document a significant loss of knowledge over this period but show a difference in the forgetting patterns of average/good students vs. poor students. Average or good students demonstrated an irreversible loss of a smaller part of their knowledge than did the poorer students. This seems reasonable. If students understand what they are learning, then they are more likely to be able to build connections to that knowledge in memory and thus more easily access it when needed. This is probably one of the differences between average/good students who have a deeper understanding of the new knowledge that they are exposed to than poorer students. Thus average/good students are better able to retain the new knowledge that is integrated with that already in their long term memory.

Much of the research we have about the decay of knowledge in the learning process comes from the psychology literature. But many teachers have anecdotal evidence to support the idea that students don't seem to remember information, specifically chemical information that they previously demonstrated competence in just days prior. Little experimental research has been done in authentic learning environments to document how long after a testing occasion this decay of knowledge occurs. Many authors who write books on good teaching practice attempt to bridge the gap between the psychological literature and authentic classroom environments about the decay of knowledge. They offer advice to teachers on possible causes for the decay and ways to decrease it through good teaching practices. Sousa (*5*) suggests that teachers take advantage of what we already know about the decay of knowledge, i.e., knowledge that is not rehearsed or retrieved will be forgotten and he suggests to teachers that they purposely revisit important concepts throughout a student's school experience to provide this rehearsal/retrieval. Bligh (*6*) suggests that people forget things that are not important to them and offers teachers ways to help make information more meaningful to students. Bligh also suggests that motivation is important because highly motivated people work carefully and recognize the importance of new material. Alternatively, he suggests that students who are motivated spend more time on task. This leads us to a discussion of what variables might affect the decay of knowledge in authentic learning environments.

There are several reasons why people experience a decay of knowledge after they have initially learned something. Decay of knowledge does not deal with the variables involved in the initial learning but rather the reasons why students experience a decay of that knowledge after they have demonstrated achievement of it on tests. These decay of knowledge variables include student aptitude, course structure, curricula approach, teaching style, and student motivation among others. The purpose of the research reported here was not to investigate *why* knowledge decay occurred but rather *if* it occurred. An investigation of the reasons for any observed decay should be the focus of future studies. However, a discussion of the scope of the variables that might affect the decay of knowledge is within the scope of this chapter and will be included.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Variables That Might Affect the Decay of Knowledge

Student Aptitude

Arzi et al. (*4*) report that average/good students retain more of their initially learned knowledge than their less able counterparts. The categorization of student s as good, average, or poor was based upon student chemistry achievement on a test on the Periodic Table used in the study. In planning a more general study of students' decay of knowledge as was done in our study, a measure of aptitude, rather than previous achievement was sought. Although SAT scores are used in many studies as a measure of aptitude, it is sometimes difficult to obtain permission to access such scores. In addition, many schools now accept either ACT or SAT scores and a reliable conversion formula between the two must be applied to interconvert the scores on these two tests. Since our study included both university and high school students, there was the added problem that not all the high school students in the study had yet taken the SAT or ACT. So an alternative measure of aptitude was needed. The Group Assessment of Logical Thinking (GALT) (*7*) has been used in other studies when a quick measurement of aptitude is needed. This test administered both electronically or as a paper and pencil test, usually takes 12-20 minutes to administer and consists of 12 questions, ten of which require both a correct answer and a correct reason for that answer. The test has been modified (Bunce and VandenPlas) to replace one question that might be considered offensive by people of different sexual orientations. The GALT test is a paper and pencil Piagetian test that measures students' ability to think logically. The questions in GALT are independent of any specific content area such as chemistry. Research that we have done (*8*) shows that the GALT test measures the same variable of student learning but just not as thoroughly as the Math SAT. Thus, this relatively easy to administer pretest can be used to measure student aptitude within a research study without putting undue strain on either the students or the research timeline. Once measured, the GALT score can be used as a way of categorizing students according to the variable of logical thinking. Logical thinking is a variable that is relevant to the successful study of chemistry.

We chose to include the GALT score in our analysis so that we could investigate whether the decay of knowledge was dependent on student aptitude as measured by this test of logical thinking. We were interested in knowing whether any observed decay of learning was more evident for low, medium, or high GALT students. The categorization of low, medium, and high GALT students was based upon the distribution of GALT scores for students in this study.

Course Structure

Course structure can either positively or negatively impact a student's engagement in a course. The structure is usually presented in the course syllabus and includes features of the course such as tests, quizzes, homework, class participation and other course aspects that will affect a student's grade. Course structure also includes a description of the type of feedback a student can expect in the course. It includes both features of the course and the schedule with which they will be implemented. For instance, how many full period tests will a course have? Is there is only a midterm and a final? In the case of infrequent testing, students may not fully engage with the course content until just before the scheduled tests. With 4 tests and a final exam in a 15-week semester, students must engage at least 5 times during the semester (about once every 3-4 weeks) to prepare for the more frequent tests as opposed to twice a semester in the case of a midterm and final only course structure. Course structure can also include the presence and number of intervening quizzes between tests. This feature can also affect how often students engage with the course material. Graded homework is another opportunity to engage students and provide feedback. If homework, quizzes, and tests are graded and returned in a timely fashion, students can use this feedback to refine or expand their understanding of chemistry concepts. McDaniel et al. (*9*) have shown the effectiveness of quizzing in promoting learning and subsequent retention of material on tests. Short answer quizzing which requires recall was more robust in their study than the use of multiple choice questions on tests which require only recognition. Timely feedback on the correctness of quiz answers was considered to be key in the positive effect of testing on retention in the McDaniel et al. study. The McDaniel study is an example of how quizzing interspersed with testing occasions helps promote the retention of knowledge.

The use of personal response devices (clickers) in class is also part of a course structure that can be used to engage students on a much shorter interval with the material being taught. Clicker questions are essentially ConcepTest questions: short, multiple-choice questions focusing on a single topic which are conceptual, rather than algorithmic, in nature (*10*). Clicker questions are inserted into a lecture or class period and their answers are recorded electronically on the teacher's computer. Graphs can be displayed immediately following student responses to the question. These graphs illustrate the range of understanding on a question dealing with a specific topic by showing the percentage of students who selected each answer to the question. Feedback in the form of a correct answer indicator provides students with immediate, anonymous feedback on their understanding of a particular question. Clicker questions are often embedded within PowerPoint presentations and can then be uploaded to course websites along with the correct answers so that students can review both question and answer at a later date. This course structure feature provides another means by which students can be engaged with the material to be learned. It is expected that such engagement facilitated through the structure of the course could have a positive influence on the retention of knowledge.

Curricula Approach

Hockley's (*2*) work on the differential decay of item information and associative information might be seen as the underlying principle behind the move in education from traditional or linear to spiral curricula. Traditional or linear curricula present each chemistry concept in its entirety and then move to the next concept. Spiral curricula, on the other hand, present a concept with as much detail as is needed to understand the question at hand. The curricula then return to the concept again and again adding more detail each time in association with other related concepts. The spiral curricula is sometimes explained as presenting knowledge on a need-to-know basis so that the learner has the knowledge that seems the most relevant at the moment. Such an approach can foster the connections between concepts, a hallmark of associative information, and thus has the potential to diminish knowledge decay. Spiral curricula although truly integrated in the minds of the curricula authors and teachers, may not seem obvious to the students. For instance in a highly regarded curricula *Chemistry in Context* (*11*) that presents information on a need-to-know basis, chemistry concepts are taught within the global issues of clean air, water and energy. Students who experience a complete story of global warming may not realize that the chemical concepts presented in this section are the same chemistry topics revisited in the discussion of the ozone layer or clean water. In order to be effective, the revisiting of concepts within spiral curricula must be obvious to students either through the materials themselves or how they are used in the classroom.

Teaching Style

A teacher's teaching approach, whether it is student-centered or content-centered, can have an effect on how engaged a student is in the process of learning. A teaching style that is student-centered will emphasize student engagement in learning including such techniques as cooperative groups, clicker questions and resources that provide immediate feedback for students. The course structure in such situations would make it possible for the student to demonstrate mastery in several different formats. Testing would be both formative and summative with demonstration of the student's mastery of the concepts as the ultimate goal. By contrast, a content-centered approach to teaching would emphasis "coverage" of chemical concepts by the teacher through lecture or other student-passive means. Student mastery would be tested primarily in a summative method meaning that you either know it or you don't. Decay of knowledge in courses where content coverage vs. student mastery of concepts is the goal, is expected to be larger.

Student Motivation

Motivation is defined as the internal state that captures, directs and sustains student behavior in achieving goals (*12*). In the academic process, motivation to learn is more specifically defined as a student's tendency to find academic activities worthwhile. It is hypothesized that the more motivated a student is, the more engaged he/she will be in the activities that encourage learning. The deeper that learning, the less likely or slower student decay of knowledge will be. Motivation is a multi-variable construct that many have tried to deconstruct so that measurements of some of the components of motivation can be made. In the literature (*13*, *14*) five key variables of motivation have been identified including self efficacy (*15*). Self efficacy is defined as belief in one's capabilities to organize and execute procedures that will produce the desired results. Self efficacy is often measured through questionnaires that ask people what they believe about their chance to succeed in a specific goal attainment. Science self efficacy instruments typically ask students to rate their confidence in their ability to succeed in science courses. Zuscho et al. (*16*) have shown that student scores on self efficacy instruments can predict grades in college chemistry courses. By definition we would expect students who freely chose to study chemistry to be more motivated, more willing to persevere when they encounter a difficult chemistry concept and more self satisfied with their level of understanding of chemical concepts. Students who are required to take chemistry as a pre-requisite for their major or for their general distribution requirement would be expected not to be as self-motivated. It is believed that the degree of self-motivation would be a factor in the decay of knowledge experienced by students in chemistry.

Purpose of This Research Study

The purpose of this research study was to investigate whether students experienced a decay of knowledge following a test in chemistry. Rather than choose one general chemistry course, we looked at the decay of knowledge in three separate courses with three different student populations namely undergraduate nursing students, undergraduate nonscience majors and high school honors students.

In addition to investigating *if* a decay of knowledge occurred, we were also interested in knowing *when* it occurred and if it did occur, how long it lasted within a two week period following a test. To further investigate this issue, we examined whether the decay was the same for students of differing logical reasoning ability.

Our research questions for this study were as follows:

1. Do chemistry students in a variety of general chemistry courses experience a decay in their knowledge following a regularly scheduled test?
2. How is this decay affected by the variables of type of course, length of time following the test, and student logical reasoning level (GALT)?

Sample

The sample chosen for this research included students in three separate general chemistry courses. This decision was made so that a range of students with different motivations to study chemistry could be examined. Both undergraduate general chemistry courses were taught by the same instructor and experienced similar teaching styles. The textbooks, curricular approach and course structure for these two courses differed and were geared to the goals of the course. The high school general chemistry course was taught by a different teacher and used a different textbook, curricular approach and course structure that was geared to the goals of the high school course.

The undergraduate courses included two semesters of a General, Organic and Biochemistry course for nursing students that were taught simultaneously in the Fall 2007 semester. Both courses (I and II) of the General, Organic, and Biochemistry course for nursing students used the same textbook, curricular approach, course structure, and teacher. Both courses also included students who were classified as first semester freshman. Students were assigned to one of these courses based upon their scores on a chemistry placement examination given the week before classes began. When tested for logical reasoning ability (GALT), there was no significant difference in the logical reasoning ability of these two courses (nursing I and II). Based upon all this information, it was decided to combine the students in undergraduate nursing I and II into a single nursing course variable. The one semester course for nonscience majors which used the *Chemistry in Context* (ACS) curriculum was also taught during the Fall 2007 semester.

The high school honors course included in this study was the second semester of a two semester course taught in Spring 08.

Methodology

Overview

The methodology used in this experiment was to repeat two free response chemistry questions that appeared on regularly scheduled tests at three given time intervals following the test and compare student achievement on these questions across these multiple testing occasions. In order to implement this methodology, some research decisions had to be made. Some of these decisions are described here.

Selection of Questions from Test

The decision was made to select open-ended questions rather than short answer or multiple choice questions because open-ended questions have the potential to offer richer insights into students' logic used to answer questions. Small changes in that logic could more easily be examined and reflected in the achievement grade with open-ended questions. The questions selected by the researchers were questions written by the individual course instructors which had been used previously on their exams. The researchers did not impose the use of specific questions on the classroom teachers but instead used questions from the teachers' tests that matched the research criteria. These criteria included the use of open-ended chemistry questions which the teachers believed their students should be able to answer correctly if they truly understood the chemistry. Within each course, the same grader who graded the tests was used to grade the quizzes using a rubric that had been developed by the researchers. Sample questions from each course are given in Table I.

Table I. Course and corresponding conceptual sample questions (*17*)

Course	*Question*
Undergraduate nursing I	The density of water at 37°C is 0.993 g/mL (the density changes as temperature increases). The density of normal urine ranges from 1.003 to 1.030 g/mL at the same temperature. Explain why the density of urine is greater than the density of water and how this can be used to diagnose some illnesses.
Undergraduate nonscience majors	Blue colored solutions transmit blue light and absorb red light. Are blue solutions absorbing light at a higher or lower energy than they are transmitting? Explain your answer.
High School honors chemistry	Even though the oxygen demands of trout and bass are different, they can exist in the same body of water. However, if the temperature of the water in the summer gets above 23°C, the trout begin to die, but not the bass. Why is this the case?

Time Intervals for Quizzes

Ideal times for the scheduling of quizzes were suggested by the researchers for 2-3, 5-7, and 13-14 days. The intervals had to be expanded to accommodate intervening weekends, scheduling of classes and other interruptions in the schools' schedules. In addition, the decision was made not to have the same students take the quiz at each of the three time intervals following a given test. This was done to avoid both undue familiarity with the questions and mental fatigue in answering the same questions several times within a short time frame. In order to accommodate the time necessary to remind students to take the quiz and the time necessary for the quiz to be taken, the time intervals had to be extended beyond those originally suggested. The final combination of time intervals used by all students across the courses was 2-5, 6-9, and 10-17 days.

Students Assigned To Take the Quizzes at Specific Time Intervals

If all students answered the same two questions both on the tests and at all three of the delayed time intervals, both question familiarity and mental fatigue might become intervening variables. To avoid this complication, students were assigned to one of three stratified random cohorts. Each cohort took only one delayed quiz following each test. Since each undergraduate course used three tests in this study, each cohort cycled through the three possible time intervals over the course of the experiment. For example, cohort 1 took a quiz in the 2-5 day interval for test #1, the 6-9 day interval for test #2 and the 10-17 day interval for test #3. The high school honors students cycled similarly but their course offered only two tests during the experiment.

Student Cohorts

Each cohort was determined through a stratified random sampling technique. The stratification was based upon GALT scores. Each cohort contained a random sample of low, medium, and high GALT students.

Students were not aware that they had been assigned to a specific cohort to complete a quiz. In the undergraduate courses (nursing and nonscience majors) email was used to notify individual students within a given cohort that it was time to take the quiz online. Quizzes were available online for the designated cohort during the specified time frame. Students were directed to complete the quizzes without the use of textbook, notes or other people. The time spent online to complete a quiz was noted by the researchers to detect any obvious violation of the directions to take the test on one's own. Students received credit for completing the quiz regardless of the score on the quiz. No test questions were discussed in class or with individual students nor were any test answer keys made available to students until all quizzes were completed.

In the high school honors course, a paper copy of the quiz was completed in class or lab by each cohort at the appropriate time.

Pretests

All students were required to complete the Group Assessment of Logical Thinking (GALT) test before the study began. GALT scores of the students within each course were then tabulated as seen in Table II.

Table II. Mean GALT Scores by Course

Course	*N*	*Mean*	*Standard Deviation*
Undergraduate Nursing, Part I	44	6.89	1.99
Undergraduate Nursing, Part II	29	6.66	2.54
Undergraduate Nonscience Majors	44	7.45	2.52
High School Honors	38	10.18	1.41

Results

In order to address the research questions, a statistical test was required to investigate student achievement on questions from test to quiz, but also how that change was affected by variables such as length of quiz delay, student logical reasoning level, and course type. Analyzing one student's test question score against his/her own subsequent quiz score is considered a within-subjects repeated measures comparison—comparing one individual to him/herself. For this study, then, achievement score on test and quiz questions are dependent variables and the comparison of test vs. quiz achievement is a within-subjects independent variable. In addition to this comparison, we also wanted to compare achievement scores among students of differing logical reasoning ability and with different lengths of quiz delay. Comparisons among students are considered to be between-subjects independent variables. In this case logical reasoning level and length of quiz delay are between-subjects variables (see below). Because two different kinds of independent variables were used, the statistical test selected for this study was a mixed between-within subjects analysis of variance (ANOVA).

A separate mixed between-within ANOVA was used for each of the three courses (undergraduate nursing, undergraduate nonscience major, and high school honors chemistry) because the textbook, curricula, curriculum framework, teaching practices and questions on tests and quizzes used in each course differed substantially, especially between the undergraduate and high school honors courses. Each ANOVA was first able to test for a significant overall decay of student knowledge across all tests and time delayed quizzes, regardless of length of decay or student GALT score within a specific course. In this analysis we did not investigate the effect that specific content had on the decay of knowledge. Following the test for global significance, we tested for more specific results by course, length of quiz delay, and logical reasoning ability (GALT score). To do this analysis for each course, we combined the data from all test questions into a single variable called "test question". The questions used on the time-delayed quizzes were also combined for each course into a single variable called "quiz question". Since students took each quiz at one assigned time interval, the effect of quiz delay intervals was measured by comparing the students in a course at time interval 1 with those in the course at time intervals 2 and 3. Assigning students to only one delayed quiz per test was done to control for the confounding effect of test familiarization and mental fatigue that might have resulted if each student took the quiz three times (once at each time interval) following a test. This methodological decision meant that the variable of time (for the time delayed quizzes) was a *between* subjects variable, rather than a *within* subjects variable. Students were also rotated through the different time intervals for the quizzes so that a given cohort of students did not experience the same time interval for more than one test.

Before an ANOVA can be used in an analysis, several assumptions regarding the data must be tested. First, the ANOVA requires that the dependent variable(s) be measured using a continuous scale, as opposed to using categorical data (*18*). In this study, questions were graded by a rubric as discussed above, and converted to a percentage (continuous) score.

ANOVA also requires independence of scores both within treatment groups and between groups (*19*). In this study we examine independce of scores by asking questions such as: does one student's achievement score affect the score of another student or would we expect scores between two students to be correlated in some way. If a researcher suspects a violation of this assumption of independence of scores then a more stringent alpha value can be used for the significance tests (*18*) to compensate for the dependence of scores. In this study based upon the experiment itself, student scores were assumed to be independent and a standard significance level of $p<0.05$ was used.

Homogeneity of variance is another assumption that must be met in order to use an ANOVA. We assume we are taking samples from populations that have equal variances (*18*). The most common test for this assumption is the Levene test for equality of variance. A nonsignificant result for the Levene test ($p>0.05$) tells you that this assumption was not violated. In this study, we must check this assumption for each of the three ANOVAs used for the three courses. The Levene test results are nonsignificant ($p>0.05$), as required for the ANOVA, for both the nursing and high school courses. In the nonscience majors' course, however, the Levene test shows that the data violate the assumption of homogeneity of variance ($p=0.000$ for both test and quiz question scores). The literature suggests that the ANOVA statistic is relatively robust in relation to violations of this assumption and that we can address such violations by setting a more stringent level of significance (*18*). Setting a more stringent level of significance serves as an added precaution against making a Type 2 error (*18*). For the results of the ANOVA with the nonscience majors'course, a more stringent level of significance ($p<0.01$) was adopted as suggested by the literature (*18*) as opposed to the more common significance level of $p<0.05$.

Because we are using a mixed between within ANOVA to analyze data, there is one additional assumption that must be met, i.e., the assumption of homogeneity of intercorrelations. To test this assumption, we must show that the intercorrelations among the levels of the within subjects' variable (in this case , test vs. quiz achievement scores), are equal (*18*). This assumption is tested with the Box's M statistic. The Box M statistic is sensitive, so it is generally adequate to accept a probability level greater than 0.001. In this case, all courses show a Box's M probability greater than 0.001 which means that the data has not violated this assumption.

To summarize, as a result of the experimental design used in this experiment, the three mixed between-within subjects ANOVAs (one for each course) had two *between-subject* factors (quiz time interval and GALT group) and one *within-subject* factor (the repeated measures of test vs. quiz achievement). This analysis enabled an overall within-subjects comparison of test question achievement vs. quiz question achievement, as well as between-subjects comparisons of the remaining variables and interaction effects among all variables for each individual course.

Main Effect for Test vs. Quiz Question Achievement

The results of the three mixed ANOVAs provide information about the main effects of each independent variable, as well as the interaction effects among variables. A main effect shows how one single independent variable affects the dependent variable(s). As previously discussed for this study, the dependent variables were student scores on both quiz and test questions. The independent variable we were most interested in was the within-subjects repeated measures comparison of test vs. quiz achievement for each course. The overall main effect for test vs. quiz achievement tests whether student scores changed from the test to the quiz question, regardless of logical reasoning ability or quiz time interval.

The overall main effect for test vs. quiz achievement was not significant for either the nursing general chemistry course ($F_{1,407}= 2.01,\ p= 0.16$) or the high school honors course ($F_{1,113} = 0.01,\ p=0.92$). This means that students in these courses showed no significant difference between their test and delayed quiz achievement. For the nonscience majors, on the other hand, the main effect for test vs. quiz achievement *was* significant. The significance level of this difference ($F_{1,221}= 26.77,\ p=0.000$) is below the conservative level of 0.01 which was chosen for this study. This significant difference exhibits a medium level of statistical power (partial eta squared = 0.108) (*18*), and shows that students experience a decrease in achievement from test (M=78.2, SD=31.4) to quiz (M=65.0, SD=31.0). Achievement means for test and quiz questions for each course are given in Table III.

Table III. Test and Quiz Question Means by Course (*17*)

Course	*Test Mean (Std. Dev.)*	*Quiz Mean (Std. Dev.)*
Undergraduate nursing	66.73 (32.78)	68.80 (31.22)
Undergraduate nonscience	78.15 (31.42)	64.99 (30.96)
High School honors	80.42 (27.43)	78.10 (32.16)

The data in Table II show that the undergraduate nonscience majors' course is the only course to show a decrease in achievement scores between test and subsequent quiz questions.

Quiz Time Interval and Achievement Interaction Effect

Interaction effects, as opposed to main effects, show us the relationship between multiple independent variables, and how they can act together to influence the dependent variable. In this study, we can examine the interaction between any combination of our independent variables. One such interaction effect between quiz time interval and test vs. quiz achievement, for example, will show us if the comparison of student test question and quiz question scores is affected by the length of time between the test and quiz.

Since the nonscience majors' course was the only course that showed a significant difference between test and quiz achievement, this was the only course where the interaction effect for quiz time interval and achievement could be investigated. This interaction effect was *not* significant ($F_{2,\ 221} = 1.92,\ p = 0.15$) for this course. This means that nonscience majors performed similarly on test and quiz questions, regardless of when the quiz was taken (2-5 days, 6-9 days, or 10-17 days after the test). Since the ANOVA did show an overall significant difference between the test and delayed quiz achievement for this course, the drop in achievement must have taken place after the test but prior to the first measured time delayed quiz (2-5 days). This is interpreted as the decay of knowledge occurring within the first 2 days (48 hours) following the test.

Main and Interaction Effects with GALT Score

The last interaction effect is that of logical reasoning ability (as measured by the GALT test) on student achievement. The mixed ANOVA provided not only an analysis of the main effect of GALT score for each course, but also how the GALT score interacted with the independent variables of test vs. quiz achievement and quiz time interval. The main effect for GALT was shown to be significant in two of the three course ANOVAs. Both the nonscience students ($F_{2,221}$=5.76, p=0.004) and nursing students ($F_{2,407}$=4.89, p=0.008) showed significant differences in test and quiz question achievement when analyzed by GALT level. This is not surprising, as GALT score has been shown to correlate with success in chemistry (*8*). The high school honors students, however, showed no significant main effect for GALT score ($F_{2,113}$=0.102, p=0.750) on the dependent variables of test and quiz question achievement. This was probably due to the homogeneity of the academic ability of the high school students who were selected for the competitive honors course. The high school honors students had the highest average GALT score and lowest standard deviation of the three courses used in this study (Table I).

The interaction effects between GALT score and the other independent variables of test vs. quiz achievement and quiz time interval in all combinations were shown to be nonsignificant. This is interpreted as student logical reasoning ability having no effect on any of the previously discussed main or interaction effects. For example, the nonscience majors' significant difference in test and quiz achievement was true for *all* students, regardless of logical reasoning ability. Low, medium, and high GALT students showed the same drop in quiz scores as compared to test scores, and the same nonsignificant interaction effect with quiz time delay in this course. While we may have originally predicted no decay for high GALT students over time, or a smaller decay over time when compared to low GALT students, we instead see that the GALT score does affect the achievement score, but does not affect the amount or time of decay. For courses where no decay was seen (nursing and high school), there is an absence of decay in *all* students regardless of logical reasoning ability. High GALT students had high achievement scores on the test and kept these high scores on the quiz. Low GALT students had low achievement scores on the test, and kept these low scores on the quiz.

Discussion

According to the data presented, only one of the three groups (nonscience majors) experienced a significant decay of knowledge following a test. This decay occurred before the first time interval of 2-5 days and did not significantly change during approximately two weeks following the test. The decay of knowledge within 48 hours is consistent with the time interval for knowledge decay reported by Wheeler et al. (*1*). Since the purpose of this experiment was to investigate if a decay of knowledge occurred, the answer is yes in some general chemistry courses a decay of knowledge is detectable within 48 hours after a test. This decay is stable over the course of this experiment (up to 17 days after a test). There is no significant effect of students' logical reasoning level on this decay. High logical reasoning students did not have a different decay of knowledge that low logical reasoning students. The only significant interaction with the decay of knowledge is due to the specific course in which it is documented.

So the natural question to ask is why do the nonscience majors experience decay in knowledge when the nursing students and high school honors students do not? Although the reasons for the decay were not specifically studied in this investigation, there is some information that was collected during the study that may partially inform the discussion.

According to the literature, additional testing or quizzing opportunities have a more significant effect on student achievement than additional studying opportunities. In this study, the course structure of both the nursing chemistry and high school honors courses included intervening quizzes between tests as part of their course structure. The nonscience majors' course did not. Two of the three courses (high school honors and nonscience majors) required the submission of homework on a regular basis. The homework was graded and returned. However, the presence of graded homework assignments does not appear to have a significant effect on the decay of knowledge. In this experiment, one course that required graded homework did not experience a decay of knowledge (high school honors) and one did (nonscience majors).

Another aspect of course structure that differed among the three courses is the presence of a cumulative final exam. The two courses (nursing and high school honors) that did have a cumulative final exam did not experience decay in knowledge. The one course that did not have a final exam (nonscience majors) did have a decay in knowledge. It is possible that students knowing that they will eventually be tested on all the material covered in the course might be at least partially responsible for the absence of a decay of knowledge immediately following the test in some (nursing and high school honors) chemistry courses.

The course curricula in all three courses claimed to be spiral curricula where concepts are revisited in ever increasing detail as the course proceeds. In two of these courses (nursing and high school honors), the topics in each chapter were listed in the textbook and students could easily identify which topics they had already seen prior to studying the new material. In the nonscience majors' curriculum, the spiral curriculum may not have been as obvious to students. The model for the *Chemistry in Context* (*11*) curriculum is to present chemistry concepts within real world problems such as global warming, ozone hole, clean water, plastics or food. It is possible that students do not separate the chemistry from the issue and therefore may not realize that polymerization is important to both plastics and proteins or that different energetic forms of radiation have different effects on bonds in both global warming and the destruction of ozone.

Although we did not purposely document the teaching style of the two instructors involved in this study, the one teacher who taught both undergraduate courses had one course (nursing) that did not experience a decay of knowledge and one that did (nonscience majors).

Student motivation is another possible reason for the difference in the decay of learning among the three courses. Nursing students may be motivated to succeed in chemistry because they regard it as a gateway course to being accepted into the clinical nursing program at the end of their freshman year. High school honors students may be concerned about their grade in all courses and the effect it will have on their overall GPA for college admission. It might be logical to think that nonscience majors, who do experience a decay in knowledge, are the least motivated to learn chemistry. Many of them are taking the course to fulfill a distribution requirement for graduation. To explore this last point that the nonscience majors in this study were not highly motivated to succeed in chemistry, we analyzed their anonymous in-house evaluations given on the last day of class.

The evaluation for nonscience majors was designed by their teacher and had been used in this course for the past ten years. Many of the questions asked for the students' rating of the effectiveness of different aspects of the course such as the use of clickers in class, TA office hours, group worksheets, labs, and the posting of annotated notes on BlackBoard following each lecture. In addition to these questions were a series of Likert scale questions that asked about the students' confidence in their ability to learn and understand chemistry both before the course started and again at the completion of the course. There were also questions asking students to evaluate their enjoyment of the course and if they would consider taking a second chemistry course if they had room in their schedules. A summary of the questions selected from the course evaluation that measure such motivation sub factors as self efficacy are included in Table IV.

Based upon this post hoc analysis, it seems unfounded to assume that the decay of knowledge in the nonscience majors' course is due solely to a decreased motivation to learn chemistry.

Table IV. Student Responses to Course Evaluation (*17*)

Question on Evaluation	*Ave. Score on Likert Scale of 1→ 5 1=Agree Strongly 5= Disagree Strongly*
1. Before this course started, I was confident that I could learn and understand chemistry	2.7
2.Now that the course is over, I feel confident that I can learn chemistry	1.8
(Average change in confidence to learn chemistry from pre to post course)	-0.9 (towards agree strongly)
6. I enjoyed this class	1.6
13. I would consider taking a second chemistry course if I had room for another science/math elective	2.4
14. This course was better than I expected it would be	1.2
15. This course was worse than I expected	4.2
Free response On a scale of 1 to 5 (1= one of the three best courses you have taken and 5= one of the worst courses), how would you rate this course?	2.0

Conclusions

There are two levels of conclusions to be made in this chapter. The first deals with the results of the investigation and the second with the decisions made during research that are often not obvious to the reader of the published research manuscript.

The results of this study indicate that there is a decay of knowledge in some general chemistry courses but not all. When this decay occurs, it takes place within the first 48 hours following a test and remains stable over a time period of at least 2 weeks. There is no differential decay of knowledge detected in any of the three general chemistry courses used in this study (undergraduate nursing, undergraduate nonscience major, and high school honors) for students of different levels of logical thinking. The reasons for why this decay of knowledge occurs in one general chemistry course and not others was not the focus of this investigation, however variables such as student aptitude, course structure, curricula approach, teaching style, and student motivation may play a role. Although not studied directly, there is some evidence that course structure such as the use of intervening quizzing between tests and the presence of a cumulative exam plus a spiral curricula approach that is obvious to students may play a role in preventing a significant decay of knowledge in some courses. The influence of these variables should be considered tenuous at this point and deserving of future, controlled experimental study.

The conclusions of this research in terms of the decisions that are made in planning and executing a chemical education research project include how to design an experiment that does not allow for competing hypotheses. One of the biggest decisions we made in planning this research was to limit the scope of the experiment to whether or not a decay of knowledge occurs following tests in general chemistry courses. We believed that this question was large enough for a single research project and that subsequent investigations could look into why such a decay occurs.

In this experiment we also chose to investigate more than one general chemistry course for the presence of the decay of knowledge. By making this choice, we have a richer data set on which to base conclusions. We chose to directly measure student aptitude using the same pretest (GALT) as opposed to using SAT scores from different administrations of the SAT exam and possible conversions between ACT and SAT scores for some students. We controlled for student fatigue in answering the same test and quiz questions multiple times within a relatively short period of time by creating stratified random cohorts of students who cycled through the quiz time intervals following tests. This cycling of student cohorts during different quiz time intervals was only possible because we collected data over an extended period of time (3 of 4 tests for the undergraduate students and 2 of 3 tests for the high school honors students).We did attempt to address intervening variables that might unduly diminish our findings by analyzing additional data that at least partially addressed the question of whether the nonscience majors were not motivated to learn chemistry and therefore a decay of knowledge for this group and not the others would be predictable. Some of the research decisions were obvious to us at the start of the study while others were made during the data collection or analysis phases. Still other decisions were a result of questions raised by the reviewers when we first submitted this manuscript for publication. The conclusions regarding chemical education research that are illustrated in this chapter are that a single investigation cannot be expected to definitely answer all aspects of the research question(s) asked. Most research studies lead to more studies to investigate additional questions raised in the course of the research. The second conclusion is that research is an iterative process that builds on what comes before it and leads to new investigations that together will provide answers to the questions we ask.

References

1. Wheeler, M. A.; Ewers, M.; Buonanno, J. F. *Memory* **2003**, *11* (6), 571–580.
2. Hockley, W. E. *J. Exp. Psychol.* **1992**, *18* (6), 1321–1330.
3. Karpicke, J. D.; Roediger, H. L. *Science* **2008**, *319*, 966–968.
4. Arzi, H. J.; Ben-Zvi, R.; Ganiel, U. *Sci. Educ.* **1986**, *70* (2), 171–188.
5. Sousa, D. A. *How the Brain Learns*; Corwin Press: Thousand Oaks, CA, 2006.
6. Bligh, D. A. *What's the Use of Lectures?* Jossey Bass: San Francisco, CA, 2000.

7. Roadrangka, V.; Yeany, R. H.; Padilla, M. J. *GALT: Group Assessment of Logical Thinking*; University of Georgia: Athens, GA, 1982.
8. Bunce, D. M.; Hutchinson, K. D. *J. Chem. Educ.* **1993**, *70* (3), 183–187.
9. McDaniel, M. A.; Anderson, J. L.; Derbish, M. H.; Morrisette, N. *Eur. J. Cog. Psychol.* **2007**, *19* (4/5), 494–513.
10. Mazur, E. *Peer Instruction: A User'S Manual*; Prentic-Hall: Upper Saddle River, NJ, 1997.
11. *Chemistry in Context: Applying Chemistry to Society*, 6 ed.; American Chemical Society/ McGraw-Hill Higher Education: Dubuque, IA, 2009.
12. Glynn, S. M.; Taasoobshirazi, G.; Brickman, P. *J. Res. Sci. Teach.* **2007**, *44* (8), 1088–1107.
13. Bandura, A. *Annu. Rev. Psychol.* **2001**, *52*, 1–26.
14. Schunk, D. H. In *Self-Regulated Learning Academic Achievement: Theoretical Perspectives*, 2nd ed.; Zimmerman, B. J., Schunk, D. H., Eds.; Erlbaum: Mahwah, NJ, 2001; pp 125−151.
15. Bandura, A. *Self-Efficacy: The Exercise of Control*; Freeman: New York, 1997.
16. Zuscho, A.; Pintrich, P. R.; Coppola, B. *Int. J. Sci. Educ.* **2003**, *25* (9), 1081–1094.
17. Bunce, D. M.; VandenPlas, J. R.; Soulis, C. *J. Chem. Educ.* **2011**, *88*, 1231−1237.
18. Pallant, J. *SPSS Survival Manual*; Open University Press: Philadelphia, PA, 2001.
19. Keppel, G. *Design and Analysis: A Researcher's Handbook*; Prentice Hall: Upper Saddle River, NJ, 1991.

Chapter 3

Student-Prepared Formula Sheets in General Chemistry: Do They Help or Hinder Learning?

Jessica R. VandenPlas,*,[1] Jack Barbera,[2] Ky Mickelsen,[3] and Ashton Hjerstedt[3]

[1]Chemistry Department, Grand Valley State University,
Allendale, Michigan 49401
[2]Department of Chemistry and Biochemistry,
University of Northern Colorado, Greeley, Colorado 80639
[3]Department of Chemistry and Biochemistry, Northern Arizona University,
Flagstaff, Arizona 86011
***E-mail: Jessica.Vandenplas@gmail.com.**

The amount of material to be utilized by students learning chemistry is great. To this end, many instructors allow students to use "formula sheets" on quizzes and exams. Do student-prepared sheets increase knowledge retention by forcing students to evaluate the vast body of information they have studied and select only the most important pieces, aiding in information encoding? Or do students, knowing they can utilize a formula sheet of their own devising on exams, rely on these sheets as an external storage method rather than studying and learning the material? In this study the benefit of allowing students to prepare their own formula sheets was investigated. Performance on exams using student or instructor prepared sheets was compared. In addition, performance on a course final, using *only* instructor-prepared sheets, was analyzed to determine the effect of formula sheet design on long-term retention.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Introduction

In order to be successful in general chemistry, one needs to integrate a large amount of information. While instructors may stress conceptual understanding of chemistry topics (the *why* of the content), students must also use definitions, conversion factors, equations, physical constants, and problem-solving strategies. And even if all of this information is committed to long-term memory, a student must still be able to recall the information into working memory in order to solve a particular problem in chemistry. In order to ease the burden on working memory, many instructors provide students with a "formula sheet" to be used during quizzes and exams. This sheet serves as a reference for students, supplying facts or figures, equations, physical constants, etc.

Yet even when offered the seemingly beneficial tool of an instructor-written formula sheet, many students frequently request the ability to construct their own "cheat sheets." Presumably students think that by making their own formula sheet they are "getting away" with something, knowing they will be able to include extra information the instructor would not provide, such as definitions or worked examples of problems students fear they may encounter in the testing situation.

When faced with the decision of whether to allow students to create their own formula sheets, instructors may believe the myth that doing so will be detrimental to student learning. Many experienced instructors will assert that allowing students to prepare their own formula sheets will no doubt increase student performance on course exams, as it contains extraneous information that would not be provided on an instructor-prepared sheet, but that in the long run, using their own formula sheets will decrease students' long term retention by giving them an external means of storage to reference rather than committing information to long-term memory.

The literature provides us with two hypotheses with which to frame our investigation of this myth. Dorsel and Cundiff (*1*) suggest the competing dependency hypothesis and coding hypothesis. The dependency hypothesis concurs with the myth above; that is, if students know they have a formula sheet of their own making to use during testing situations, they will rely on this sheet as external storage rather than using time encoding the material to long-term memory. This does not predict if students with their own formula sheets will do better or worse than students with only instructor-prepared sheets on exams where formula sheet use is allowed. However, it *does* suggest that students who have relied on formula sheets for external storage in the past will not be be successful on exams where formula sheet use is forbidden.

Conversely, the coding hypothesis tells us that students who make and use their own formula sheets may actually be *better* at encoding this information into long-term memory than students who do not go through this process. Theoretically, choosing the most important information to include on a formula sheet from the vast body of information studied forces students to evaluate the utility of each piece of information, perhaps seeing its place in the larger body of knowledge, and even committing it to long-term memory. In this case, students who construct and use their own formula sheets will be more successful than students who use instructor-prepared sheets not only on immediate tests with the formula sheets, but even on delayed tests where formula sheets may not be allowed.

This leads us to two research questions. First, does allowing students to create their own formula sheets affect their performance on exams in a general chemistry course where their use is allowed? To answer this, performance wias compared on course exams for students using an instructor-prepared formula sheet and students using formula sheets of their own design. Second, will using instructor-prepared or student-prepared formula sheets affect long-term retention of information in general chemistry? To answer this, students who have used both kinds of formula sheets in a semester-long general chemistry course were given a cumulative final exam, during which students could only rely on an instructor-prepared formula sheet.

Methods

Two sections of a General Chemistry I (first semester general chemistry) course, taught by the same instructor, and two sections of a General Chemistry II (second semester general chemistry) course, taught by a second instructor, took part in this study. For each course, students in one section were assigned an instructor-prepared formula sheet for each course exam, while students in the second section were asked to prepare their own formula sheets for these exams. The number of students in each section are given in Table I.

Table I. Number of Students by Course and Section

Formula Sheet Type	*Gen Chem I*	*Gen Chem II*
Student-Prepared	n = 75	n = 87
Instructor-Prepared	n = 69	n = 83

Data were collected over the course of a single semester, such that students did not overlap in these courses. In order to ensure students in the two sections of each course were equivalent prior to the beginning of the intervention (use of formula sheet on exams), student performance was measured on three in-class quizzes without the use of formula sheets. Quizzes were given in each course once per week and averaged 5 questions each. Questions were written by each instructor, and included multiple-choice, calculation, or short answerquestions. Student scores were calculated based on a maximum 25 points for each quiz. Amultivariate analysis of variance (MANOVA) was used to compare section performance across the three quizzes for each course.

Instructor-prepared formula sheets contained formulas, physical constants, and conversion factors the instructor deemed pertinent to each course exam, while student-prepared sheets were allowed to contain any information the students could fit on their page. The student-prepared sheet size was limited to one half-sheet of standard 8.5” x 11” paper per exam (0.5 page for exam 1, 1.0 page for exam 2, 1.5 page for exam 3). While the exams themselves were not cumulative, the instructors felt that the cumulative nature of chemistry itself necessitated allowing students cumulative formula sheetsas the semester progressed. In addition, the instructor-prepared formula sheets became lengthier as the course material became more difficult, and allowing students extra space on their own sheets removed formula sheet length as a variable between sections.

In order to characterize the information present on both student and instructor sheets, formula sheets were qualitatively coded for the presence of facts, formulas, and example problems. The category "facts" included items such as definitions, physical constants, conversion factors, and other statements. The category "formulas" included all mathematical formulas, such as $PV = nRT$. Finally, the "example problems" category encompassed all worked examples, including examples from lecture, text, or homework.

To test the effect of formula sheet type (instructor- vs. student-prepared), section performance was compared on three hour-long course exams in each section. Course exams were a mix of instructor written true/false, multiple-choice, calculation, and short answer questions, and generally covered 3-4 chapters of material. In both courses, these exams were a mix of algorithmic and conceptual questions. Student scores were calculated out of 100 points for each exam, and a MANOVA was again used to compare section performance across all three exams for each course.

Finally, to investigate the long-term effects of formula sheet use in a course, students completed a cumulative final exam. This exam was given approximately two weeks after the last hour-long course exam, and included material from all previous exams. The final exams for each section included50 multiple-choice questions and were departmental standard exams, not written by the course instructors. The final exams covered the same topics as the hour exams, but no questions were duplicated from earlier exams. All students, regardless of section, were given a departmentally-prepared formula sheet, equivalent to the instructor-prepared formula sheets used on the hourly exams. An analysis of variance (ANOVA) was conducted to determine if students performed differently on this exam based upon their previous experience in the course using either instructor- or student-prepared formula sheets.

Results

As discussed above, the researchers wanted to ensure that the sections in each course were matched in terms of chemistry knowledge prior to the start of the experiment. To do this, student scores were taken on three in-class quizzes, and compared using a MANOVA. One MANOVA was run for each course, with "section" (instructor- vs. student-prepared sheet) as a between-subjects variable, and "quiz score" (for each of 3 quizzes) as the dependent variables. The results of this test showed no significant difference between the two sections of General Chemistry I ($F_{3,140} = 0.452$, $p = 0.715$) or between the two sections of General Chemistry II ($F_{3,166} = 0.406$, $p = 0.749$) for the main effect of section (instructor- vs. student-prepared sheet). This tells us that the sections in each course performed equally well on the quizzes overall, and were therefore matched prior to the start of the intervention.

After completing the hourly course exams, students were asked to turn in their formula sheets for evaluation. Results of the qualitative analysis of formula sheet content are given in Tables II and III. These results show that student-prepared sheets included roughly equal numbers of formulas as theinstructor-prepared sheets. As expected, however, the student-prepared sheets included a larger number of facts, sometimes by orders of magnitude, and example problems, which were not included on theinstructor-prepared sheets. This tells us that, in general, students preparing their own sheets have more information, in terms of facts, formulas, and worked examples, at their fingertips while taking the hourly course exams.

Table II. General Chemistry I, Mean Number of Items per Formula Sheet

Exam	*Sheet*	***Facts*** *Mean (SD)*	***Formulas*** *Mean (SD)*	***Examples*** *Mean (SD)*
Exam 1	Student	16 (12)	3 (3)	4 (5)
	Instructor	2	5	0
Exam 2	Student	10 (8)	8 (4)	6 (8)
	Instructor	2	8	0
Exam 3	Student	24 (14)	7 (4)	12 (13)
	Instructor	4	4	0

Table III. General Chemistry II, Mean Number of Items per Formula Sheet

Exam	*Sheet*	***Facts*** *Mean (SD)*	***Formulas*** *Mean (SD)*	***Examples*** *Mean (SD)*
Exam 1	Student	5 (5)	23 (10)	1 (2)
	Instructor	3	18	0
Exam 2	Student	12 (8)	12 (5)	2 (3)
	Instructor	4	13	0
Exam 3	Student	11 (7)	8 (4)	3 (3)
	Instructor	2	14	0

To investigate the impact of this additional information, a MANOVA was used for each course to compare section scores on the hourly course exams. On these exams, each course had one section of students using an instructor-prepared formula sheet, and one section using a student-prepared formula sheet. For these MANOVAs, "section" (instructor- vs. student-prepared sheet) was again used as a between-subjects variable, with “exam score” (for each of 3 hourly exams) as the dependent variables. The results of these tests showed no significant difference between the two sections of General Chemistry I ($F_{3,140} = 0.933$, $p = 0.427$) or between the two sections of General Chemistry II ($F_{3,166} = 1.495$, $p = 0.218$) for the main effect of section (instructor- vs. student-prepared sheet) on exam score. The results show that students in each course performed equally well on the hourly exams, regardless of the type of formula sheet used. Means and standard errors for each section’s performance on these exams are given in Table IV.

Finally, to study the effects of formula sheet design on long-term retention, student scores on the cumulative final exam were compared. On this test, all students used an instructor-prepared formula sheet, but sections were again compared based on the type of formula sheet used throughout the semester. In this case, one ANOVA was completed for each section, using final exam score as the dependent variable, and “section” (instructor- vs. student-prepared sheet) as the independent variable. Mean and standard errors for the final exam scores are given in Table V below. Although students who had used student-prepared formula sheets during the semester averaged a few points higher on the final exam in both courses than those who had used instructor-prepared sheets, this difference was not found to be significant in either section of General Chemistry I ($F_{1,142} = 0.241$, $p = 0.624$) or General Chemistry II ($F_{1,168} = 0.068$, $p = 0.794$). This suggests that use of either formula sheet design during the semester did not help or hinder students’ long-term retention or performance on the final exam.

Table IV. Student Hourly Exam Scores by Formula Sheet Type

Course	*Exam*	*Sheet*	*Mean*[a]	*Std. Error*
Gen Chem I	Exam 1	Instructor	76.7	1.8
		Student	78.0	1.8
	Exam 2	Instructor	73.4	2.1
		Student	77.0	20
	Exam 3	Instructor	68.9	2.1
		Student	68.2	2.0
Gen Chem II	Exam 1	Instructor	81.4	1.7
		Student	80.4	1.6
	Exam 2	Instructor	74.1	1.8
		Student	70.9	1.8
	Exam 3	Instructor	76.6	2.1
		Student	78.0	2.0

[a] 100 points possible

Table V. Student Final Exam Scores by Formula Sheet Type

Course	*Sheet*	*Mean*[a]	*Std. Error*
Gen Chem I	Instructor	82.2	3.2
	Student	84.4	3.1
Gen Chem II	Instructor	96.9	3.2
	Student	98.0	3.1

[a] 150 points possible.

Discussion

Overall, students did not show increased performance on course exams when given the opportunity to create and use their own formula sheets. While this has been demonstrated in other studies (*2–5*) for subjects less mathematical than chemistry, there has been pervasive "conventional wisdom" that allowing students to prepare their own formula sheets gives students an unfair advantage over students using an instructor-prepared sheet. This myth seems to be untrue, at least as far as instructor-written course exams are concerned. Although students who wrote and used their own formula sheets had much more information at their disposal during exams, these students did not perform statistically significantly different than students without access to this extraneous information.

In addition, the myth that students who use formula sheets will rely on this external storage method rather than learning the content themselves also seems to be a myth. In this case, students did not show a significant change in performance on long-term testing measures regardless of previous formula sheet experience. While it could be argued students *knew* they would not be able to use their own formula sheets on the final exams, and therefore studied differently for the final than they had for the hourly exams, the cumulative final, given only two weeks after the third hourly exam, contained a large amount of material with little time to study. It is unlikely that students would have performed as well as they did on the final exam if they had relied on their own formula sheets as a form of external storage rather than commiting the knowledge to their long term memories.

Conclusions

Although formula sheets were not shown to affect either current or long term course achievement, the use of formula sheets may affect other aspects of student performance. Future studies will further investigate what, if any, effect formula sheet design and use might have on student knowledge attainment. For example, formula sheet design may affect performance on conceptual questions differently than on algorithmic questions, a difference that could have been lost when exams were analyzed as a whole. And if formula sheets do not affect students' academic achievement, does allowing students to prepare their own formula sheets affect some other variable that makes the use of formula sheets advantageous? Student self-efficacy, test anxiety, and cognitive load are all variables that may be affected, positively or negatively, by formula sheet design. Based on the results of this study, it seems that allowing students to produce their own formula sheet does not give them an unfair advantage during use, or hurt students' long-term retention of chemistry content. If student-prepared formula sheets can be shown to affect another variable, by decreasing test anxiety for example, their use may actually be beneficial, contrary to what the myths regarding formula use may suggest.

References

1. Dorsel, T. N.; Cundiff, G. W. *J. Exp. Educ.* **1979**, *38*, 39–41.
2. Dickson, K. L.; Miller, M. D. *Teach. Psychol.* **2005**, *32*, 230–233.
3. Hindman, C. D. *Teach. Psychol.* **1980**, *7*, 166–168.
4. Trigwell, K. *Assess. Eval. Higher Educ.* **1987**, *12*, 56–65.
5. Whitley, B. E. *Coll. Student J.* **1996**, *30*, 489–493.

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.ch005

Chapter 5

Development and Refinement of a Research Study Assessing Student Attention in General Chemistry

Kelly Y. Neiles,[1] Elizabeth A. Flens,[1] Diane M. Bunce,[*,1] and Michael Ferguson[2]

[1]Chemistry Department, The Catholic University of America, Washington, DC 20064
[2]Science Department, DeMatha Catholic High School, Hyattsville, Maryland 20781
[*]E-mail: bunce@cua.edu

Planning an experiment and collecting and analyzing data is only half the process of doing research. Progression from research ideas to a published manuscript is an iterative process. The purpose of this chapter is to describe the decisions in that process including the development and refinement of the research questions, methodology, data analysis, and conclusions that must be completed before an article is likely to be accepted for publication. These decisions will be discussed within the context of a research experiment dealing with an innovative use of a measurement tool to collect data on student attention lapses in general chemistry. Use of reviewer comments in the journal submission process to improve the manuscript is discussed in terms of the changes that were made in the final version of the manuscript prior to acceptance for publication.

Introduction

Journal articles often sound as if the researchers progressed in a straight path from the research idea to conclusions with everything else impeccably planned and executed. This can be daunting to a new researcher. The purpose of this chapter is to provide the reader with a more realistic view of how one of our published manuscripts progressed from initial idea to published article. Inherent in this

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

iterative process is a series of decisions that we made as researchers, sometimes in response to theory and other times in response to the reality of the research situation. When we thought we had a well written and polished product, we submitted it to a journal for review. It was surprising to us that the reviewers had so many suggestions for improvement on what we had considered to be a finished product. After we reflected on the reviews, we too agreed that we could improve our product.

The following section discusses how we transitioned from an initial idea to a research question based upon the literature on this topic.

Theory and Literature Review

As teachers, we have probably all used lectures as part of our teaching approach. It is curious that student achievement does not always reflect the quality of those lectures. The question arises as to whether or not the students were paying attention in lecture. Lapses in student attention have been thought to be a result of passive lectures where students are not actively involved in the learning process. Books on how to lecture effectively suggest that if lectures are passive presentations of information then student lapses are more likely to occur (*1*). The possible reasons for these lapses in student attention may be a result of how the brain operates and how learning takes place. For instance, student attention lapses may be a function of an overload of the working memory; student lack of interest in the topic; or the passive role of the student in the teaching pedagogy used.

Limited working memory, one of the possible reasons for student attention lapses, is the term used to describe the process by which incoming information is integrated with information retrieved from the student's long term memory. Too many stimuli presented at once without allowing time for students to process the incoming information can overload the working memory capacity (*2*). The working memory has been shown to have a capacity of 7±2 items. This capacity can quickly be exceeded in a 50 minute lecture through the normal process of listening, taking notes and thinking about multiple concepts without time to process and clear the working memory space (*3*). If working memory capacity is exceeded, then students will have a great deal of difficulty paying attention. Fluctuations in students' attention may reflect the students' efforts to process the continuing flow of information while dealing with the limitations of working memory.

Another reason for students' inability to maintain attention in lecture may result from students' level of interest in the topic. Here the teacher's role is crucial. The teacher may be able to influence students' initial interest level in a topic by helping them see its relevance. Interest according to Ainley, Hidi, & Berndorff (*4*) is characterized by focused attention, increased cognitive and affective functioning, and persistent effort. As a result, if students are interested in a topic, this may result in increased attention during lecture.

Even if students are interested in a topic, the length of a typical lecture can be overwhelming. Students in a passive mode can experience attention lapses even if they are interested in the topic. Brophy (*5*) attributes this lack of attention to students not being required to use their cognitive abilities during lecture. Others agree that attention capacity can vary depending on students' motivation to learn the information presented (*1*). A passive learning experience allows students to become uninvolved. Active learning experiences draw students into the learning process and encourage them to learn the information (*6*). These learning experiences prevent students from becoming passive and thus may be responsible for helping engage students in learning concepts that they may not have been initially motivated to learn. In essence, the teacher in an active learning process has involved students in a manner that a passive lecture may not (*6*).

Other variables affecting student attention can be more personal in nature and may include lack of sleep, poor nutrition, general health, or complications from personal relationships, among others. These variables are more individualistic in nature and harder to measure.

Previously attention lapses were identified by researchers who either observed or videotaped students during lecture (*7*). Here the occurrence of attention lapses were determined by observing a student's facial expressions. A possible threat to the validity of this approach is the fact that the student's facial expression may not be a true reflection of an attention lapse. Student input in the determination of attention lapses is crucial. Measuring student attention lapses must include accurate identification of lapses. In addition the measurement technique should include unobtrusive methods of data collection and data measured over an extended period of time, occasions, and with different teaching pedagogies. Such parameters call for new methods of data collection using more automated systems so that multiple data points can be recorded with minimal interference to the learning process.

Development of Questions and Methodology

As teachers we sometimes notice that students do not appear to pay attention during the entire length of a class. Even as faculty members in a seminar audience we may find it difficult to pay attention during the entire seminar. Are such lapses part of human nature? If so, then how often do such attention lapses occur? We were interested in investigating these questions in regards to students enrolled in general chemistry classes. We were also interested in investigating whether the number of student attention lapses is affected by using different teaching pedagogies within a given class.

A search of the literature produced few studies where student attention lapses were actually measured but quite a few references where definitive advice was offered on how to increase student attention during lecture (*1*, *8*). This situation increased our interest in developing both a way to measure student attention lapses during a class and the effects of different teaching methods on these lapses. Once we decided the main focus of our research, we developed the following four main research questions (*9*):

1. Does student attention remain constant during a general chemistry class?
2. Are there differences in the length of attention lapses reported?

3. Is there a difference in student attention during different teaching pedagogies within a given class?
4. Is there a difference in student attention during the beginning, middle or end of a class?

Although we were satisfied with our progress in developing the research questions to this point, it soon became obvious that a number of decisions regarding the research project were in order. These decisions included developing operational definitions of parameters involved in this study. To develop valid terminology that students could relate to, we invited general chemistry students to meet with us and discuss their understanding of the descriptions we would later use to describe our study to the students in the study. As a result of this meeting, we learned that the term “attention lapse” was not as effective in communicating our plan to students as was the phrase “zoning out”. The students we interviewed also discussed what might constitute short, medium or long lapses of attention in a class. This discussion helped insure that the student directions we wrote for the study would be better understood by the student participants.

We realized too, that we needed to differentiate among commonly used terms such as “lecture”, “lecture segment”, “class”, and “course” for this study. Although these terms are sometimes used interchangeably in casual conversation, in our experiment, they had specific and unique meanings. We recognized the need to be consistent in our use of these operationally defined terms. Our operational definitions for these terms are presented here (*9*):

Lecture: traditional pedagogical approach involving the teacher presenting information to an audience. The flow of information proceeds from teacher to student.

Lecture segment: portion of the class devoted to lecture pedagogy.

Demonstration: use of chemicals or models to present a visual presentation of the chemistry concept being presented.

Clicker Question: ConcepTests (*10*) presented electronically through the use of personal response devices.

Class: the full length teaching session. In this study, all classes had a duration of 50 minutes.

Course: the semester-long curricula, which for this research included general chemistry for engineering students; general, organic and biochemistry for nursing students; and chemistry for nonscience majors.

Pedagogical approach: any unique presentation or interaction between teacher and student during a class. Typical examples include lecture, demonstration, clicker questions, working in student groups/pairs, inclusion of real world applications, personal vignettes, or announcements.

Sample

To increase the generalizability of the study, we decided to sample a range of general chemistry courses including general chemistry courses for engineering, nursing and nonscience majors. In our institution, these three populations have separate general chemistry courses which we will refer to as Chem I (engineering students) (n= 74); Chem II (nursing students) (n= 68); and Chem III (nonscience majors) (n= 44). Including all three of these courses in our research had the added advantage of using courses taught by more than one teacher. Chem I was taught by teacher A and teacher B taught both Chem II and Chem III. The research idea was presented to the two teachers and their cooperation was secured.

Institutional Review Board (IRB)

All research involving human participation is required to be reviewed and approved by an Institutional Review Board (IRB). Research is judged as being either exempt or nonexempt from the full IRB committee review according to the guidelines set by the National Institutes of Health. Our research was determined to be exempt because only normal educational practices were employed in the study and no student was identified through the collection and reporting of data. Another key element of our exempt status was that all participation was voluntary and participant identity was not revealed to the teachers in the study.

Methodology and Data Collection

Personal Response Devices (Clickers)

In an effort to measure self-reported student attention lapses during class over an extended period of time, a data collection process was devised that would cause minimal disruption in each of the courses. Our familiarity with personal response devices (clickers), led us to the idea that students' reports of attention lapses could be collected through the use of a dedicated class set of clickers

A research proposal was submitted to the clicker company who agreed to a no cost one-semester loan of a class set of clickers plus a radio frequency receiver for use in this study. Responses from these clickers were recorded by a receiver on a laptop tablet PC in the back of the room. This tablet PC was also used by the researchers to record different pedagogies used by the teacher during each class. Student clicker responses were surveyed every 30 seconds by setting up a PowerPoint presentation on the tablet PC that automatically advanced a PowerPoint slide every 30 seconds. Changes in teaching pedagogy were recorded with the tablet PC stylus on the appropriate slide in that PowerPoint file.

In two of the courses studied, students owned a personal clicker and used it regularly in class. In the third course, students did not own personal clickers. All three courses were provided with a set of clickers on lanyards for this research study. Students used the lanyard clickers exclusively to record attention lapses. This class set of lanyard clickers was distributed at the beginning and collected at the end of each class.

Directions to Students

Students were instructed to select Button #1 on the lanyard clicker if they believed their attention lapse was 1 minute or less; Button #2 if the lapse was 2 to 3 minutes and Button #3 if the lapse was 5 minutes or more. To help students understand what typical lapses of these durations might be, examples of each type of attention lapse were suggested. These suggested attention lapses were based on examples that the students we consulted at the beginning of the project provided. For example, Button #1 (lapse of 1 minute or less) might include looking at a clock/watch, reading a text message, or daydreaming. Button #2 (2 to 3 minute lapse) might be typing a response to a text message. Button #3 (5 minutes or more) might include working on assignments for other classes or falling asleep.

Students were reminded in class on a regular basis to use their lanyard clickers to record attention lapses when they occurred. Reminders were delivered verbally when clickers were handed out at the beginning of class, by the instructor during class, or as a footnote on the teacher's PowerPoint slides used in class.

Identifying Teaching Pedagogies

Inter-rater reliability was calculated on the start time and duration of the different pedagogies used during the class. The use of several identifiable pedagogies within a single class period resulted in a class being divided into pedagogical segments. For example, a single class could contain several segments of different pedagogies including multiple segments of a specific pedagogy. During the first week of data collection, inter-rater reliability did not meet acceptable research standards. Meetings were held with the researchers to review what constituted a pedagogical change. This resulted in an acceptable inter-rater reliability statistic during the remaining data collection.

Other Variables

In two of the three courses (Chem II (nursing) and Chem III (nonscience)), data on additional variables were collected. In these courses, students routinely completed a diagnostic test of logical thinking called the Group Assessment of Logical Thinking (GALT) test during the first two weeks of the course (*11*). The GALT is a 12-question Piagetian test that measures students' use of logical thinking. The GALT test was delivered electronically in this experiment. The possible scores range from 0 to 12, and for 10 of the 12 questions include the selection of a correct answer and a correct reason for that answer. The last two questions involve the ability to group variables. Student GALT scores were graphed and natural breaks in the score distribution were used to establish low (1-6), medium (7-9), and high (10-12) GALT scores.

In these same two courses (Chem II (nursing) and Chem III (nonscience)), final course averages were used as indicators of achievement. Student averages were graphed and natural breaks in the distribution were used to establish low (0-75%), medium (76%-85%), and high (86%-100%) achievement scores. Gender was also noted in these two courses.

GALT, gender and achievement could not be used for the third course (Chem I (engineering)) because clicker use in this course was not part of the regular teaching scheme and no clickers were registered to any particular student. Thus, identifying data could not be recorded and specific students could not be tracked.

Timeline

Data were collected in each course three days a week for 6 weeks. Data from the first two weeks of the experiment were used in a formative manner to both help refine the methodology and familiarize students with the use of clickers. The remaining 4 weeks of data collection were used in the analysis.

Data Analysis

Data Reduction

Data collected on some days were eliminated from the analysis due to incomplete data sets. This was a result of the teacher scheduling short lab experiences during class, use of group worksheets, or researcher error in data collection on that day. The long data collection time period of this research helped ensure that such glitches did not seriously affect the integrity of the data.

This particular experimental design provided a relatively large data set. Although this was a plus, it also meant that we had a large, almost overwhelming data set to analyze. As in most experiments, a method had to be devised to reduce the data in a way that was in accord with the prerequisites of the statistical methods used. Data reduction involved refinement of research questions and the definition of conditions under which these questions were addressed.

To start the data reduction process, we reviewed the number and frequency of pedagogies used in the three courses of this study. We documented fourteen types of pedagogies and made the decision to analyze only the three most frequently used pedagogies. These three pedagogies common to at least two of the three courses were lecture, clicker questions, and demonstrations. Since students recorded more than one response per class over multiple days, the appropriate statistic to analyze these data is a repeated-measures ANOVA (*12*). This statistical procedure is predicated on the analysis of equal segment lengths. The longest common segment of lecture, clicker question session, or demonstration that was common to all classes within a course was chosen as the unit of analysis. When different pedagogies were compared to lecture segments, a new common segment length for both the pedagogy and lecture segment was established as the unit for that analysis. An in-depth discussion of the analyses used in this study can be found in Bunce, et al (*9*).

Refining the Research Questions

Once we looked at the data initially, we developed additional questions we could ask of the data. This led us to a series of subquestions that could be addressed through the statistical procedures we would use. The additional questions that we asked of the data in relation to the four general questions are in bold face and included in the following (*9*):

1. Does student attention remain constant during a general chemistry class?
    - a. **Is this attention different for different gender, GALT or achievement levels?**

2. Are there differences in the length of attention declines reported?

3. Is there a difference in student attention during different teaching pedagogies within a class?
    - a. **Does the use of different pedagogies affect the attention lapses reported during subsequent pedagogies?**

4. Is there a difference in student attention during the beginning, middle or end of a class?
    - a. **Is there a difference in the number of short, medium and long duration attention lapses during short, medium, and long lecture segments?**

Checking Assumptions of Statistical Procedures

If statistical procedures are used without first checking that the data do not violate the assumptions of the statistical procedure, misleading interpretations of the statistic can result. Violation of assumptions for a statistical procedure can range from a minimal effect on interpretations of results to the fact that the data may not be appropriate for a given statistic. Good practice requires that researchers be aware of violations of assumptions if present and either choose a different statistic or modify their conclusions accordingly.

In this experiment, the choice of the repeated measures ANOVA required checking the following data assumptions: 1) independence of observations; 2) normal distribution of data; and 3) homogeneity of variance (*12*). Our data set met all three assumptions as described in Bunce, et al (*9*).

Results and Conclusions

The results presented here are a summary of those reported in Bunce et al. (*9*). Repeated measures ANOVAs were used to address the questions we investigated as follows (*9*):

1. Does student attention remain constant during a general chemistry class?

The three repeated measures ANOVAs (one for each course, Chem I (engineering), Chem II (nursing), and Chem III (nonscience)) produced the following results presented in Table I:

Table I. Student Reported Attention Lapses During Class

Course	*F (df)*[a]	*Significance*
Chem I (engineering)	1.22 (26, 2863)	0.205
Chem II (nursing)	1.60 (19, 1313)	**0.047**
Chem III (nonscience)	0.86 (9, 641)	0.571

[a] Degrees of freedom reflect the number of student responses over the course of the experiment and not the number of students enrolled in the course. Bolded value is significant at $p<0.05$.

The results show that there is a significant difference in self-reported student attention lapses during a class in only one course (Chem II (nursing)). This result may be due to the fact that the largest student participation occurred in this course compared to the others. Chem II had an average daily participation of 56% compared to Chem I (engineering) of 23% and Chem III (nonscience) of 27%. This significant result in Chem II can be interpreted as showing that there is a fluctuation between attention and attention lapse during the class.

1a. Is there an interaction between student attention and each of the following: gender, GALT and achievement?

In this study, the one course (Chem II (nursing)) that demonstrated a significant effect and which had the largest student participation was further analyzed for interaction effects with gender, GALT and achievement. The results indicate that there is no significant difference between male and female reports of attention lapses in Chem II, $F(1,49)=0.01$, $p=0.917$. When GALT is used as a control variable, there is no significant difference in the reporting of attention lapses among low, medium, and high GALT groups for Chem II, $F(2,49) = 1.02$, $p=0.369$. When achievement is entered as a control variable, no significant difference in self-reported attention lapses was found among low, medium, and high achieving groups for Chem II, $F(2,49)=2.21$, $p=0.121$. These results suggest that attention lapses are independent of gender, GALT level, or achievement for this course.

2. Are there differences in the length of attention declines reported?

Since Chem II (nursing) was the only course that showed significant differences in the overall reporting of attention lapses throughout lecture segments,

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.ch005

only data from this course was used to investigate the question regarding length of lapses. Students registered differing lengths of attention lapses by pressing one of three buttons on their research lanyard clickers. A repeated-measures ANOVA was used to compare the means of these responses of different attention lapse length. The results indicate that the shortest attention lapse (1 minute or less) was reported by students at a significantly higher rate than the medium (2-3 minute) and long lengths (5 minutes or longer), (F(4,324)=8.30, p=0.000). This suggests that the most prevalently reported attention lapse is of short duration (1 minute or less).

3. Is there a difference in student attention during different teaching pedagogies within a class?

To analyze this question using a repeated measures ANOVA, students' reporting of attention lapses during student-centered pedagogies (demonstrations and clicker questions) were compared to their reporting of attention lapses during lecture segments within the same class. The pedagogies used in each course are presented in Table II. In one of the courses, Chem III (nonscience), all three pedagogies (lecture, demonstration and clicker questions) were used by the teacher. In Chem I (engineering), the teacher used only lectures and demonstrations, and in Chem II (nursing), the teacher used lectures and clicker questions. This resulted in analyses of only one student-centered pedagogy (demonstration or clicker question) for Chem I (engineering) and Chem II (nursing) compared to lecture segments. In Chem III (nonscience) both demonstrations and clicker question pedagogies were compared to lecture segments.

Table II. Pedagogies Used in Each Course[a]

	Lecture	*Clicker Questions*	*Demonstrations*
Chem I	x		x
Chem II	x	x	
Chem III	x	x	x

[a] Reprinted with permission from Bunce, D. M.; Flens, E. A.; Neiles, K. Y. How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education*, **2010**, *87* (12), 1438-1443. Copyright 2011, American Chemical Society.

In Chem II (nursing) and Chem III (nonscience), the effect of the clicker pedagogy was significant (Chem II (nursing), F(1,67)=26.71, p=0.000, and Chem III (nonscience), F(1,67)=6.93, p=0.011). In Chem I (engineering) and Chem III (nonscience), the effect of the demonstration pedagogy was significant (Chem I (engineering), F(1,86)=7.22, p=0.009 and Chem III (nonscience), F(1,67)=5.46, p=0.022). These results indicate that the use of student-centered pedagogies (demonstration or clicker questions) significantly reduces the number of attention lapses reported by students when compared to the number of lapses reported during lecture segments.

3a. Does the use of student-centered pedagogies affect attention lapses during subsequent lecture segments?

To address this question, only portions of lecture segments both before and after student-centered pedagogies (demonstrations or clicker questions) were used in the analysis. Lecture segments had to be equivalent in length to that of the student-centered pedagogy to satisfy the prerequisites of the statistic. A repeated measures ANOVA was used in this analysis. In Chem II (nursing) where clicker questions were used, a significant decrease in the number of attention lapses was reported in lecture segments that occurred *after* versus *before* the use of a clicker question (F(1,66)=8.70, p=0.004). In Chem III (nonscience) where demonstrations were used, a significant decrease in the number of reported attention lapses was also found in lecture segments that occurred *after* versus *before* a demonstration (F(1,64)=4.25, p=0.043). These results indicate that the use of student-centered pedagogies decreases the number of reported attention lapses not only during the pedagogy itself, but also in lecture segments that immediately follow the student-centered pedagogy compared to those in the lecture segments prior to the examined pedagogy.

4. Is there a difference in student attention within a lecture segment at the beginning, middle or end of a class?

In Chem II (nursing), a repeated measures ANOVA showed an overall significant difference in the reporting of attention lapses in the lecture segments at the beginning, middle and end of class (F(2,324)=12.78, p=0.000). A post hoc analysis indicated a significant difference between the beginning and middle of a class with a significantly higher number of attention lapses reported during the middle of a class. A similar effect is seen between attention lapses reported in the lecture segments in the middle vs. the end of a class though it is not a significant difference. This is interpreted as the middle part of a class having a significantly higher number of self-reported attention lapses than a lecture segment at the beginning of a class.

4a. Is there a difference in the length of self-reported attention lapses during lectures of differing lengths?

In Chem II (nursing), the length of student reported attention lapses was compared within lectures of differing length. This analysis investigated whether the length of attention lapse was affected by the length of the lecture segment. The results of this analysis are presented in Table III. Overall, there is a significant difference in the length of attention lapses reported in lecture segments of varying lengths (F(4,324)=8.30, p=0.000). Due to the small n and the large number of possible comparisons which would result in a high Type I error, post hoc analysis was not pursued. Based on the means of attention lapses in lectures of varying lengths, the trend appears to show that short attention lapses are the most prevalent type of attention lapse in lectures of different lengths. Together with the previous analysis that showed a significantly higher occurrence of short attention lapses throughout the length of an entire class, this result showing a trend for short attention lapses during lecture segments of varying lengths appears valid.

Table III. Difference in the Length of Attention Decline in Short, Medium, and Long Lectures in Chem II (Nursing)

Length of lecture	*Length of attention decline*[a]	*Mean*	*F (4,324)*	*Sig.*
Short	1	1.62	8.30	**0.000**
	2	0.41		
	3	0.09		
Medium	1	5.18		
	2	1.65		
	3	0.63		
Long	1	3.23		
	2	1.12		
	3	0.59		

[a] Length of attention decline: 1=1 minute or less; 2=2-3 minutes; 3=5 minutes or more Bolded value is significant at $p<0.05$.

Discussion of Results

In general, the analysis of the data was limited to one course (Chem II nursing) that showed a significant overall effect for attention lapses over time (*9*). This was also the course with the highest daily average participation. Similar but nonsignificant trends were also evident in the two other courses Chem l (engineering) and Chem III (nonscience).

Limiting the discussion to Chem II (nursing), we see that student attention fluctuates over time during the length of a typical class. This fluctuation peaks with the largest number of self-reported attention lapses typically occurring during the middle of the class. The peak during the middle of class is significantly higher than self-reported attention lapses at the beginning. Based on this result, teachers should be aware that student attention lapses are most likely to occur during the middle of a class and plan to re-engage students at this point.

Within a lecture segment, the majority of self-reported attention lapses were characterized by students as short duration (1 minute or less) regardless of the length of the lecture segment studied. This means that when students experience attention lapses, these lapses are likely to be short and students should be able to

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.ch005

reestablish their attention quickly. Since students operate individually in terms of their pattern of attention and attention lapses, a teacher is faced with some students being engaged at a given moment and others who are not. Teachers may not be aware of the attention and attention lapse cycle of the students in the class because they may be focusing on only those students who appear to be fully engaged or non-engaged in the lecture at any one time.

When gender, GALT (logical reasoning ability) level and achievement variables are investigated for two courses (Chem II (nursing) and Chem III (nonscience)), none of these three variables significantly affects the reported attention lapses. For example, women do not report more or fewer attention lapses than men at any given time within a lecture segment. Students who perform at a low GALT (logical reasoning ability) level do not report a significantly different number of attention lapses than students who perform at a high or middle GALT level. This is also true for students of differing achievement levels. All students regardless of gender, logical reasoning ability, and achievement report comparable number of attention lapses during lecture segments.

In the analysis concerning the use of different teaching pedagogies (demonstrations and clicker questions), data from all three courses were used. Here, self-reported attention lapses for the two student-centered pedagogies compared to lecture segments of comparable length, are significantly fewer than during lecture segments. This significant reduction in the number of attention lapses is still evident in lecture segments that follow the use of these pedagogies. This suggests that these alternatives to lecture (demonstration and clicker questions) are more successful at engaging students and reducing the occurrence of attention lapses. This positive effect carries over to subsequent lecture segments. Teachers who use a variety of student centered pedagogies during a given class could expect increased student attention overall.

The Review Process

No matter how perfect a manuscript appears to the authors when submitted for publication, objective reviewers are often able to point out inconsistencies or deficiencies in the manuscript. Authors very often are resistant to these critiques but if a manuscript is not able to convey a complete and convincing scenario to someone not directly involved in the research, then there is room for improvement. Knowing how to interpret and respond to such reviews is an important learning process for the researcher.

In this manuscript, the description of the analysis and results was complex. The reviewers asked pointed questions that helped us understand where we were losing the reader. We were able to develop a more expanded description of what we did statistically and explain it so that someone not familiar with the statistics used could better understand the analysis. The data reduction process in the originally submitted manuscript was one area that also needed more explanation. In the methodology section, we had to more fully explain our method of data collection that used research clickers and a receiver to collect data automatically every 30 seconds within a PowerPoint framework. Questions concerning the reported percentage of student responses also required more clarification. The usual formatting and overlooked grammatical errors were included in the reviews. Although seemingly painful to construct, the resulting revised manuscript was considered an improvement by the authors as well as the reviewers.

Published reports of research may present a polished, coherent picture executed through a series of seemingly sequential steps. The reality of doing research from conception, data collection, analysis, submission, revision and ultimately to publication do not always follow such a direct path. New researchers may feel that they are not successful when their research does not seem to follow a direct path. In reality, very few people's research does. The secret to becoming a successful researcher is to keep trying and learn something new with each turn in the road to completing and publishing research.

Acknowledgments

We would like to thank the teachers and students who participated in this study and Turning Technologies, Inc. for the loan of the class set of clickers used in this study.

References

1. McKeachie, W. J. *McKeachie's Teaching Tip*; Houghton Mifflin Company: New York, 2002.
2. Miller, G. A. *Psych. Rev.* **1956**, *63*, 81.
3. Dehn, M. J. *Working Memory and Academic Learning: Assessment and Intervention*; John Wiley & Sons, Inc: Hoboken, NJ, 2008.
4. Ainley, M.; Hidi, S.; Berndorff, D. *J. Ed. Psychol.* **2002**, *94*, 545–561.
5. Brophy, J. *Motivating Students to Learn*, 2nd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, 2004.
6. McCombs, B. L.; Miller, L. *Learner-Centered Classroom Practices and Assessments*; Corwin Press: Thousand Oaks, CA, 2007.
7. Johnstone, A.; Percival, F. *Educ. Chem.* **1976**, *13*, 49–50.
8. Bligh, D. A. *What's the Use of Lectures?*; Jossey-Bass: San Fransisco, CA, 2000.
9. Bunce, D. M.; Flens, E. A.; Neiles, K. Y. *J. Chem. Educ.* **2010**, *87*, 1438–1443.
10. Mazur, E.; Somers, M. *Am. J. Phys.* **1999**, *67*, 359–360.
11. Roadrangka, V. The Construction and Validation of the Group Assessment of Logical Thinking (GALT). EdD. Dissertation, University of Georgia, 1985.
12. Weinberg, S. L.; Abramowitz, S. K. *Statistics Using SPSS: An Integrative Approach*, 2nd ed.; Cambridge University Press: New York, 2008.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.

Chapter 13

Diverse Methodologies Used To Challenge Myths

Jessica R. VandenPlas[*,1] **and Diane M. Bunce**[2]

[1]**Chemistry Department, Grand Valley State University, Allendale, Michigan 49401**
[2]**Chemistry Department, The Catholic University of America, Washington, DC 20064**
[*]**E-mail: vandenpj@gvsu.edu**

The chapters in this book present a wide range of research questions and methodologies challenging commonly held myths of teaching and learning. There are many valid research designs that can be used to conduct quality research including quantitative methods, qualitative methods, test analysis theories, and in many cases utilizing multiple experiments to address complicated questions. Each chapter in this book can be viewed independently as an example of a typical type of research. This final chapter will revisit the research presented previously and discuss it in terms of the research methodologies used. The purpose is to help the reader better see the infrastructure of the research presented and how the methodology employed matches the research questions asked.

Introduction

One of the purposes of this book was to provide reports of research that addressed some popular myths about teaching and learning. The questions investigated included the following: How long do students pay attention in lecture? How long do they retain knowledge after a test?; Can students do inquiry even if their teachers don't think they can? Do students do better in semester-long or intensive summer school chemistry courses? Do student-prepared formula sheets help or hinder students' achievement on tests compared to formula sheets prepared by teachers? Is the same amount of content taught in an activity-based vs. a lecture-based course? Can we analyze test questions to help identify topics that are difficult for students?; Can standardized tests be used to gauge whether we as a field have made progress teaching conceptual understanding? What are effective ways to choose and implement different visualization methods in chemistry? What is the process of developing valid and reliable assessments? How can surveys be used to track changes in student use of curriculum options over time?

This is a diverse list of questions that are tied to myths (cultural beliefs) that research can investigate and challenge. The take home message of the book is that any belief that appears to explain some aspect of the teaching/learning process is worth investigating in an attempt to place it on a firmer footing than hearsay alone. The goal is to put the process of teaching and learning on a firm foundation supported by research results. This is easier said than done. Teaching and learning are complicated processes that encompass a large number of variables, only some of which we may be aware of. These variables may include student aptitude; student self confidence in pursuing understanding of difficult topics; time available to learn the material; the way the material is presented in terms of visual, audio or kinesthetic approaches; the framework of the course and the materials provided to support learning; the empathy and support provided by the teacher; and the appropriateness and accessibility of testing measures. In some situations, we may not even be aware of what the variables are as well as not knowing how to adequately measure or control them.

In reading the research presented in this book, we can analyze both the types of questions asked and the research methodologies employed. No two studies reported here are alike. They all ask different questions and use different methodologies that help address these unique questions. Rather than just looking at the results of these studies, this chapter will analyze the methodologies used to address the questions raised by the researchers. The purpose of this chapter, then, is different from the rest of the book. Here we are more interested in *how* the questions were addressed rather than *what* the results of the research were. To this end, we will revisit the chapters analyzing the research methodologies employed and how these methodologies were used to address the questions asked.

Quantitative, Statistical Research Designs

One of the most common research designs, shared by several studies in this book, involves the collection of quantitative data that can be analyzed using inferential statistics (*1*). Quantitative data can be collected in many ways, from traditional in-class quizzes and exams, to novel methods, such as the use of clickers described in the chapter by Neiles et al. (Chapter 5). Before any quantitative data is collected, researchers have many decisions to make in how to process and analyze the data to best address the research question being asked. One of the most common ways of dealing with quantitative data is to use significance testing to identify differences, or the lack thereof, among datasets. Such methods were used to address several research questions in this book including topics such as student achievement, course design, and attention allocation. Each study applied these methods in a unique manner that was determined by the question that was being asked.

Studies Utilizing Quantitative Methods

The study by Bunce and VandenPlas (Chapter 2) was conducted to investigate how student knowledge changed over time. In order to address such a research question, the researchers must first operationally define terms such as "knowledge" for the purposes of the study. In this case, the researchers used student performance on in-class quizzes and exams as a way of approximating student knowledge. The quiz and exam questions used for the study were short answer questions, providing students' free responses as data for the investigation. These data could be analyzed several ways, ranging from a direct qualitative analysis of the language used in the answers, to the more quantitative methods analyzing achievement employed in this study. In this case, the student answers were rubric-scored by the researchers, a common method of converting free responses (qualitative data) into numerical (quantitative) data for the purpose of statistical analysis. This is one way to simplify a large data set for a more manageable analysis. Here it also allowed the researchers to compare student scores across testing occasions in a statistical manner. This provided a clearer picture of the changes that occurred and how these were influenced by variables such as course type and students' logical reasoning ability.

Because these data were analyzed in a quantitative manner, the researchers were able to document the trends observed in student scores (no significant decay for nursing and high school students, but immediate decay for nonscience majors). Although the trends could be identified, the *reasons* for these trends could not. This would require asking a different research question: *why* student knowledge changed over time rather than *whether* it changed over time. Clearly, one must document *if* a change occurs before asking *why* it occurs. Collecting student quiz/exam responses was a reasonable method of determining whether students' knowledge changed over time, but additional data would be needed to answer the question of why this happened in some classes and not others.

As a start towards addressing competing explanations for the resulting data, the researchers collected and analyzed additional qualitative data in the form of student evaluations. This was used to investigate the hypothesis that student motivation may play a role in how knowledge changes over time. The evaluation did help to eliminate some possible explanations for why the data provided the trends it did. Because different measurement tools were used, the study by Bunce and VandenPlas is considered a mixed methods design (*2*), utilizing both quantitative analyses (statistical analyses of student quiz and exam scores) and qualitative analyses (rubric-scoring student free responses and a survey to collect data on student motivation). Mixed methods designs allow for triangulation of data. Data triangulation is the process of collecting data from more than one source or instrument and analyzing it to reveal inconsistencies, identify contradictions, or to provide complementary data to support interpretation (*3*). Triangulation also helps to control overall threats to validity by balancing the individual threats of different data collection methods. Mixed methods designs can therefore often provide a more complete picture of the research questions posed, and the soundness of the data interpretation by the researchers.

In the study conducted by VandenPlas et al. (Chapter 3), the authors also addressed a research question regarding student achievement. In this case, the authors had competing hypotheses regarding how students would perform in both the short-term and long-term after using formula sheets on examinations. Both competing hypotheses predicted students would do better on examinations when they had prepared their own formula sheets, either because students had spent extra time organizing their knowledge or because the formula sheet served as an external storage device. The research hypotheses differed, however, on the long-term effects of allowing students to use their own formula sheets. The overarching question in this study was: would students learn more overall because of the time spent organizing their sheets and identifying important concepts, or would they learn less because they had these sheets to rely on? To test these competing hypotheses a two-part study was devised.

As in most achievement studies, control of variables was important. Because formula sheet use on a single exam was hypothesized to affect a student's knowledge structure on the topics covered by that exam, and due to the integrated nature of chemistry concepts, students were assigned to one research treatment for the entire semester. Course sections were randomly designated as "experimental" and "control" groups, and all students enrolled in those sections experienced the same presentation of material and exam questions. In the "experimental" section, students were allowed to prepare their own formula sheets, whereas in the "control" section students were given the standard instructor-prepared formula sheets. Confounding effects can arise when entire sections are sampled in this way, including self-selection into certain class sections, time-of-day effects, and other variables. To control for these possible effects, the researchers had to demonstrate that the sections being compared were equivalent prior to beginning the study. It is common to make this comparison statistically on some pertinent achievement measure or demographic prior to the research intervention. In this study, the researchers used student achievement without formula sheets on several quizzes prior to the start of the research to demonstrate that the sections performed equivalently under similar test conditions.

The study by VandenPlas et al. is an example of a quantitative study, relying only on quantitative data and statistical analyses to address the research question. Although the results of the statistical tests for this study were all non-significant, showing that the intervention (student-prepared formula sheets) did not affect student performance and thus failing to confirm the initial research hypotheses, the outcome did address the question asked. In this study, the authors set out to *prove* a common classroom myth regarding how formula sheets hinder student learning. In the end the non-significant statistical results helped to *disprove* the myth instead.

The study by Hall et al. (Chapter 8) is another example of a quantitative study based on research questions regarding student achievement. In this study, the authors compared chemistry content knowledge between students taking chemistry courses during a 3-week summer session with those in a traditional 15-week semester. In addition, chemistry content knowledge of students in these two course types was compared by life experience (time since high school graduation). As previously discussed, the authors first had to define "chemistry content knowledge" in order to devise a method of addressing their research questions. In this case, student achievement on an in-class exam was used as a measure of chemistry content knowledge. This provided quantitative data, and allowed for a statistical comparison between groups of students.

A unique aspect of this study is that instead of collecting a new set of data to address their research question, the authors conducted a secondary analysis of existing longitudinal data. A secondary analysis is an examination of data that goes beyond what was originally intended when the data were collected. The researchers discuss the pros and cons of using a secondary analysis in their chapter, but the main advantage is clear: data *already existed* that would allow the researchers to investigate their research questions. The data only needed to be analyzed in a different way. This is a good use of longitudinal data, which is costly both in terms of time and money to collect. This study utilized four years worth of data, which would have required a significant amount of effort for the researchers to collect. Instead, the researchers were able to utilize data that had already been collected for a different purpose. Although this gave the researchers less freedom to design the study, the tradeoff in the amount of data collected was deemed worthwhile.

Results of this study showed significant achievement differences between students in the two courses. The statistical results demonstrate that students enrolled in the 3-week summer session gained more chemistry content knowledge than those in the traditional 15-week semester. Students who had graduated from high school more than 5 years prior to taking the course were also shown to possess more chemistry content knowledge than those who graduated less than 5 years prior to taking the course. As in the Bunce and VandenPlas (Chapter 2) study, the existing data could not address *why* these differences occurred, but the data did allow the researchers to address their initial research questions regarding *if* the differences existed. The authors suggest that a mixed-methods study, using qualitative data such as interviews in addition to the quantitative analysis of student performance, would have been a good methodology to address these additional research questions. Using existing data restricted this study to a quantitative design.

The article by Neiles et al. (Chapter 5) is an example of a project devoted not to chemistry content, but to student behavior. This study sought to identify the pattern of student attention during class specifically, at what point attention lapses occurred and the length of these lapses. Because the research question dealt with timing issues, it was important for the researchers to devise a real-time data collection method. While the researchers could have collected retrospective reports from students after class or as part of a daily or weekly survey, this would not have adequately addressed the time-sensitive nature of the research questions, such as whether attention remained constant during lecture, or varied based on teaching pedagogy. Instead, the researchers used a novel data collection method—utilizing student response systems ("clickers") to collect student self-reports of attention lapses in real-time during class.

The data collected in this study is qualitative in nature (student self-reports of attention lapses), but the use of clickers converts this to quantitative, numerical data (number of lapses, the time each lapse occurred, and student-report of lapse length). From there, the authors use statistical methods to analyze the data specifically, significance testing used to compare the occurrence of attention lapses throughout the courses studied. This use of significance testing methods shows that it is not limited to achievement data or test scores as described in previously discussed studies. The same statistical methods can be applied to any set of quantitative data as in this case where behavioral data reported by students themselves was used.

In this study, two of the three courses investigated showed nonsignificant results. The third course showed more variation in attention lapses reported by students. This demonstrates the importance of gathering a robust data set, such as the three courses, rather than just one, used in this study. The courses were analyzed separately due to the inherent differences among them, but these separate analyses also allowed the researchers a finer level of investigation. Here the significant results found for a single course may have been lost if data from these three disparate courses had been aggregated.

Qualitative Research Designs

In many cases, quantitative data is not appropriate for addressing a given research question. Researchers sometimes seek to investigate research questions describing a particular classroom situation, discussing *why* a particular effect occurs, or the process by which educational changes occur. To answer questions like these, qualitative data is often a more suitable and powerful tool. Qualitative data can take the form of interviews, surveys, observations, document analysis, or field notes (*3*). Rather than using inferential statistics, qualitative coding and descriptive statistics (describing the frequency of certain occurrences, for example) are more appropriate methods of analyzing qualitative data. Many studies described in this book include at least some qualitative methods to aid in data triangulation resulting in mixed-methods designs. The studies described here, however, focus exclusively on qualitative designs.

Studies Utilizing Qualitative Methods

The chapter by Daubenmire et al. (Chapter 7) explored two research questions regarding whether a curricular intervention would result in high school chemistry teachers changing their instructional practices to include more inquiry activities. If so, then a second question concerning if the teachers would perceive shifts in student abilities to perform such inquiry activities was asked. These research questions focus not on student knowledge acquisition itself, but rather on classroom practices and how to bring about change in such practices. To answer these questions, the researchers used qualitative research methods involving the practices of teachers and students at multiple institutions.

The first method used to gather data in this study was the use of focus groups. Conducting focus groups at various points throughout the study enabled the researchers to gather feedback from multiple teachers simultaneously, without having to interview each teacher individually. The researchers then qualitatively coded the teacher feedback using a theoretical framework appropriate for their investigation called SWOO (Strengths, Weaknesses, Obstacles, and Opportunities of program implementation). This data provided evidence that teachers were indeed moving in the direction of adopting more inquiry in the classroom as a result of the support provided within the teacher training and curricular initiative.

To triangulate this data, the researchers used two surveys of the teachers themselves. The first was a self-assessment developed specifically for this project, asking teachers to rank their own progress and that of their students towards achieving specific program benchmarks (including conducting inquiry activities). The second was an existing survey, called the Survey of Enacted Curriculum (SEC), which was given to both teachers taking part in the curricular and training intervention being studied and those who were not. This survey asked teachers to identify the time they spent in various classroom activities, including focusing on specific concept or skill developments. Data from both of these surveys agreed with the earlier focus group data, and showed that teachers taking part in the curricular intervention spent more time on average using inquiry activities or developing inquiry skills than those who were not part of the initiative.

Finally, to further investigate teacher and student use of inquiry, researchers analyzed reports made by program coaches from the training initiative. The program coaches worked with teachers one-on-one on a regular basis to help them adopt the new curricular approach and to mentor them in the classroom. The coaches provided regular reports to the researchers, and these reports were analyzed for both the occurrence and frequency of representative quotes, which the researchers used in their analysis. This study is a good example of a qualitative research design, but perhaps more importantly, shows the importance of data triangulation. Although the researchers could have collected only one type of data (focus group feedback, for example), they collected multiple data sets from multiple sources, including teachers involved in the study, teachers outside the study, and program coaches to inform their conclusions. This provided a rich dataset that the authors used to directly address their research questions. The end result demonstrated that teachers (and their students) exposed to the particular curricular and teacher training intervention were, in fact, successful in moving towards more inquiry based activities in the classroom.

The study by Pienta (Chapter 9) is another example of a study that focused on non-content material related to chemistry, and which used qualitative methods, in this case, surveying. In this study, the author was interested in how students valued various course components, such as lecture or textbooks, in their learning. This study also serves as an example of longitudinal research, in that data was collected for 10 years using a single instrument. This allowed the researcher to identify trends in student responses and how these trends changed over time. While data from a single data collection may sometimes be skewed by mitigating factors, a longitudinal study provides a robust dataset for analysis and increases confidence in the interpretation of the results of the analyses by the author.

Pienta's research used a Likert-scale survey to gather information from students regarding self-reports of the relative value and time spent interacting with each course component. Because of the nature of the research questions being investigated, inferential statistical analyses and significance testing were not appropriate for this study, even though the qualitative data collection method (surveying) used provided numerical data (Likert scores). Instead, descriptive statistics, including frequency counts of student responses in each Likert category, are reported and graphed, allowing the researchers to identify the students' value of each component. Changes in these trends over time were also investigated. Additional descriptive statistics such as mean and standard deviation were reported for student self-reports of time spent on various course components and non-course related activities.

Using Test Analysis Theories To Address Research Questions

One thing that many of the previously discussed studies have in common is the use of an instrument, whether quantitative or qualitative, to collect data. When chemists hear the word "instrument" they no doubt think of an HPLC or GC-MS, but for chemical educators, "instruments" are data collection tools such as attitudinal surveys and chemistry content exams. While many of the studies discussed in this book make use of existing instruments, such as the GALT or TOLT exams, several also develop their own exams or surveys. Although the selection, development, and validation of these instruments are not discussed in depth in these individual studies, the chapter by Barbera and VandenPlas (Chapter 11) provides a discussion of the considerations that should be made when conducting research that utilizes such instruments. The authors discuss existing literature on how to develop valid and reliable instruments for use in research, including how existing instruments may need to be evaluated before use in a study. For further discussion on designing tests and surveys, consult Scantlebury and Boone's chapter (*4*) in a previous symposium series volume.

Once a robust instrument has been built or selected, there are many statistical tools available to analyze the data collected. Classical Test Theory (CTT) and Item Response Theory (IRT) are two frameworks that can be used to analyze data collected from an instrument, and two studies in this book, utilizing these models are discussed below.

Classical Test Theory

The study by Holme and Murphy (Chapter 12) is an example of research utilizing Classical Test Theory (CTT) to analyze student performance data on a developed instrument. In this study, the researchers looked at a large data set collected from students who had taken the American Chemical Society's Paired-Questions exams (first- and second-semester) at different institutions. In this case, item level statistics were used not only to develop the instrument itself (which was not the focus of this particular study), but also to separately investigate a research question regarding student achievement on two different types of chemistry content questions, namely, conceptual and algorithmic questions.

Although the researchers discuss how item statistics from CTT, such as difficulty and discrimination, were used to select the items that ultimately became part of the final instruments, the more interesting use of these statistics comes in analyzing student performance on the final instruments themselves. Item difficulty, in particular, is used to compare traditional and conceptual questions on a single topic to identify where students have difficulty. The item level statistics provide a picture of the general trends in student achievement for both conceptual and algorithmic questions.

This study is an example in which overall student achievement on the exam, as used in some of the studies discussed previously, could not directly address the proposed research questions. Even pooling all traditional questions and comparing them to the pool of all conceptual questions would have lost the fine-grained nature of this analysis, and provided an unconvincing answer to the research questions presented in the study. Analyzing individual item pairs allowed the researchers to compare traditional versus conceptual questions for individual topics in chemistry, and to tease out which topics were more difficult for students. The authors used these results to address the question of whether students were more likely to successfully answer traditional or conceptual questions today, as compared to results from earlier research.

Item Response Theory

The research presented by Shurmeier et al. (Chapter 10) is another example of analyzing a large dataset collected over many years on a large number of students. In this study Shurmeier et al. used Item Response Theory (IRT) in a manner similar to that of Holme and Murphy's use of CTT as discussed previously (Chapter 12), in order to answer research questions regarding student achievement and capabilities in chemistry.

IRT, like CTT, can be used to determine item difficulties and discriminations, but IRT can also assign students to individual ability levels. The authors indicate that results from early IRT analyses were used to eliminate poorly performing questions from the final instrument and to measure test reliability. Here, the statistics were used to directly answers the main research question, namely, are some topics in chemistry more difficult for students than others?

Using IRT, the authors were able to identify topics that were consistently easy for students, such as unit conversion and balancing equations, and those that were consistently difficult, such as the particulate nature of matter and intermolecular forces. The analysis also allowed the researchers to analyze students' performance by their ability, and to predict which topics would be difficult for different ability-level students. This reflects a fine grain of detail, and allows the research to identify the best questions for sorting students by letter grade (A, B, C, D, and F). For example, questions dealing with the particulate nature of matter were routinely missed by students who received a low C, D, or F on the exam, while A, B, and high C students did not demonstrate difficulty with these questions. This same analysis was carried out for other types of questions and in this way, it was possible to identify topics that were both difficult for *all* students enrolled in the course and which were difficult for only lower performing students.

To address this research question, Shurmeier et al. needed a research design that would allow them to compare student achievement on questions across a large number of topics. In this case, achievement was defined as student performance on course exams, which allowed for multiple questions on each topic to be tested. Data were collected on multiple versions of the in-class exams over several years, using several thousand students. While there may be other research methodologies capable of addressing the research question in this study, the large amount of data collected was handled well by IRT thus providing a robust statistical analysis.

Assembling a Body of Literature To Address Complex Research Questions

Occasionally we are faced with a research question for which a large body of research already exists. A thorough analysis of the literature is a time consuming process, but synthesizing existing work can frequently provide satisfying results. In situations where individual research studies each add a piece to the puzzle, it may take several years of research, synthesized into a single body of work, to address more complex research questions. Two chapters in this book provide discussions of how the researchers have compiled a large body of work through their own research, and that of others, to best address long-standing myths from multiple angles.

Integrating Multiple Experiments

The chapter by Oliver-Hoyo (Chapter 4) is one such example. In this case, the author presents a review of her own research, consisting of four separate studies on the use of a new curriculum entitled $cAcL_2$ (concept Advancement through chemistry Lab-Lecture). The benefit of a review of the literature in this case is that it allows the data from multiple independent studies to be compared simultaneously and considered as a whole. This provides triangulation for the individual datasets, and offers compelling evidence to address the research question of how this particular curricular intervention (a lab-lecture hybrid) compares to the effectiveness of covering content with traditional lecture-based methods.

This chapter discusses four separate studies that have previously been published, and which provide evidence for the efficacy of the $cAcL_2$ method. The first study is an achievement study, comparing exam scores between students experiencing the $cAcL_2$ method and those in a traditional lecture class. This was a quantitative design, in which exam scores were compared using statistical significance tests. The results of these tests suggested that the $cAcL_2$ students performed significantly better than the traditional students on many exams.

Two attitudinal studies were also conducted, the first using surveys developed specifically for this project, and the second using an array of qualitative data collection methods including interviews, surveys, student self-reports via journaling, and field note entries from the researchers themselves. Although these qualitative data on student attitudes could not address how the $cAcL_2$ method influenced students' understanding of chemistry, they were used to support the observation that students viewed the $cAcL_2$ method more positively as the semester progressed.

Finally, a study was described in which the researcher studied students' higher order cognitive skill development. To do this, qualitative data were collected in the form of student reflections on their own problem solving methods after completing several conceptual chemistry problems. The problem solutions themselves were not studied. Instead, the retrospective reflection on the *methods* the student used to solve the problems was qualitatively coded by the researcher using the theoretical framework of Bloom's taxonomy. These codes were compared over time resulting in the emergence of a general trend in which students exposed to the $cAcL_2$ method showed increased use of higher order (abstract) reasoning to solve problems.

Taken together, these four studies provide a well-rounded view of the effects of using the $cAcL_2$ method in the classroom. While each study individually is a crucial piece to the puzzle, combining multiple studies in this way provides a much clearer overall picture of the body of research used to address the over arching research question.

Addressing Pervasive Myths

Williamson (Chapter 6) provides another review of the literature, including many of her own studies, in her chapter discussing the use of visualizations in the chemistry classroom. This review aggregates over a decade's worth of research on the topic, and uses this existing data to challenge several myths about the use of visualizations, such as animations, in chemistry.

This chapter demonstrates how pervasive myths can be in the area of teaching and learning despite existing proof to the contrary. For example, the first myth addressed by Williamson centers on how much time must be sacrificed from lecture in order to use animations. To counter this myth, Williamson cites her own 1995 study, in which students' conceptual understanding of the particulate nature of matter was compared after viewing animations for varying amounts of time. This early study provides compelling evidence that viewing animations does not require sacrificing large amounts of class time to reap significant benefits. Yet this study completed over a decade ago and providing solid statistical evidence to demonstrate this point, has not been sufficient to completely put this particular myth to rest.

Williamson discusses six myths in total, each with a body of literature that should effectively challenge the myth. The fact that these myths are so pervasive, even in the face of such compelling research, is a testament to the power that myths have within the culture of education. This chapter shows that simply conducting and publishing good research is not enough. Researchers must advocate against myths and continue to work to educate chemical educators of the research done to disprove them.

The Williamson chapter is also a good example of the type of literature search that should be undertaken in the process of forming one's own research questions.

Producing a review of the literature of this scope allows one to identify common theoretical frameworks in the field, holes in the existing research, and prevents the researcher from reinventing the wheel. This type of review aids in the development of robust research questions that can serve as the starting point for quality research. Such research not only challenges common myths, but adds value to the body of literature as well.

Conclusions

This book encompasses a wide variety of research questions. The researchers who addressed these questions each chose a different methodology including quantitative, qualitative and mixed methods designs. Some researchers chose questions that rely heavily on statistical analysis to answer. while others used the wealth of information in the literature to help challenge myths. Each of these methods provides different advantages and disadvantages. The researcher's role is to match the questions asked to the methodology that best addresses those questions. Since myths are tied to beliefs about teaching and learning, myths are pervasive in the culture of education and are difficult to isolate and confront. It requires carefully crafted research questions and well designed studies to confront these myths. Only with the data from thoughtful experiments, such as those presented in this book, can teaching and learning myths be effectively challenged.

References

1. Sanger, M. J. Using Inferential Statistics to Answer Quantitative Chemical Education Research Questions. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2007; pp 101−133.
2. Towns, M. H. Mixed Methods Designs in Chemical Education Research. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2007; pp 135−148.
3. Bretz, S. L. Qualitative Research Designs in Chemistry Education Research. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2007; pp 79−99.
4. Scantlebury, K.; Boone, W. J. Designing Tests and Surveys for Chemical Education Research. In *Nuts and Bolts of Chemical Education Research*; Bunce, D. M., Cole, R. S., Eds.; ACS Symposium Series 976; American Chemical Society: Washington, DC, 2007; pp 149−169.

Subject Index

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.ix002

July 1, 2012 | http://pubs.acs.org
Publication Date (Web): September 26, 2011 | doi: 10.1021/bk-2011-1074.ix002

In Investigating Classroom Myths through Research on Teaching and Learning; Bunce, D.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2011.